

Topics in Current Chemistry Collections

Henning Jacob Jessen *Editor*

Phosphate Labeling and Sensing in Chemical Biology

 Springer

Topics in Current Chemistry Collections

Journal Editors

Massimo Olivucci, Siena, Italy and Bowling Green, USA

Wai-Yeung Wong, Hong Kong

Series Editors

Hagan Bayley, Oxford, UK

Kendall N. Houk, Los Angeles, USA

Greg Hughes, Codexis Inc, USA

Christopher A. Hunter, Cambridge, UK

Seong-Ju Hwang, Seoul, South Korea

Kazuaki Ishihara, Nagoya, Japan

Barbara Kirchner, Bonn, Germany

Michael J. Krische, Austin, Texas

Delmar Larsen, Davis, USA

Jean-Marie Lehn, Strasbourg, France

Rafael Luque, Córdoba, Spain

Jay S. Siegel, Tianjin, China

Joachim Thiem, Hamburg, Germany

Margherita Venturi, Bologna, Italy

Chi-Huey Wong, Taipei, Taiwan

Henry N.C. Wong, Hong Kong

Vivian Wing-Wah Yam, Hong Kong

Chunhua Yan, Beijing, China

Shu-Li You, Shanghai, China

Aims and Scope

The series *Topics in Current Chemistry Collections* presents critical reviews from the journal *Topics in Current Chemistry* organized in topical volumes. The scope of coverage is all areas of chemical science including the interfaces with related disciplines such as biology, medicine and materials science.

The goal of each thematic volume is to give the non-specialist reader, whether in academia or industry, a comprehensive insight into an area where new research is emerging which is of interest to a larger scientific audience.

Each review within the volume critically surveys one aspect of that topic and places it within the context of the volume as a whole. The most significant developments of the last 5 to 10 years are presented using selected examples to illustrate the principles discussed. The coverage is not intended to be an exhaustive summary of the field or include large quantities of data, but should rather be conceptual, concentrating on the methodological thinking that will allow the non-specialist reader to understand the information presented.

Contributions also offer an outlook on potential future developments in the field.

More information about this series at <http://www.springer.com/series/14181>

Henning Jacob Jessen
Editor

Phosphate Labeling and Sensing in Chemical Biology

With contributions from

G. Michael Blackburn • Kayla Borland • Ilya Captain • Amit K. Dutta
Susanne Ermert • Dorothea Fiedler • Stephan M. Hacker
Itaru Hamachi • Henning Jacob Jessen • Jacek Jemielity • Yi Jin
Joanna Kowalska • Patrick A. Limbach • Alan M. Marmelstein
Marcin Warminski • Andreas Marx • Robert W. Molt Jr.
Javier Moreno • Akio Ojida • Adolfo Saiardi • Pawel J. Sikorski
Miranda S. C. Wilson • Jirarut Wongkongkatep



Springer

Editor

Henning Jacob Jessen
Institute of Organic Chemistry
Albert-Ludwigs-University of Freiburg
Freiburg im Breisgau
Baden-Württemberg, Germany

Originally published in *Top Curr Chem (Z)* Volume 375 (2016),
© Springer International Publishing Switzerland 2017

ISSN 2367-4067 ISSN 2367-4075 (electronic)
Topics in Current Chemistry Collections
ISBN 978-3-319-60356-8 ISBN 978-3-319-60357-5 (eBook)
DOI 10.1007/978-3-319-60357-5

Library of Congress Control Number: 2017945302

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Preface	vii
Fluorescence Sensing of Inorganic Phosphate and Pyrophosphate Using Small Molecular Sensors and Their Applications	1
Jirarut Wongkongkatep, Akio Ojida, Itaru Hamachi	
Metal Fluorides: Tools for Structural and Computational Analysis of Phosphoryl Transfer Enzymes	35
Yi Jin, Robert W. Molt Jr., G. Michael Blackburn	
Importance of Radioactive Labelling to Elucidate Inositol Polyphosphate Signalling	67
Miranda S. C. Wilson, Adolfo Saiardi	
Applications and Advantages of Stable Isotope Phosphate Labeling of RNA in Mass Spectrometry	89
Kayla Borland, Patrick A. Limbach	
New Synthetic Methods for Phosphate Labeling	105
Amit K. Dutta, Ilya Captain, Henning Jacob Jessen	
Phosphate-Modified Nucleotides for Monitoring Enzyme Activity	153
Susanne Ermert, Andreas Marx, Stephan M. Hacker	
Chemical Approaches to Studying Labile Amino Acid Phosphorylation	179
Alan M. Marmelstein, Javier Moreno, Dorothea Fiedler	
Applications of Phosphate Modification and Labeling to Study (m)RNA Caps	211
Marcin Warminski, Pawel J. Sikorski, Joanna Kowalska, Jacek Jemielity	

Preface

The essential role of phosphate in biology has inspired researchers throughout the decades. The 1987 classic article “Why nature chose phosphates”¹ by F. Westheimer remains one of the most insightful analyses in this area. This important and fundamental question has led to several seminal sequels: “Why nature really chose phosphate”² by A. Warshel and “Why nature chose phosphate to modify proteins”³ by T. Hunter.

This topical collection addresses the focused and practical questions: How can phosphate be labeled and which key labeling techniques have emerged preeminent over the years? Although there are countless applications of phosphate labeling in chemical biology, there are select techniques upon which the most significant studies are based. The practical recurrence of these techniques is a hallmark of this collection; these examples of phosphate labeling serve as a *passee-partout* pertinent to understanding chemical biology approaches to the many phosphorylated natural products that could not be covered in this collection. In this sense, I hope that the choice of examples will be useful for researchers from other fields of research as well.

Two recent and insightful collections in *Topics in Current Chemistry*, edited by J.-L. Montchamp (*Phosphorous Chemistry I & II, 2015*) motivated the choice towards complementary topics with a more biological perspective. The term “labeling” in the title is interpreted in a broad sense as can be seen in the different contributions.

I. Hamachi covers phosphate and pyrophosphate sensors, in which the “labeling” occurs through hydrogen bonding, coordination chemistry, aggregation induced phenomena, and chemical reactions. M. Blackburn then discusses the concept of “nuclear mutation” in which the phosphate group is replaced with metal fluorides for studies into the function of different enzymes and how QM calculations aid this field of research. A. Saiardi discusses the use of radioactive phosphorous as a powerful analytical beacon and its application to understand inositol polyphosphate

¹ Westheimer F. H. (1987), *Science*, 235, 1173-1178.

² Kamerlin S. C., Sharma P. K., Prasad R. B., Warshel A. (2013), *Q. Rev. Biophys.*, 46, 1-132.

³ Hunter, T. (2012), *Phil. Trans. R. Soc. B*, 367, 2513–2516 .

metabolism. P. Limbach reviews another highly useful, stable isotopic replacement: ^{18}O vs ^{16}O -phosphates and their value in studying RNA biology by mass spectrometry. My postdocs, Amit Dutta and Ilya Captain, (and I) provide an overview of recent synthetic approaches towards organophosphates and analogues thereof. This contribution is also meant as an entry to the following contributions, which make extensive use of such modified probes. S. Hacker and A. Marx discuss the development of phosphate-modified nucleotides and how such tools can be used to monitor enzymatic activity. The contribution by D. Fiedler covers synthetic approaches to phosphorylated unstable amino-acids and peptides and how such unstable modifications can be replaced by stable analogues that enable further biological studies. J. Jemielity finally shows how many of the previously described labels and approaches were successfully applied to understand the rich biology of the RNA cap structure.

Clearly, investigations into the chemistry and biology of phosphate(s) offers many research opportunities. Such studies are at the intersection of disciplines where many exciting questions are waiting to be solved. How do cells control phosphate acquisition and distribution? How do cells sense phosphate availability? How has the phosphate group become one of the basic motifs in life? How can one design and make probes to study these processes? This topical collection offers a good starting point for those who wish to improve their understanding into the mysteries of this important modification.

Thanks are due to the many prominent researchers that have contributed to this work, some of them with articles, others with highly judicious comments. I personally would like to thank the editorial staff of *Topics in Current Chemistry* and the editorial board for choosing me to help assemble this collection.



Fluorescence Sensing of Inorganic Phosphate and Pyrophosphate Using Small Molecular Sensors and Their Applications

Jirarut Wongkongkatap¹ · Akio Ojida² · Itaru Hamachi³

Received: 28 October 2016 / Accepted: 1 February 2017 / Published online: 1 March 2017
© Springer International Publishing Switzerland 2017

Abstract The aim of this contribution is to provide an introduction and a brief summary of the principle of fluorescence molecular sensors specific to inorganic phosphate (Pi) and inorganic pyrophosphate (PPi) as well as their applications. In our introduction we describe the impact of both Pi and PPi in the living organism and in the environment, followed by a description of the principle of fluorescence molecular sensors and the sensing mechanism in solution. We then focus on exciting research which has emerged in recent years on the development of fluorescent sensors specific to Pi and PPi, categorized by chemical interactions between the sensor and the target molecule, such as hydrogen bonding, coordination chemistry, displacement assay, aggregation induced emission or quenching, and chemical reactions.

This article is part of the Topical Collection “Phosphate Labeling and Sensing in Chemical Biology”; edited by Henning Jessen.

✉ Itaru Hamachi
ihamachi@sbchem.kyoto-u.ac.jp

Jirarut Wongkongkatap
jirarut.chu@mahidol.ac.th

Akio Ojida
ojida@phar.kyushu-u.ac.jp

- ¹ Department of Biotechnology, Faculty of Science, Mahidol University, 272 Rama 6 Road, Bangkok 10400, Thailand
- ² Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan
- ³ Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510, Japan

Keywords Fluorescence detection · Molecular sensor · Phosphate · Pyrophosphate · Imaging

1 Introduction

Phosphates play a central role in the building of the most fundamental molecules in living organisms, such as DNA and RNA. Along with proteins and carbohydrates, DNA and RNA constitute the three major macromolecules essential for all known forms of life. Phosphates are also major constituents of membrane lipids (in the form of phospholipids) and are involved in many biological processes, including skeletal development and bone integrity, energy metabolism, cell sensing, and regulation of protein synthesis [1]. Approximately 85% of total body inorganic phosphate (Pi) is found in bone, primarily in association with calcium in hydroxyapatite crystals deposited onto the collagen matrix [2]. The remainder is in soft tissue, with only approximately 1% in extracellular fluids [3]. Prolonged Pi deficiency results in hypophosphatemia with accompanying serious biological consequences, such as impaired mineralization of bone resulting in osteomalacia or rickets, dysfunction of the central nervous system, increased erythrocyte membrane rigidity, abnormal function of leukocytes and platelets, weakness of rhabdomyolysis and muscle, and cardiac dysfunction and respiratory failure [4–7]. At the other end of the spectrum, hyperphosphatemia is now recognized to decrease life expectancy and lead to seizures, cardiac dysrhythmias, chronic kidney disease, muscle weakness and tetany, decreased visual acuity, soft tissue calcification, and eventually death [8–13]. Therefore, controlling the Pi concentration is critical for the well-being of the organism.

Inorganic pyrophosphate (PPi), the dimeric form of Pi and a by-product of cellular hydrolysis of ATP, DNA polymerization, and other metabolic processes, is a biologically important target given its role in many crucial reactions [14, 15]. The difference in PPi concentrations in a variety of biological environments may be a diagnostic marker for various clinical conditions. For example, abnormally high levels of PPi in synovial fluid are observed for patients with calcium pyrophosphate crystal deposition disease [16–19]. Hence, the sensing and imaging of PPi has become an important research target. Phosphates are not only essential factors in living organisms, they are also important as components of several medicinal drugs and fertilizers. Eutrophication in the aquatic ecosystem is often related to pollution from phosphates and phosphorylated compounds [20, 21]. Due to their importance in both biological and environmental fields, great efforts have been made to develop systems capable of selectively sensing phosphates and their related compounds. Several comprehensive reviews describing the recognition of phosphate molecules by artificial receptors have been published [22–32].

Molecular sensors for phosphate anions usually consist of a Pi/PPi binding (receptor) site covalently linked to a sensing unit, such as a fluorophore (Fig. 1). One of the main difficulties in designing appropriate binding sites for Pi arises from the high hydration energy of this anion (-2765 kJ/mol) that places it near

the bottom of the Hofmeister selectivity series [33]. Moreover, most of the phosphate-type anions exist in water, at different pH, in different protonated states bearing different negative charges: H_3PO_4 , H_2PO_4^- , HPO_4^{2-} , and PO_4^{3-} . For these reasons, most of receptors prepared in early studies work only in organic solvents, and even in such examples, it has been pointed out that the binding of the Pi/PPi anions to artificial receptors requires optimization of both electrostatic and hydrogen-bond interactions through topological complementarity. One of the strategies adopted in the design of receptors for Pi/PPi useful in aqueous solution involves the usage of easily protonated polycations, such as polyammonium, imidazolium, and guanidinium moieties. An early pioneer in this area was A.W. Czarnik, who designed and synthesized fluorescent chemosensors for PPI [34]. The sensing unit can be attached to the receptor (binding unit) directly so that the receptor is part of the conjugated π system of the fluorophore, or it can be separated from the receptor via a covalent linker (Fig. 1). Pi/PPi sensing is based upon a variety of signal transduction mechanisms, such as binding-induced modulation in the fluorescence/absorbance using a conventional molecular recognition means (e.g., hydrogen bonding and coordination chemistry). In recent years, more elaborate methods of fluorescence sensing have emerged which involve a metal displacement assay in which the metal ion in the complex is specifically removed by Pi/PPi, and the aggregation induces emission/quenching upon binding to Pi/PPi. Sensors exhibiting covalent bond formation/breaking upon coupling Pi/PPi binding have also been reported.

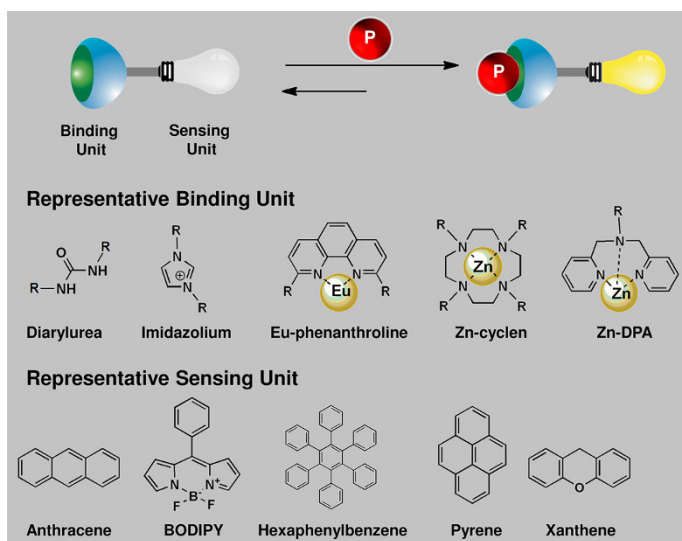


Fig. 1 Schematic overview of fluorescence sensing using a molecular sensor for phosphate (P) consisting of a binding and a sensing unit. The sensor is shown as a “Turn-ON” fluorescent sensor, but “Turn-OFF” sensors are also plausible. The chemical structures of the representative binding and sensing units are displayed

2 Molecular Sensors Based on Conventional Host–Guest Chemistry and Their Applications

Several interactions, such as electrostatic interaction, hydrogen bonds, metal ligand complexation or coordination chemistry, and hydrophobic interactions, have been employed for Pi/PPi binding, many of which are incorporated into the receptor site. From the viewpoint of the development of sensing systems in the last decade, the majority of the chemosensors for phosphate anions are rationally conjugated to an appropriate sensing unit, so that the binding events can be experimentally measured by spectroscopic changes resulting from excimer formation [35–39], intramolecular charge transfer processes (ICT) [40, 41], metal-to-ligand charge transfer processes (MLCT) [42] or photo-induced electron transfer processes (PET) [43, 44]. Receptor design still remains challenging, in particular for efficient sensing toward the complicated phosphate derivatives or the related biological events with high selectivity and sensitivity in aqueous solutions. Monitoring events over an appropriate time is also critical.

2.1 Hydrogen Bonding

The hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H, in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation [45]. Common hydrogen bond donors generally used as a receptor for phosphates include amide, ammonium, imidazolium, guanidinium, pyridinium, and urea. The strength of the hydrogen bond is generally determined by electrostatic interaction, which is typically stronger than van der Waals force but weaker than the covalent or ionic bond. Therefore, most receptors contain multiple hydrogen bond donors that are carefully designed to provide a preorganized binding unit suitable for several different anions, including Pi/PPi (Fig. 2). Unfortunately, these receptors do not work well in aqueous solution, only showing an excellent sensing capability in pure organic solvent or an aqueous/organic solvent mixture because in organic solvent, the solvation energy is not that strong and supramolecular interactions play a more important role.

2.1.1 Pi Sensing

Kumar and Srivastava [38] reported a protonated sensor built on a pyridine-2,6-bis-carboxamide framework as a binding unit and pyrene as fluorophore **1** (Table 1). This system showed a blue fluorescence corresponding to the pyrene monomer and a green fluorescence resulting from pyrene excimer upon the addition of NaH₂PO₄ or NaHSO₄ in CH₃CN. In the presence of perylene monoimide-based red emitter (PMI) and in response to the addition of these oxyanions, the sensor gave rise to white light emission visible to the human eye due to the panchromatic emission. UV/Vis and ¹H-nuclear magnetic resonance (NMR) spectroscopy studies suggested the involvement of hydrogen bonding in addition to electrostatic interactions in the

stabilization of the anion–sensor complex [38]. Because no evidence of energy transfer between the pyrene and PMI was found, the emission perceived as white light is simply a composite emission of the two fluorophores emitting independently.

A selective detection of H_2PO_4^- in CH_3CN by a complete Turn-OFF fluorescence emission with tetraamide-based receptors bearing quinolyl moieties (**2**; Table 1) was investigated in 2013 by Kondo and Takai [46]. Similar but less significant fluorescence and UV–Vis changes of sensor **2** were recorded upon the addition of CH_3COO^- , HSO_4^- , and Cl^- . The results of the UV–Vis and fluorescence titrations of sensor **2** imply that the nitrogen atoms of the quinolyl groups play a crucial role in the discrimination between H_2PO_4^- and CH_3COO^- because they act as hydrogen bond acceptors for hydroxy groups of H_2PO_4^- while four amide NH groups act as hydrogen bond donors to recognize anionic oxygen atoms of H_2PO_4^- and CH_3COO^- . The high selectivity of H_2PO_4^- over CH_3COO^- was achieved because CH_3COO^- cannot form such hydrogen bonds due to the lack of a hydroxy group in CH_3COO^- . The fluorescence quenching induced by the association with H_2PO_4^- over CH_3COO^- could be attributed to PET.

A series of macrocyclic sensors based on benzimidazolium and urea appended with acridine **3** (Table 1) for ratiometric sensing of H_2PO_4^- were reported in 2013 by Martinez and Gao [47]. Adding 3.0 eq. of H_2PO_4^- tetrabutyl ammonium salt to the solution of sensor **3** resulted in quenching of fluorescence by 68% at 430 nm and enhancement of fluorescence by 4.3-fold at 501 nm, which could be attributed to the anion-induced acridine excimer, resulting in the fluorescent color change from blue (430 nm) to green (501 nm) (Fig. 3). However, HSO_4^- was able to give rise to the excimer peak of the acridine derivative **3** at 501 nm. Ratiometric sensing seems to be a versatile principle for the class of benzimidazolium–urea-based receptors, showing that they could be used as ratiometric fluorescent sensors for H_2PO_4^- via the mechanism of anion-induced fluorophore dimer formation [47, 48].

Luis, Vila, and coworkers [49] investigated an acridine-based pseudopeptidic receptor which showed a high fluorescence Turn-ON specific to H_2PO_4^- in CHCl_3 . The macrocyclic **4** (Table 1) was found to display an increase of fluorescence emission corresponding to a new band centered at 510 nm, whereas the original fluorescence at 420 nm disappeared in the presence of H_3PO_4 . The response toward other anions was practicably negligible. As already mentioned, the solvent has an important effect on the recognition process. In water, the solvation energy is larger

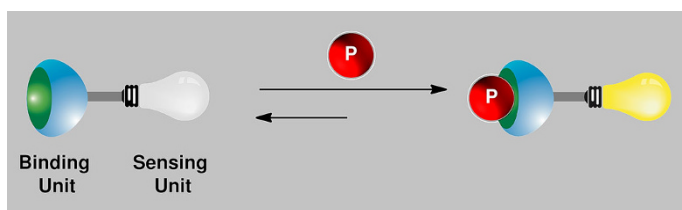


Fig. 2 Schematic overview of the fluorescent sensors utilizing hydrogen bonding for specific binding with phosphates (P)

Table 1 Inorganic phosphate sensors based on hydrogen bonding interactions

Chemical structures	Fluorescence sensing properties Proposed sensing mechanism	Binding stoichiometry (Sensor:Pi)	Binding constant (Solvent system)	Applications	Ref.
	Emission ON (~ 35-fold at 480 nm) $\lambda_{ex} = 344$ nm Excimer formation	1:1	$6 \times 10^4 \text{ M}^{-1}$ (CH ₃ CN) ($1.2 \times 10^5 \text{ M}^{-1}$ for sulfate anion)	NA	[38]
	Emission OFF (nearly perfect quenching at 355 nm) $\lambda_{ex} = 318$ nm PET	1:1	$2.76 \times 10^6 \text{ M}^{-1}$ (CH ₃ CN)	NA	[46]
	Ratiometric emission red-shift ($F_{430 \text{ nm}}/F_{501 \text{ nm}}$) $\lambda_{ex} = 357$ nm Excimer formation	1:1	$2.9 \times 10^6 \text{ M}^{-1}$ (CH ₃ CN)	NA	[47]
	Emission ON (~ 120-fold at 510 nm) $\lambda_{ex} = 357$ nm Formation of acridinium cation	NA	NA (CHCl ₃)	NA	[49]
	Emission ON (~ 100-fold at 500 nm) $\lambda_{ex} = 360$ nm Excimer formation	2:2	$3.0 \times 10^6 \text{ M}^{-1}$ (CH ₃ CN)	NA	[50]

NA not available

than the host–guest binding energy. Thus, spectra from titrations in water show an isosbestic point which results from an equilibrium observed between the triprotonated and the diprotonated receptor, and no supramolecular species are detected. In contrast, in chloroform, solvation energy is not as important, and supramolecular interactions play an important role. The behavior of compounds of **4** in acidic

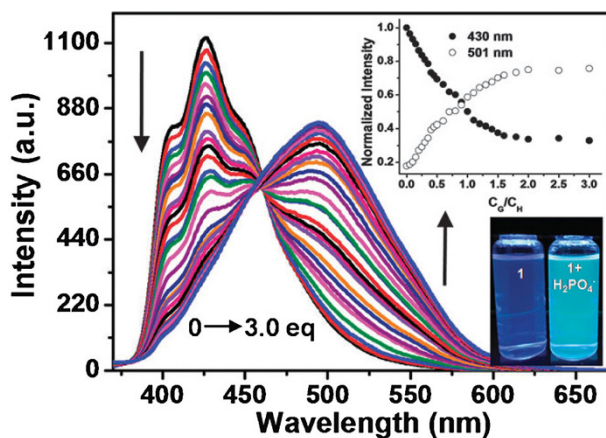


Fig. 3 Fluorescence titrations of 5×10^{-6} M sensor **3** (Table 1) excited at 357 nm with H_2PO_4^- in CH_3CN . *Insets:* *Top* Normalized emission developments at the wavelength of monomer and excimer peaks as a function of H_2PO_4^- equivalents, *bottom* fluorescent color of sensor **3** in the absence or presence of H_2PO_4^- under a UV lamp excited at 365 nm. Reproduced from Zhang et al. [47] with permission from The Royal Society of Chemistry

medium is comparable to the that shown by acridine in this study. Therefore, it can be concluded that the fluorescent moiety in the supramolecular complex formed between **4** and H_2PO_4^- in CHCl_3 is always the acridinium fluorophore, which was confirmed experimentally via fluorescence pH titrations, fluorescence lifetime, $^1\text{H-NMR}$, and X-ray studies as well as by computational calculations.

A selective sensing of H_2PO_4^- in CH_3CN driven by the assembly of anthryl amidopyridinium ligand **5** (Table 1) was demonstrated by Gong, Ning, and coworkers [50]. The ligand **5** showed a relatively low fluorescence intensity, which was ascribed to the quenching effect of a PET process from the anthracene moieties to the charged pyridinium ring. With an increase in the concentration of **5** in CH_3CN from 10^{-5} to 10^{-4} M, a new excimer emission peak centered at 539 nm was observed, which might be due to the relative proximity of the anthracene moiety at higher concentrations. This result indicated that **5** has a tendency to aggregate to some extent at high concentration. Upon the addition of various inorganic anions, including Pi, to the ligand solution, the anthryl group of **5** exhibited a strong excimer emission via H_2PO_4^- -directed assembly, while other anions showed a negligible effect. The Job plot and elemental analysis confirmed the 1:1 stoichiometry for the **5**-Pi complex. Considering there is only one anthracene fluorophore in the **5** structure and the appearance of the excimer emission between anthracene fluorophores upon addition of Pi, a plausible 2:2 stoichiometry between **5** and H_2PO_4^- was proposed, supported by absorption spectra and density functional theory (DFT) calculation [50].

2.1.2 PPI Sensing

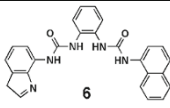
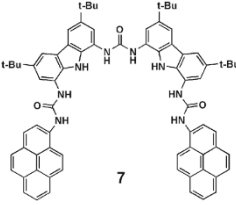
Caltagirone et al. reported fluorescent symmetric bis-ureas [51] and asymmetric bis-urea receptor **6** [52], as shown in Table 2 for PPI recognition in dimethyl sulfoxide (DMSO) and in pure water once embedded in cationic surfactant cetyltrimethylammonium bromide (CTAB) micelles. A strong interaction between PPI and **6** was observed by $^1\text{H-NMR}$, and a 1:1 receptor/anion molar ratio of the adduct in the solid state was confirmed by crystallography of the analog of **6** and $\text{H}_2\text{PPI}^{2-}$. Upon addition of PPI, the decrease in the emission band of the naphthalene fragment at 376 nm was observed to be concurrent with the formation of a new emission band centered at 476 nm. The addition of other anions, such as Pi and F^- , caused the formation of a new band at 476 nm, but its intensity was negligible compared to that observed in the case of PPI. The formation of a new red-shifted band can be attributed to a possible anion-assisted intermolecular π - π interaction involving the indole and the naphthalene groups, as suggested by theoretical calculations. Compound **6** is also able to selectively sense HPPI^{3-} in pure water by means of fluorescence quenching once embedded in CTAB micelles.

Molina et al. [53] reported a bis(carbazoly)urea bearing two pyrene fluorophores (**7**; Table 2) as a selective receptor for the recognition of PPI in anhydrous CH_3CN or $\text{CH}_3\text{CN-H}_2\text{O}$ 85/15. In anhydrous CH_3CN , **7** showed a broad and red-shifted emission band at 496 nm, assigned to the excimer emission of the pyrene moiety, and two sharp bands at 416 and 394 nm, arising from the monomer emission. The intensity ratio of excimer to monomer, $I_E/I_M = 1.06$, was barely changed in the concentration range of 10^{-7} – 10^{-5} M, indicating that the excimer emission results from an intramolecular excimer but not from an intermolecular interaction. Binding with PPI perturbs the intramolecular excimer emission of pyrene because the presence of the PPI interacting with the binding cavity forces the side chains to open in order to accommodate the guest into the cavity, thereby disabling any possibility of forming an intramolecular excimer. On the other hand, in aqueous mixture, $^1\text{H-NMR}$ and quantum chemical calculations suggest that the $7\cdot(\text{H}_2\text{O})(\text{PPI})_2$ complex displays quasi- C_2 symmetry and features all eight NH groups pointing inward to the cavity and an efficient pyrene–pyrene parallel stacking which induces an increase in excimer emission of the receptor upon binding to PPI.

2.2 Coordination Chemistry

While hydrogen bonding interactions have been widely utilized in the development of selective anion receptors and sensors in organic solvents [54], it has shown only limited success in selective anion recognition in aqueous systems [52, 53]. Because most phosphate anions and their derivatives are in water, it is crucial to bind and sense these species in aqueous medium. It is now regarded that one of the most powerful strategies for phosphate anion recognition and sensing in water is the utilization of coordination chemistry, where one or two vacant coordination sites of metal complexes are employed for binding anionic guest molecules [55–62]. In many cases, multiple metal ions are positioned on an organic scaffold at appropriate

Table 2 Inorganic pyrophosphate sensors based on hydrogen bonding interactions

Chemical structures	Fluorescence sensing properties Proposed sensing mechanism	Binding stoichiometry (Sensor:PPi)	Binding constant (Solvent system)	Applications	Ref.
 6	Ratiometric emission red-shift $(F_{476\text{ nm}}/F_{376\text{ nm}})$ $\lambda_{\text{ex}} = 328\text{ nm}$ Excimer formation	1:1	NA but experimental evidence suggests strong interaction (DMSO)	NA	[52]
	Emission OFF (~40% at 364 nm) $\lambda_{\text{ex}} = 326\text{ nm}$ NA	NA	NA	(micellar aqueous solution of cationic surfactant CTAB)	
 7	Emission OFF (~71% at 500 nm) $\lambda_{\text{ex}} = 345\text{ nm}$ Excimer deformation	1:1	$\text{Log } K_{11} = 7.00 \pm 0.57$ (anhydrous CH_3CN)	NA	[53]
	Emission ON (~3.5-fold at 500 nm) $\lambda_{\text{ex}} = 345\text{ nm}$ Excimer formation	1:2	$\text{Log } \beta_{12} = 13.60 \pm 0.63$ (CH_3CN -water 85/15 v/v)		

λ_{ex} excitation wavelength, NA not available

distances to allow an anion guest to bridge the metal centers, providing a means of introducing selectivity for a specific guest (Fig. 4).

While a number of metal ions have been used in such receptors, including those of the main group, transition metals, and lanthanides, Zn^{2+} is among the most commonly employed, particularly where the guests are phosphate derivatives [28]. Kimura and his group were the first to demonstrate that the macrocyclic Zn^{2+} complex exhibits excellent binding affinities to phosphate anions and the derivatives [63, 64], inspired by the binding sites of metalloenzymes, in which phosphates act as substrates or inhibitors by reversibly coordinating to one or more Zn^{2+} ions in the active site [65]. More recently, Hamachi and his group [43, 44] proposed the Zn^{2+} -dipicolylamine (DPA) complex as a versatile binding motif for phosphate anions in water and biological fluids. The DPA is a tridentate ligand comprised of three nitrogen donors that provides good selectivity toward Zn^{2+} over other biologically relevant metal ions and leaves coordination sites free for Pi/PPi binding. One of the advantages of the Zn–DPA receptor is its simplicity in terms of synthesis and easy modification for optimization of sensor design. Subsequent to the studies of Hamachi et al. [43, 44], many types of Zn–DPA-based molecular/supramolecular sensors have been reported [28].

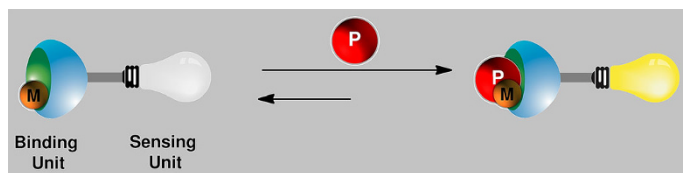


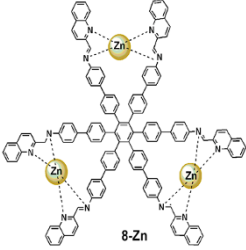
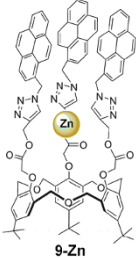
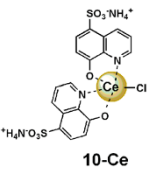
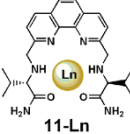
Fig. 4 Schematic overview of the fluorescent sensors utilizing coordination chemistry for the specific binding toward phosphate (*P*). *M* Metal

2.2.1 *Pi* Sensing

Bhalla, Kumar, and coworkers [55] reported the recognition of AMP and H_2PO_4^- in ethanol–tetrahydrofuran (THF; 3/1) using a Zn ensemble of a unique propeller-shaped hexaphenylbenzene derivative appended with quinoline moieties **8** (Table 3). By itself, sensor **8** does not exhibit any fluorescence emission in ethanol-THF (3/1) due to PET from imino nitrogen to a photoexcited hexaphenylbenzene moiety [66]. Upon addition of increasing amounts of Zn^{2+} ions (0–20 eq.) a strong fluorescence emission band appeared at 438 nm. This fluorescence emission band is attributed to the interaction between Zn^{2+} ions and imino nitrogens and the nitrogen atoms of the quinoline moieties as a result of which the PET is suppressed, resulting into fluorescence enhancement. Upon the addition of H_2PO_4^- between 0 and 6 eq. the emission band at 438 nm was drastically quenched, but continued addition of Pi to 6–12 eq. resulted in an increase in a new blue-shifted band at 366 nm. This result indicates the weakening of the existing **8**–Zn bond due to the interaction of H_2PO_4^- with Zn. On the other hand, the addition of AMP led to the enhancement of emission intensity along with a slight blue shift in the signal from 438 to 431 nm while no significant fluorescence change was observed upon addition of other anions. A multichannel molecular keypad system based on three different chemical inputs (Zn^{2+} , H_2PO_4^- , AMP) that switches between two different fluorescent outputs at 431 and 366 nm has been successfully constructed [55].

Yamato and collaborators [56] investigated a pyrenyl-linked triazole-modified homooxacalix [3] arene based ratiometric fluorescent receptors **9** selective for H_2PO_4^- ions (Table 3). This system exhibited a cascade signal output for the ligand toward Zn^{2+} and consequently H_2PO_4^- through switching of the excimer/monomer emission of pyrene at 485/396 nm from the “ON–OFF” to the “OFF–ON” type. The smaller downfield chemical shift of triazole proton suggests that the Zn^{2+} ion of complex **9** may be located in the negative cavity formed by the nitrogen-rich triazole ligand and the carbonyl group. In particular, the coordination force of the Zn^{2+} ion would prevent the three pyrene moieties from maintaining π – π stacking for the excimer emission and instead leads to a concomitant increase of the monomer emission of the pyrene in the fluorescence spectra. $^1\text{H-NMR}$ results suggested that the receptor **9** and the H_2PO_4^- anion not only have strong coordination and electrostatic interactions but also have strong hydrogen bonding interactions. As a result, the ratiometric signal of I_{485}/I_{396} of complex **9** changes from the OFF to the ON state upon addition of the H_2PO_4^- ion. A design of

Table 3 Inorganic phosphate sensors based on coordination chemistry

Chemical structures	Fluorescence sensing properties Proposed sensing mechanism	Binding stoichiometry (Sensor:Pi)	Binding constant (Solvent system)	Applications	Ref.
 8-Zn	Emission OFF (~90% at 438 nm) λ_{ex} = 287 nm PET	NA	NA (EtOH-THF, 3/1 v/v)	Potential bioprobe and multichannel keypad system	[55]
 9-Zn	Ratiometric emission red-shift ($F_{485\text{ nm}}/F_{396\text{ nm}}$) λ_{ex} = 343 nm Excimer formation	1:1	$1.01 \times 10^5 \text{ M}^{-1}$ (CH ₃ CN-CH ₂ Cl ₂ -H ₂ O, 1000/1/5 v/v)	Potential molecular traffic signal with an R-S latch logic circuit	[56]
 10-Ce	Emission ON (~2-fold at 496 nm) λ_{ex} = 460 nm ICT	1:1	$3.0 \times 10^6 \text{ M}^{-1}$ (aqueous solution, pH 8)	Analysis of Pi content in fertilizer and tap water	[59]
 11-Ln	Eu complex Emission OFF (~95% at 618 nm) λ_{ex} = 276 nm NA	1:1	$\text{Log } K_b = 5.4 \pm 0.01$ (10 mM HEPES, pH 7.4)	Staining of microalgal cell (<i>Chlorella vulgaris</i> CCNM 1017)	[60]
	Tb complex Emission OFF (~98% at 543 nm) λ_{ex} = 278 nm NA	1:1	$\text{Log } K_b = 5.5 \pm 0.01$ (10 mM HEPES, pH 7.4)		

λ_{ex} , excitation wavelength, NA not available

molecular logic gate using Zn^{2+} and H_2PO_4^- ions as the chemical inputs and the fluorescence emission at 396 and 485 nm as an output signal was also reported [56].

Ganjali et al. [59] reported a bis(8-hydroxy quinoline-5-sulphonate) cerium(III) complex **10-Ce** as a novel fluorescence Turn-ON sensor specific for Pi recognition

(Table 3). The significant fluorescence enhancement might be a result of the electrostatic interaction between Pi and the **10**-Ce complex, in which two oxygen atoms of the Pi bridge interact with the center of Ce^{3+} to reduce the magnitude of the electron withdrawal via partial neutralization of the charge on the Ce^{3+} ion, thereby increasing the electron-donating character of the quinolone-5-sulphonate moiety and finally resulting in an increased efficiency of ICT [67].

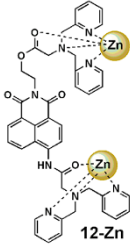
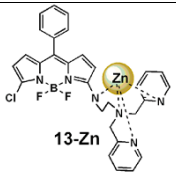
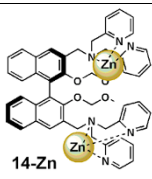
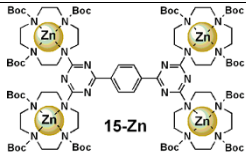
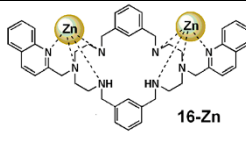
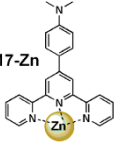
The lanthanide (III) complex of bis-amine-substituted phenanthroline-based chiral ligand (**11**; Table 3) was subsequently synthesized and characterized in 2014 by Subramaniam, Mishra, Albrecht and their coworkers [60]. Although water is known to be a quencher of lanthanide complexes, both complexes show intense luminescence in HEPES buffer (10 mM, pH 7.4), which is attributable to a $J = 2$ transition (${}^5D_0 \rightarrow {}^7F_2$) [68]. Complexes **11**-Eu and **11**-Tb show red and green emission, which displayed a selective sensing of Pi through 95 and 98% fluorescence quenching over other anions such as CH_3COO^- , NO_3^- , and SO_4^{2-} . A strong fluorescence quenching was also observed in the case of NO_2^- , ATP, ADP, and AMP. The same selectivity was observed in the case of the bis-imine-substituted phenanthroline-based Eu complex reported previously [58]. Both complexes were shown to be useable as a cell-staining reagent to monitor phosphates present in biological membranes, as demonstrated in the case of the green microalga (*Chlorella vulgaris* CCNM 1017).

2.2.2 PPI Sensing

Kim and co-workers [69] reported a 1,8-naphthalimide-DPA-Zn (II) complex **12**-Zn (Table 4) as a PPI-selective fluorescence Turn-OFF probe. The sensor **12**-Zn exhibits significant fluorescence quenching upon binding to both PPI (approx. 52%) and ATP (approx. 31%) over other anions in 95% aqueous solution containing CH_3CN . However, the binding of PPI induced a blue shift (approx. 23 nm), while a blue shift was not observed in the case of ATP. In contrast to the bis[DPA-Zn(II)] complexes discussed in other reports [70–72], where binding to PPI generally involved two Zn(II) ions, time-dependent DFT calculations at the level of B3LYP/6-31G (d) suggests that only one of the DPA-Zn(II) centers in **12**-Zn binds directly to PPI. However, no experimental data to support the proposed binding mode between **12**-Zn-PPI complex was provided in the study [69]. Molecular orbital studies predict that available electron density in either of the chelating moieties may participate in PET, which causes fluorescence quenching in the ligand. Binding of Zn^{2+} with chelating moieties eliminates the available electron density from the chelating moieties in **12** and, therefore, PET is not possible in the **12**-Zn complex. Upon binding with PPI, the photo-induced charge transfer occurs from negatively charged PPI to either the DPA unit or a fluorophore moiety. A biological application was also attempted, namely, to monitor intracellular PPI by addition of **12**-Zn to C2C12 cells with exogenous PPI.

Wu et al. [70] demonstrated a PPI selective BODIPY-based fluorescence probe **13**-Zn complex (Table 4) and its imaging in a living RAW264.7 cell line. Fluorescence quenching with 10-nm blue shift was observed upon the addition of PPI, whereas the other anions, including nucleoside polyphosphate (NPP), caused

Table 4 Inorganic pyrophosphate sensors based on coordination chemistry

Chemical structures	Fluorescence sensing properties Proposed sensing mechanism	Binding stoichiometry (Sensor:PPi)	Binding constant (Solvent system)	Applications	Ref.
 12-Zn	Emission OFF (~ 52% at 504 nm) λ_{ex} = 360 nm PET/PCT	NA	NA (CH ₃ CN-20 mM HEPES, pH 7.4, 5/95 v/v)	Fluorescence imaging using C2C12 cells with 500 and 1000 eq. exogenous PPi.	[69]
 13-Zn	Emission OFF (~ 72% at 549 nm) λ_{ex} = 465 nm NA	1 : 1	$6.9 \times 10^4 \text{ M}^{-1}$ (CH ₃ OH-5 mM HEPES, pH 7, 1/49 v/v)	Fluorescence imaging using RAW264.7 cells with Zn ²⁺ followed by 1 eq. exogenous PPi	[70]
 14-Zn	Emission ON (~ 4-fold at 383 nm) λ_{ex} = 316 nm PET	1:1	$6.7 \times 10^3 \text{ M}^{-1}$ (CH ₃ OH-10 mM HEPES, pH 7.4, 1/99 v/v)	Fluorescence imaging using HeLa cells with 2 eq. exogenous PPi	[71]
 15-Zn	Emission OFF (~ 90% at 385 nm) λ_{ex} = 290 nm Dimer formation	1:2	$> 10^6 \text{ M}^{-1}$ (25 mM HEPES, pH 7.4)	NA	[73]
 16-Zn	Emission ON (~ 21-fold at 368 nm) λ_{ex} = 302 nm ESICT	NA	6.22 log unit (20 mM MOPS, pH 7.4)	NA	[74]
 17-Zn	Emission ON (~ 500-fold at 591 nm) λ_{ex} = 440 nm NA	3:1	NA (10 mM HEPES, pH 7.4)	Fluorescence imaging of HeLa cells without addition of exogenous PPi. Hydrogel coated paper strips for PPi	[75]

 λ_{ex} excitation wavelength, NA not available

only minor fluorescence quenching or even induced a slight increase in fluorescence emission in a water–methanol mixture. $^1\text{H-NMR}$ spectroscopy and DFT calculation suggested that Zn^{2+} binding occurs mainly through the four nitrogens at the *N,N*-di(pyridin-2-ylmethyl)-ethane-1,2-diamine substituent and confirmed the binding of PPI with the **13**–Zn complex. These observations indicate that PPI binding to the Zn^{2+} results in weaker binding between **13** and Zn^{2+} . Living cell imaging using the RAW264.7 cell line showed that the green fluorescence of the BODIPY-based **13**–Zn complex disappeared when 1 eq. of PPI was applied.

The Zn complex of 1,1'-bi-2-naphthol bearing DPA units **14** (Table 4) that make PPI visible was investigated by Li, Yu, and their coworkers [71]. A fluorescence enhancement of **14** at 383 nm in 1% methanol/HEPES buffer was observed upon the addition of Zn^{2+} due to the suppression of PET from the lone pair of electrons on the DPA group. The fluorescence of the **14**–Zn complex was subsequently increased by approximately fourfold with the addition of PPI, wherein the intensity was saturated at 0.3 eq. Among the other anions tested, ATP was found to also enhance the fluorescence of **14**–Zn, and the selectivity coefficient for PPI and ATP was calculated to be 4.1/2.8, suggesting that the enhancement is dependent on both the bulkiness of the organic moiety and the number of phosphates. When the concentration of the **14**–Zn complex was raised from 1.0×10^{-5} to 2.5×10^{-3} M it was possible to detect the recognition of anions by the naked eye through precipitation formation. The formation of small particles (0.8–1.3 μm), observed by scanning electron microscopy (SEM), indicated that **14** was uniformly dispersed in the solution, due to its poor solubility in the HEPES solution, but that upon the addition of Zn^{2+} to the system, the solution became clear and no small particles (turbidity) were observed. Once PPI was added, rough and scaly solids with large surfaces were observed which confirmed the interaction between the **14**–Zn complex and PPI. The fluorescence imaging was demonstrated in HeLa cells treated with **14**, followed by Zn^{2+} /pyrithione and PPI.

König and Hamachi's group [73] reported a rigid luminescent bis-Zn(II)-bis-cyclen complex **15**–Zn specific for PPI (Table 4) in a completely aqueous solution. **15**–Zn displayed a strong fluorescence quenching which could arise from a similar dimerization resulting in π – π stacking interactions of planar benzene–triazine moieties. Electrospray ionization mass spectrometry (ESI-MS) measurement of the complex in the presence of an excess amount of PPI showed that the major species for the complex corresponded to the receptor/PPI = 1:2 complex, and the presence of a dimer of the complex **15**–Zn with PPI was clearly observed, enabling the isotope distribution to be compared with the predicted one.

Mateus, Delgado, and their coworkers [74] investigated a Zn complex of diethylenetriamine-derived macrocycle **16**–Zn, bearing 2-methylquinoline arms and containing *m*-xylyl spacers (Table 4). At pH 7.4, the ligand exhibits almost no fluorescence emission. Single-crystal X-ray diffraction studies indicated that the ligand is bound to each Zn center by three nitrogen atoms of a diethylenetriamine subunit and a quinolyl nitrogen atom, and the metal coordination sphere is completed by the binding of oxygen atoms of carbonate anions. Upon the addition of 1 eq. of PPI to **16**–Zn, a remarkable 21-fold enhancement of the fluorescence intensity at 368 nm was observed. The increase in fluorescence quantum yield of the

complex can be attributed to a significant increase of the radiative decay constant of the complexed quinoline fluorophore. Further addition of PPI caused a decrease in fluorescence intensity, which may be ascribed to the removal of Zn^{2+} from the complex by the PPI anion. No further fluorescence change was observed in the presence of 1 eq. of the other anions studied, including Pi, NPPs, and phenyl phosphate, indicating that the **16**-Zn receptor acts as a selective fluorescent sensor for PPI which leads to an important effect on excited-state intramolecular charge transfer (ESICT) in the quinoline ring. The selectivity of the **16**-Zn complex is most likely related to a steric hindrance effect caused by the bulky quinoline pendant arms.

In 2014 a long wavelength detection of PPI using a simple terpyridine- Zn^{2+} complex **17**-Zn (Table 4) was reported by Rissanen et al. [75]. The crystal structure of the **17**-Zn complex confirmed the formation of the 1:1 metal complex in which Zn^{2+} adopts a distorted trigonal bipyramidal N_3Cl_2 coordination. The **17**-Zn complex demonstrated an extraordinary sensitivity because the probe was able to detect PPI at nanomolar level with the lowest limit of detection of 0.8 nM, while the sensitivity of the other probes was invariably limited to micromolar levels. Although the unusual binding stoichiometry of 1:3 (PPI: **17**-Zn) was confirmed unambiguously by a Job plot analysis and the binding constant of the **17**-Zn and PPI was not available, a bright orange-yellow emission of **17**-Zn-PPI was imaged in a single HeLa cell due to the excellent sensitivity of the receptor. Almost the entire cell could be mapped for PPI, even regions with minimal PPI concentrations (Fig. 5). The maximum emission intensities were recorded at the nuclei, and emissions were also observed from the membranous cytoplasmic structures. The receptor is highly suitable for monitoring the biological events occurring in a single cell.

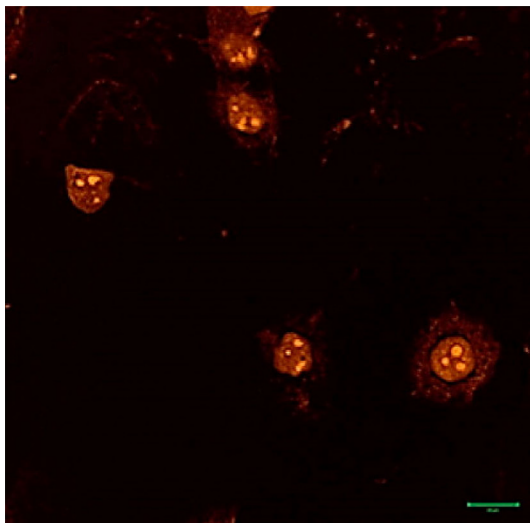
3 Molecular Sensors Based on New Sensing Mechanisms and Their Applications

Recently, several alternative approaches to Pi/PPI sensing involving metal displacement, aggregation-induced emission or quenching, and chemical reactions have emerged as simple and convenient techniques with high sensitivity. In contrast to sensing based on supramolecular chemistry such as hydrogen bonding or coordination chemistry, when the maximum fluorescence change can be recorded with a time resolution in the order of seconds, the kinetics of reaction/displacement assay-based sensing strongly depends on the nature of reaction, which is several seconds in some cases [76] while in other cases it can be very slow and more than 1 h is required to reach the highest fluorescence change [77].

3.1 Displacement Assay

In this approach, a specific metal complex is designed such that Pi/PPI has a stronger affinity for the metal ion, resulting in displacement of the metal upon Pi/PPI binding which induces a change in fluorescence (Fig. 6). Most of the metal displacement sensors demonstrate a good reversibility because the addition of the

Fig. 5 Confocal fluorescence microscopy image of HeLa cells incubated with probe **17**-Zn (50 μ M). Reprinted with permission from Bhowmik et al. [75]. Copyright 2014 American Chemical Society



metal ion can reverse the process of sensing. Another approach is the indicator displacement assay (IDA) in which an indicator is allowed to bind reversibly to a receptor in the first step. Pi/PPi with a strong affinity to the sensor is subsequently introduced into the system, causing the displacement of the indicator from the receptor, which in turn modulates the sensing unit (Fig. 6). IDA has become a popular method for converting almost any synthetic receptor into an optical sensor along the lines originally developed by Anslyn et al. [78, 79].

3.1.1 Pi Sensing

Meng, Zhang and their co-workers [80] investigated the fluorescein bearing 2-[(pyridin-2-yl-imino)methyl]phenol moiety **18**-Fe complex (Table 5) as a receptor for the highly selective detection of Pi in THF-HEPES buffer mixture. The fluorescence emission of the **18**-Fe³⁺ complex was completely quenched, which could be ascribed to the paramagnetic quenching effect of Fe³⁺ and/or MLCT. The ligand **18** displayed a high affinity to Fe³⁺ ($K_a = 1.40 \times 10^6 \text{ M}^{-1}$) over Cu²⁺ and other metal ions. The specific interaction between Pi and the **18**-Fe³⁺ complex led to the liberation of fluorophore **18**, and thus the fluorescence was approximately 9.6-fold recovered when 18 eq. Pi was added. The **18**-Fe³⁺ complex showed a negligible fluorescent response upon the addition of diverse anions but rather displayed reasonable performance (less than fourfold) with other polyphosphate species, including NPP and PPI. Studies on the quantitative deposition of **18** in MDA-MB-231 cells and its fluorescence “ON-OFF-ON” response in living cells were conducted using a flow cytometer. Staining of the Pi-pretreated MDA-MB-231 and U-343 MGa cells with the **18**-Fe³⁺ complex resulted in bright intracellular fluorescence images showing that **18** is able to display a fluorescence Turn-ON response to Pi in the living cells.

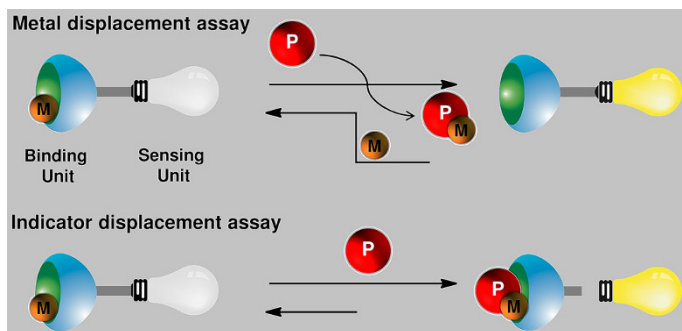


Fig. 6 Schematic overview of fluorescent sensors utilizing the metal and indicator displacement assay for the specific binding toward phosphate (*P*). *M* Metal

Table 5 Displacement assays specific to inorganic phosphate

Displacement assays	Fluorescence sensing properties Proposed sensing mechanism (solvent system)	Applications	Ref.
<p>18-Fe $\xrightleftharpoons[\text{Fe}^{3+}]{\text{Pi}}$ 18 + $\text{Fe}^{3+}\text{-Pi}$</p>	Emission ON (~ 10-fold at 515 nm) $\lambda_{\text{ex}} = 430 \text{ nm}$ Paramagnetic quenching effect of Fe^{3+} and/or MLCT (THF-20 mM HEPES, pH 7.4, 3/7 v/v)	Fluorescence imaging using MDA-MB-231 and U-343 MGA cells treated with 30/100 eq. Pi. Flow cytometry analysis of MDA-MB-231 cells treated with 30 eq. Pi	[80]
<p>19-Cu $\xrightleftharpoons[\text{Cu}^{2+}]{\text{Pi}}$ 19 + $\text{Cu}^{2+}\text{-Pi}$</p>	Emission ON (~ 4-fold at 405 nm) $\lambda_{\text{ex}} = 290 \text{ nm}$ PET (CH_3OH)	Molecular logic gate	[81]
<p>20-Zn $\xrightleftharpoons[\text{Zn}^{2+}]{\text{Pi}}$ 20 + $\text{Zn}^{2+}\text{-Pi}$</p>	Emission OFF (~ 80% at 450 nm) $\lambda_{\text{ex}} = 365 \text{ nm}$ PET (CH_3OH)	Molecular logic gate	[82]

Another report of metal displacement assay in CH_3OH was demonstrated by Hu, Ju, and their coworkers [81] in 2015. A steroid-coumarin conjugate **19** (Table 5) displayed a significant fluorescence quenching only in the presence of 9 eq. Cu^{2+}

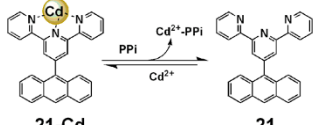
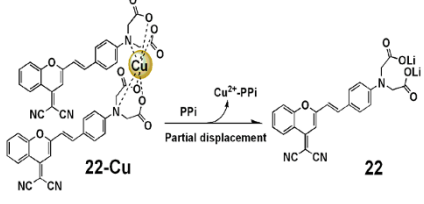
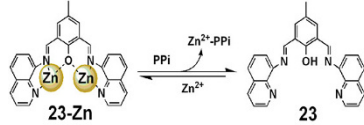
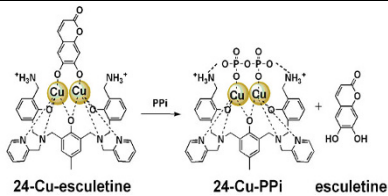
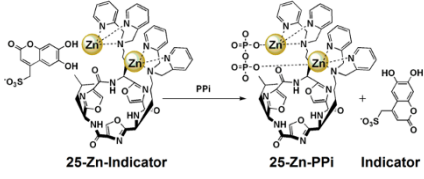
via PET resulting from the excited coumarin fluorescence in the presence of Cu^{2+} ions [67]. Furthermore, a 1:2 binding complex between the 1,2,3-triazole motif of **19** and the Cu^{2+} ion was shown using UV–Vis titration and ESI MS. When 2.0 eq. H_2PO_4^- ion was added, the absorption curve showed a dramatic change and almost reversed to the state of free compound **19**, suggesting the removal of Cu^{2+} ion from the complex by H_2PO_4^- anion. Conversely, there was no change in fluorescence in the presence of other anions, including F^- , Cl^- , Br^- , I^- , CH_3COO^- , NO_3^- , and HSO_4^- . The Turn-ON recognition ability in fluorescence for histidine was also observed as a result of the release of the free **19** after Cu^{2+} was removed from the complex to form a stable Cu–histidine complex due to the universal interaction between Cu^{2+} and amino acids. Its application as a molecular logic gate was studied with Cu^{2+} and H_2PO_4^- as input signals and the fluorescence intensity at 405 nm as output. The same research group also reported steroid–salen conjugate **20** for Zn ion recognition and cascade recognition for H_2PO_4^- in CH_3OH , showing the logic gate properties of Zn^{2+} and H_2PO_4^- [82].

3.1.2 PPI Sensing

Li, Yu, and their coworkers [83] investigated a Cd(II)-terpyridine-based complex **21**–Cd (Table 6) as a reversible ratiometric fluorescent probe for PPI detection in CH_3CN solution. The ligand has strong fluorescence emission at 421 nm, and there was a new emission peak at 553 nm upon the addition of Zn^{2+} ($I_{553}/I_{421} = 10$) or Cd^{2+} ($I_{553}/I_{421} = 30$) that did not respond to Li^+ , Na^+ , K^+ , or Ca^{2+} in CH_3CN . The red shift of the fluorescence emission of the ligand after coordination with Cd^{2+} can be ascribed to the electron-donating ability of the anthracene group, which causes ICT from the anthracene to the Cd(II)-coordinated terpyridine (see structure in Table 6). The 2:2 binding stoichiometry of the ligand and Cd^{2+} was verified by a single-crystal structure analysis and mass spectrometry. Upon the addition of PPI to the **21**–Cd complex in CH_3CN solution, the ratio of the I_{431}/I_{560} increased from 0.26 to 7.46, whereas the addition of ATP increased this ratio only from 0.26 to 2.66. The fluorescent titration, UV–Vis absorption, and high-resolution mass spectrometry results suggested that PPI could completely de-complex the **21**–Cd and liberate the ligand. This ratiometric sensing behavior of red and blue shifts upon addition of Cd^{2+} followed by the addition of PPI in fluorescence emission was reversible over five successive cycles. Finally, Li et al. [83] demonstrated the confocal fluorescence imaging of **21**–Cd complex with PPI in living RAW264.7 cells.

Zhu et al. [84] reported a near-infrared (NIR) fluorescent Turn-ON sensor **22**–Cu complex, which employs dicyanomethylene-4*H*-chromene as the fluorophore and the iminodiacetic acid group as the receptor (Table 6), for the selective detection of PPI in aqueous solution. Since α -amino acid shows a good chelating property toward Cu^{2+} [85], the incorporation of a lithium iminodiacetate group in **22** has a double effect, namely, the coordination ligand and the improvement of water solubility. There was no significant change in the fluorescence emission in the presence of other metal ions—only the addition of 5 eq. Cu^{2+} resulted in an obvious decrease in fluorescence. Job plots and mass peak analysis revealed that **22** forms a 2:1 complex with Cu^{2+} . Meanwhile, the fluorescence intensity in the NIR region of

Table 6 Displacement assay specific to inorganic pyrophosphate

Displacement assays	Fluorescence sensing properties Proposed sensing mechanism (solvent system)	Applications	Ref.
Metal displacement assay			
 <p>21-Cd $\xrightleftharpoons[\text{Cd}^{2+}]{\text{PPI}}$ 21 + $\text{Cd}^{2+}\text{-PPI}$</p>	Ratiometric emission blue-shift $(F_{431\text{ nm}}/F_{560\text{ nm}})$ $\lambda_{\text{ex}} = 326\text{ nm}$ ICT (CH ₃ CN-10 mM HEPES, pH 7.4, 1/1, v/v)	Fluorescence imaging in RAW264.7 cells treated with Cd ²⁺ followed by PPI	[83]
 <p>22-Cu $\xrightarrow{\text{PPI}}$ 22 + $\text{Cu}^{2+}\text{-PPI}$ (Partial displacement)</p>	Emission ON (~4.6-fold at 675 nm) $\lambda_{\text{ex}} = 450\text{ nm}$ NA (10 mM MOPS, pH 7.0)	Fluorescence imaging of to KB cells (human nasopharyngeal epidermal carcinoma cell) supplemented with 3 eq. PPI	[84]
 <p>23-Zn $\xrightleftharpoons[\text{Zn}^{2+}]{\text{PPI}}$ 23 + $\text{Zn}^{2+}\text{-PPI}$</p>	Emission OFF (~complete quenching at 480 nm) $\lambda_{\text{ex}} = 350\text{ nm}$ ICT (CH ₃ OH-1 mM HEPES, pH 7.4, 3/2, v/v)	Molecular logic gate Detection of DNA amplification Estimation of bacterial cell numbers through PCR	[86]
Indicator displacement assay			
 <p>24-Cu-esculetine $\xrightarrow{\text{PPI}}$ 24-Cu-PPI + esculetine</p>	Emission ON (~30-fold at 465 nm) $\lambda_{\text{ex}} = 380\text{ nm}$ Association and dissociation of the indicator with 24-Cu complex (10 mM HEPES, pH 7.0)	Monitoring hydrolysis of PPI catalyzed by pyrophosphatase Detecting PPI released during PCR	[87]
 <p>25-Zn-Indicator $\xrightarrow{\text{PPI}}$ 25-Zn-PPI + Indicator</p>	Emission ON (~8-fold at 480 nm) $\lambda_{\text{ex}} = 347\text{ nm}$ Association and dissociation of the indicator with 25-Zn complex (5 mM HEPES, pH 7.4/saline solution/Krebs buffer solution)	NA	[88]

 λ_{ex} excitation wavelength, NA not available

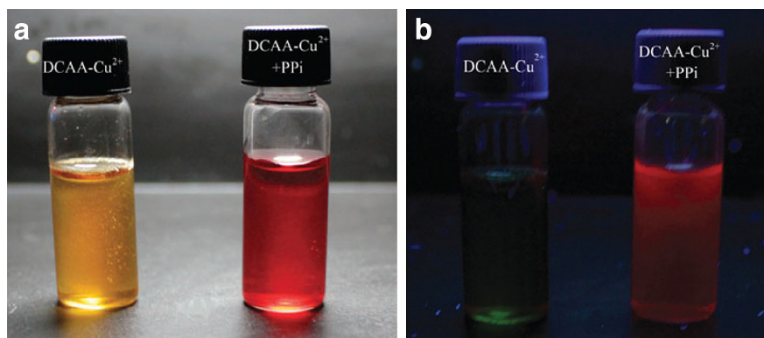


Fig. 7 Photographic images observed of the **22**–Cu complex (10^{-4} M) with the addition of inorganic pyrophosphate (PPi) at 5 eq. **a** Color change, **b** fluorescence emission change irradiated at 365 nm by a portable fluorescent lamp. Reproduced from Zhu et al. [84] with permission of The Royal Society of Chemistry

675 nm was obviously enhanced and finally stabilized upon the addition of 15 eq. of PPi. The associated color change can be easily differentiated by the naked eye (Fig. 7). Here, the fluorescence can not be recovered to the original intensity of **22**, which indirectly indicates that PPi does not form a sufficiently strong complex to completely remove Cu^{2+} from the **22**–Cu complex. To test its usefulness in a biological context, the sensor **22**–Cu was then applied to KB cells (human nasopharyngeal epidermal carcinoma cell) supplemented with 3 eq. PPi and monitored.

A diformyl–quinoline-based reversible receptor **23** (Table 6) prepared by Ramesh, Das, and coworkers [86] exhibits selective fluorometric enhancement upon the addition of Zn^{2+} in methanol aqueous solution. It is possible that the weak fluorescence of **23** is a consequence of ICT and free rotation around the azomethine ($\text{C}=\text{N}$) carbon, which brings flexibility within the ligand. In the presence of Zn^{2+} , **23**–Zn forms a rigid phenoxo-bridged binuclear Zn complex which restricts the free rotation around the azomethine carbon, thus inhibiting the ICT process, which results in enhancement of fluorescence intensity. The formation of the binuclear phenoxo-bridged metal complex **23**–Zn provides an anion binding site in each of the metal centers. The complete quenching of fluorescence was selectively observed toward PPi, while other monovalent anions and NPP showed negligible response. Job's plot and ESI–MS data revealed a 1:1 binding stoichiometry between the **23**–Zn complex and PPi, and the apparent association constant was determined to be $1.76 \times 10^4 \text{ M}^{-1}$. The change in the emission spectrum of the **23**–Zn complex upon the addition of PPi can be understood by considering the strong binding affinity of PPi toward Zn^{2+} , resulting in the release of free ligand in solution. Application of **23** as a molecular logic gate with two inputs (Zn^{2+} , PPi) and one output (emission intensity at 480 nm), and the detection of DNA amplification by PCR assay were demonstrated. The estimation of bacterial cell numbers by indirect measurement of PPi generated from the PCR reaction of the bacterial 16s rRNA universal primer was also performed in vitro.

A dinuclear-copper(II) complex **24**-Cu (Table 6) with two ammonium arms based on bis-2-[(pyridin-2-ylmethylamino)methyl]phenol as the coordinated unit was reported by Xie, Chen, and coworkers [87] in 2016. Using esculetine as a fluorescent indicator, IDA was carried out to obtain the high affinity with PPI over other anions in 10 mM HEPES buffer (pH 7.0). A fluorescence quenching of esculetine at 465 nm was observed upon the addition of 1 eq. of the **24**-Cu complex. With the addition of PPI, the fluorescence of the solution containing **24**-Cu-esculetine was recovered, indicating the liberation of esculetine from the ensemble. The ensemble of **24**-Cu-esculetine was successfully applied to monitoring the hydrolysis reaction of PPI and the released PPI from the PCR process, exhibiting its potential application in enzyme activity screening and DNA sequencing.

A family of cyclic peptide receptors bearing the DPA-Zn complex highly selective to PPI was investigated by Butler and Jolliffe [88] using indicator displacement assays in water, saline solution, and Krebs buffer. All receptors strongly bound the coumarin-based indicator, but hexapeptide scaffold receptor **25**-Zn (Table 6) had the highest affinity toward the indicator and PPI, with $\log K_a$ values of 7.3 and 9.8, respectively. The DPA-Zn unit was capable of quenching the fluorescence of the coumarin sulfonate derivative [89, 90], and the addition of 1 eq. of PPI resulted in almost complete restoration of its fluorescence intensity while the addition of ATP or ADP resulted in fourfold and 3.5-fold fluorescence enhancements, respectively. On the other hand AMP, cAMP, phosphoserine, phosphotyrosine, Pi, and the polycarboxylates acetylglutamate and Ac-Glu-Gly-Glu were not able to displace the indicator from the receptor to an appreciable extent. The enhanced selectivity of the **25**-Zn complex observed for PPI in the routinely used Krebs physiological buffer solution ensures the further development of **25**-Zn complex in a biological assay for PPI.

3.2 Aggregation Induced Emission and Quenching

Some organic molecules that are almost nonfluorescent in solution, exhibit a high capability to emit a strong fluorescence upon aggregation (Fig. 8). This exciting fluorescence phenomenon was first noted by Tang et al. [91] in 2001 from a solution of 1-methyl-1,2,3,4,5-pentaphenylsilole. They referred to this phenomenon as aggregation-induced emission (AIE) and subsequently showed that the restriction of intramolecular rotation (RIR) in the aggregates was the main cause of AIE phenomenon. Unhindered intramolecular rotation of AIE molecules under the free state leads to efficient nonradiative decay of the corresponding excited states, making them nonemissive. In view of such interesting fluorescence behaviors, AIE phenomenon have been successfully utilized to design sensitive and selective chemosensors suitable for the detection of PPI in living cells.

Chao and Ni [92] reported a terpyridine-Zn complex **26**-Zn (Table 7) for selective nanomolar PPI detection over ATP and ADP in DMSO aqueous mixture based on AIE and ICT. Complex **26**-Zn bears a donor-acceptor (D-A) structure in which the carbazole group is the donor and the terpyridine-Zn part is the acceptor; therefore, an ICT effect was observed. In fact, terpyridine-Zn complexes bearing

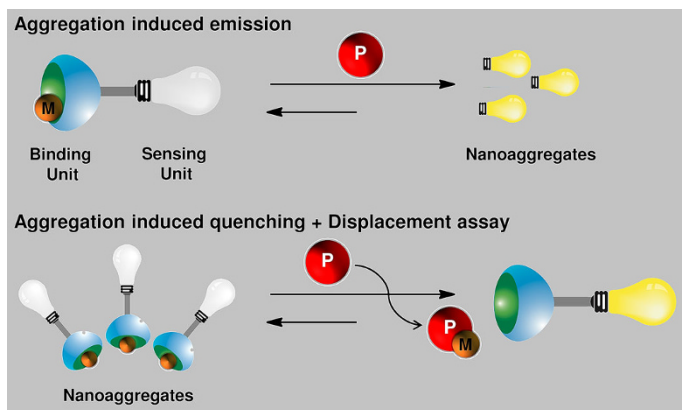
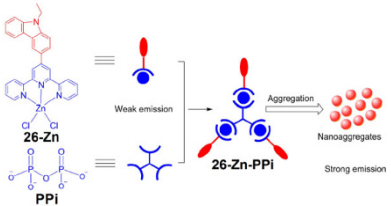
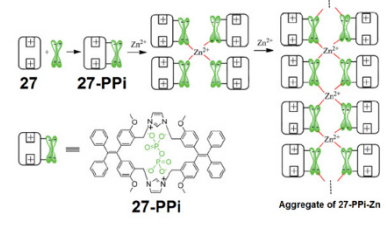
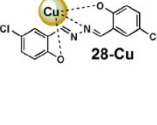
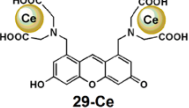


Fig. 8 Schematic overview of the sensors working in aggregation-induced emission and aggregation-induced quenching in combination with displacement assay upon binding with phosphate (*P*). *M* Metal

the D–A structure displayed different emission in solvents with various polarities, and especially weak emission was found in aqueous solution due to the strong polarity of water. When the terpyridine–Zn–PPi ensemble (3:1 binding mode) formed nanoaggregates with an average particle size of about 300 nm, as confirmed by dynamic light scattering (DLS), SEM and transmission electron microscopy, the quenching effect caused by a polar solvent, such as water, can be inhibited efficiently, leading to the strong emission at 515 nm, which is known as AIE. Other anions, including Pi, led to negligible changes for the emission of **26**–Zn but NPP induced less effects on the emission than PPI. The sensor **26**–Zn has been successfully employed for nucleus staining in living HeLa cells (Fig. 9).

A tetraphenylethylene imidazolium macrocycle **27** (Table 7) was reported to be a selective fluorescence Turn-ON sensor of PPI by Zheng et al. [93] in 2015. This positively charged macrocycle showed a maximum AIE effect in DMSO aqueous mixture, in the presence of a half eq. of Zn²⁺ with 2.4 eq. of PPI, while other common inorganic anions gave almost no response. In the presence of Zn²⁺, the UV–Vis titration of **27** with PPI was very similar to that of **27** without Zn²⁺, indicating that the metal ions did not directly interact with the macrocycle but only with the PPI. The association constant of the complex, according to a 1:1 molar ratio between **27** and PPI, as confirmed by ESI MS, was calculated to be $1.41 \times 10^4 \text{ M}^{-1}$. The crystal structure of **27** indicates that the distance between the nitrogen atoms of each imidazolium unit was approximately 5.836 Å, while the longest distance between the oxygen atoms of each phosphate unit of PPI was 5.36 Å. Therefore, the cavity of **27** is composed of two imidazolium units and is suitable for the inclusion of one molecule of PPI to form a **27**–PPI complex, driven by an electrostatic attraction. In the presence of Zn²⁺, the two component **27**–PPI complex was transformed into a five component (**27**–PPI)₄–Zn complex due to one Zn cation being coordinated by two **27**–PPI complexes. With continued coordination, an aggregate of (**27**–PPI)₄–Zn complexes formed, with the average diameter of

Table 7 Aggregation induced emission/quenching specific to inorganic pyrophosphate

Chemical structures	Fluorescence sensing properties Proposed sensing mechanism (solvent system)	Applications	Ref.
Aggregation induced emission			
	Emission ON (~ 4500-fold at 515 nm) λ_{ex} = 400 nm RIR/ICT (DMSO-10 mM HEPES, pH 7.4, 3 /7 v/v)	Nucleus staining in HeLa cells	[92]
	Emission ON (~ 25-fold at 472 nm) λ_{ex} = 347 nm RIR (water containing 0.5% DMSO)	NA	[93]
	Emission ON (~ 40-fold at 570 nm) λ_{ex} = 388 nm RIR (DMSO-10 mM Tris-HCl, pH 7.0, 2/8 v/v)	Analysis of PPI in a fetal bovine serum sample	[95]
Aggregation induced quenching			
	Emission ON (~ 318-fold at 529 nm) λ_{ex} = 500 nm AIQ/self-quenching of the xanthene fluorophore in the aggregated state (CH ₃ OH-25 mM MES, pH 6.8, 1/1 v/v)	Detection of viral infection using nucleic acid amplification reaction	[76]

λ_{ex} excitation wavelength, NA not available

aggregates being up to 2800 nm as confirmed by DLS; therefore, the strong fluorescence was probably due to a RIR mechanism [94].

Another AIE fluorescence mechanism using 5-chlorosalicylaldehyde azine **28** induced by Cu²⁺ (Table 7) followed by PPI was reported by Tong et al. [95] in 2015. The structure of salicylaldehyde azine derivatives provided the AIE characteristic, as well as the potential chelating sites for metal ions [96, 97]. Compound **28** showed a strong fluorescence in the DMSO volume fraction range of 10–50%, which was considered to be AIE fluorescence in its aggregated state. In contrast, in a high DMSO volume fraction range of 70–90%, a weak fluorescence

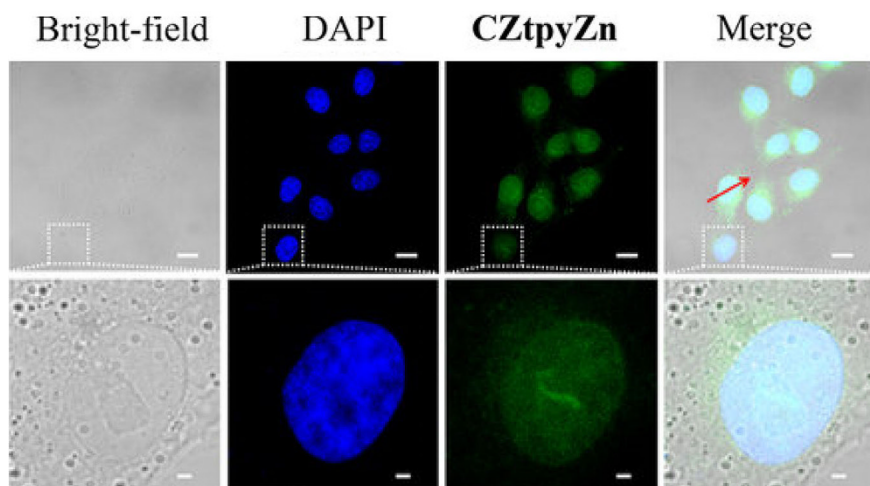


Fig. 9 Confocal fluorescence images of HeLa cells incubated with **26**-Zn (5×10^{-6} M) for 30 min and then further incubated with 4',6-diamidino-2-phenylindole (DAPI; 5×10^{-6} g mL $^{-1}$) for 10 min. The *blue* channel for DAPI excited at 405 nm and the *green* channel for **26**-Zn excited at 488 nm. Scale bar: 20 μ m. This work is licensed under a Creative Commons Attribution 4.0 International License

was observed due to the free rotation of N–N single bond in solution state. According to DLS data, the mean diameter of the fluorescent aggregates of **28** was 533.5 nm. The addition of Cu $^{2+}$ considerably reduced the size of the fluorescent aggregates to 279.4 nm, which might be due to the better solubility of the **28**–Cu complex. The addition of PPI led to the generation of fluorescent aggregates once again, with a mean size of 506.0 nm, which was close to that of **28** in the same solvent system. The fluorescence response of ADP and ATP was 25% and nearly 100%, respectively, of that of PPI. Negative charges at neutral pH for these NPPs were supposed to be responsible for their coordinating ability with Cu $^{2+}$. An application to the analysis of PPI in a protein removed from 50-fold diluted fetal bovine serum sample was also demonstrated.

On the other hand, a unique combination of displacement assay and aggregation-induced quenching (Fig. 8) was first introduced by Ojida and Wongkongkatap et al. [76] in 2014. The xanthene probe–Ce $^{3+}$ complex **29**–Ce (Table 7) was reported as a selective fluorescence sensor of PPI in methanol aqueous solution. The **29**–Ce complex forms an aggregated polymer through bridging coordination interactions between **29** and Ce $^{3+}$ ions; therefore the large fluorescence decrease of **29** upon complexation with Ce(NO $_3$) $_3$ is ascribed to self-quenching of the xanthene fluorophore in the aggregated state and the quenching effect of the coordinated Ce $^{3+}$ ions. When PPI was added to a solution of the **29**–Ce complex, its fluorescence increased drastically ($I/I_0 = 318$) upon addition of over 20 eq. PPI. ESI–MS showed that **29** mainly exists as the free ligand in the presence of 20 eq. PPI, indicating that coordination exchange occurs between **29**–Ce and PPI to form a Ce $^{3+}$ –PPI complex and to liberate the fluorescent ligand **29**. Other phosphate species, including NPP, and other oxoanions scarcely induced an increase in

fluorescence except for Pi ($I/I_0 = 28$) and uridine-5'-triphosphate ($I/I_0 = 30$). The fluorescence detection of DNA polymerase-catalyzed nucleic acid amplification by the loop-mediated isothermal amplification method for viral infection diagnosis using the **29**-Ce complex was demonstrated (Fig. 10).

3.3 Chemical Reactions

Not only molecular recognition, but the binding interaction between host and guest molecules can be based on an irreversible formation of a covalent bond. Such indicators are described as a chemodosimeter [98]. Chemodosimeters require at least two functional units, namely, the reaction site, where the host binds to the analyte covalently, and the sensing unit, which is dependent on the interaction with the analyte. The analyte reacts with the sensor to create a new molecule with different optical properties (Fig. 11). Because of their high sensitivity and selectivity, the design, synthesis, and application of chemodosimeters in luminescence bioimaging of Pi/PPi have attracted increasing attention and become an active research field.

3.3.1 Pi Sensing

Zhou et al. [99] reported the fluorescence properties of compound **30** containing an oxalate moiety linked via an ester bond to the hydroxyl group of coumarin fluorophore. The selective reaction of **30** toward Pi led to the cleavage of the ester bond and liberation of the fluorophores (Table 8). This unique Pi-induced hydrolytic reaction was effective in DMSO-HEPES buffered solutions that produce a colorimetric change associated with a 62-nm red-shift in the UV-Vis absorption maximum and up to a 780-fold enhancement in the fluorescence intensity; in contrast the addition of uridine monophosphate (UMP) and guanosine monophosphate (GMP) to solutions of **30** resulted in minor fluorescence enhancement (UMP 5.3-fold, GMP 3.6-fold), while common other anions, including NPP, cysteine, glutathione, and glutamic acid did not promote any change in emission. DFT calculations depicted the energy changes occurring in the hydrolytic reaction to be about -19.3 kJ. Mass spectrometry analysis of a mixture of **30** and 100 eq. of Pi in DMSO-HEPES buffer after stirring for 12 h contained a peak at m/z 151.08 corresponding to the cyclic diphosphate (Table 8). Fluorescence imaging studies were carried out using HeLa cells and *Caenorhabditis elegans* with the addition of 4 eq. exogenous Pi, or ATP and apyrase. A clearly detectable bright blue fluorescence was emitted, indicating the increase in endogenous Pi as a result of ATP hydrolysis catalyzed by Apyrase. Another application was performed using Sf9 adherent cells derived from *Spodoptera frugiperda* pupal ovarian tissue and treated with innexin 2 or 3 as an apoptosis inducer. Treatment of Sf9 cells with **30** and innexin 3 in the absence of Apyrase resulted in a clearly detectable fluorescent image, thereby demonstrating that innexin 3 caused dephosphorylation of Akt (protein kinase B) in hemichannel-closed cells that led to apoptosis.

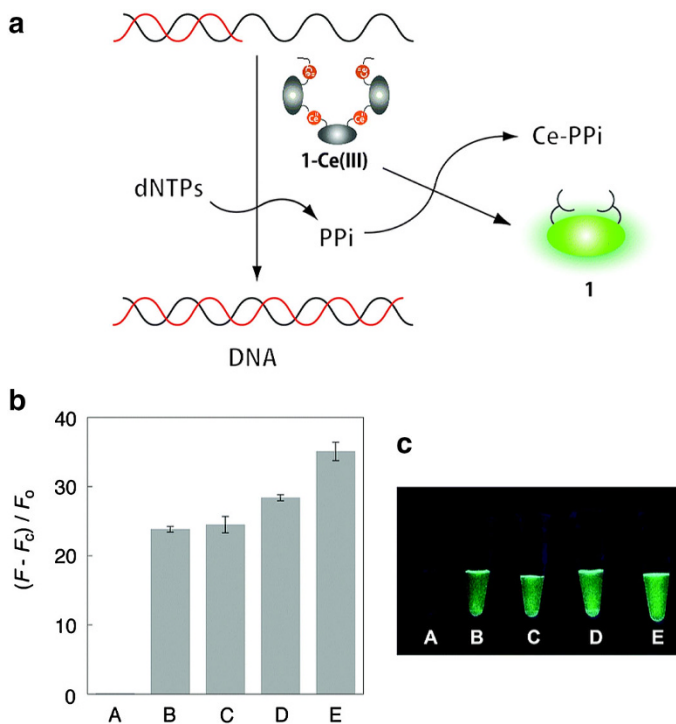


Fig. 10 **a** Mechanism of the fluorescent loop-mediated isothermal amplification (LAMP) assay using the **29**-Ce complex. **b** Changes in fluorescence of the **29**-Ce complex upon addition of the LAMP reaction solution. Each reaction was conducted in the presence of 0 (A), 2 (B), 20 (C), 200 (D), and 1000 (E) copies of white spot syndrome virus plasmid DNA. Conditions [29]: 5×10^{-6} M, 0.27 mM $\text{Ce}(\text{NO}_3)_3$, 25 mM MES (pH 6.8)- CH_3OH (1:1), 25 °C, $\lambda_{\text{ex}} = 500$ nm. The *y*-axis indicates the ratio of the fluorescence increase of each sample $[(F - F_c)/F_0]$, where *F*, *F_c*, and *F₀* are the fluorescence intensity of the reaction sample, of the control reaction sample without plasmid, and of the solution of the **29**- Ce^{3+} complex, respectively. **c** Photographs of solutions of the **29**-Ce complex upon addition of LAMP reaction solutions. Adapted from Kittiloespaisan et al. [76] by permission of The Royal Society of Chemistry

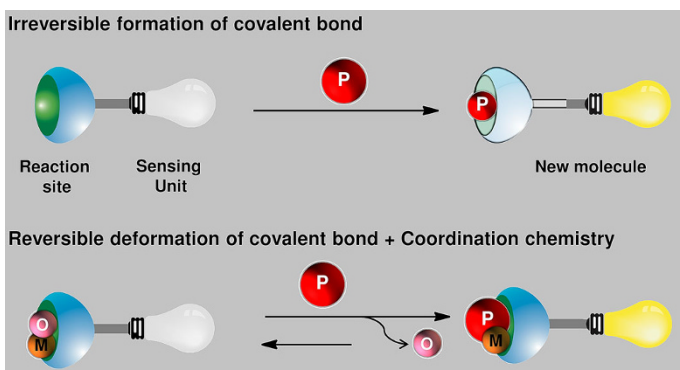
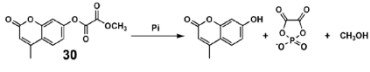

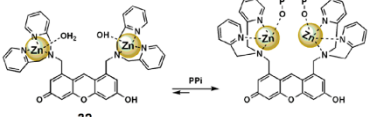


Fig. 11 Schematic overview of the reaction-based sensing mechanism. *M* Metal, *O* covalently bound oxygen, *P* phosphate

Table 8 Sensing through reactions specific to inorganic phosphate/inorganic pyrophosphate

Chemical reactions	Fluorescence sensing properties	Applications	Ref.
 <p>Reaction of compound 30 with Pi to form a fluorescent product, PPi, and CH_3OH.</p>	Emission ON (~ 780-fold at 455 nm) $\lambda_{\text{exc}} = 385 \text{ nm}$ NA (DMSO-0.02 M HEPES, pH 7.4, 9/1 v/v)	Bioimaging using HeLa cells and <i>C. elegans</i> treated with exogenous 4 eq. Pi or 4 eq. of ATP + apyrase Bioimaging of Sf9 adherent cells treated with innexin 2 or 3 with/without apyrase	[99]
 <p>Reaction of compound 31 with PPi and H_2O to form salicylaldehyde and H_2N.</p>	Emission ON (~ 36-fold at 500 nm) $\lambda_{\text{exc}} = 350 \text{ nm}$ MLCT (10 mM Tris, pH 7.4)	Fluorescence imaging of endogenous PPi in mitochondria of HeLa cells incubated with the ANK protein inhibitor probenecid.	[100]
 <p>Reaction of compound 32 with PPi to form a fluorescent product.</p>	Emission ON (~ 55-fold at 523 nm) $\lambda_{\text{exc}} = 488 \text{ nm}$ Binding-induced recovery of the conjugated form of the xanthene fluorophore (50 mM HEPES, 10 mM NaCl, 1 mM MgCl_2 , pH 7.4)	Fluorescent staining of intracellular ATP storage in live Jurkat cells [101] Fluorescent assay of hydrolysis pathway of diadenosine tetraphosphate [102] Organelle-localized multicolor fluorescence probes useful for imaging of NPP dynamics in living cells [103]	[101-106]
		Detection of nucleic acid amplification reaction useful for diagnosis of viral infection [104] Detection of pathogenic spore-forming bacteria through the intracellular ATP pool [105] Visualization of red blood cell CR1-mediated ATP release [106]	

3.3.2 *PPi* Sensing

Zelder et al. [100] reported a metal disassembly strategy concurrent with reaction for imaging endogenous PPi in the mitochondria of HeLa cells by using the square pyramidal Fe^{3+} -salen complex **31** (Table 8). Non-fluorescent **31** displayed an approximately 36-fold increase in emission intensity upon the addition of PPi in aqueous solution (pH 7.4), with a saturation of the intensity at approximately 25 min. Minor increases in emission were observed following the addition of NPP and Pi , but no other anions showed any effects. UV/Vis spectroscopy suggested that the presence of PPi led to a disappearance of the MLCT band of **31**. Job plot analysis indicated a 2:1 stoichiometry for the reaction between **31** and PPi at physiological pH, and an equilibrium constant value ($\log K$) of 7.06 ± 04 was calculated. The reaction-based release of fluorescent salicylaldehyde upon

disassembly of **31** in the presence of PPI was confirmed with ESI MS, and the fluorescence properties of salicylaldehyde were compared with the reaction mixture.

A reversible covalent bond destruction leading to a change in optical properties of the sensing unit upon specific binding to the target molecule is also possible (Fig. 11). This unique sensing mechanism was first noted by Ojida and Hamachi et al. [101] and was used for an excellent design of a Turn-ON fluorescent PPI/NPP sensor. Sensor **32** comprises two DPA-Zn moieties as a binding motif for the PPI and the xanthene fluorophore as a sensing unit (Table 8). In the absence of PPI/NPP, the bridging water between the two Zn^{2+} centers attacked the ring of xanthene fluorophore breaking down a fluorescent π -conjugated ring under a wide pH range of 6–9 in 100% aqueous solution. Upon molecular recognition of PPI/NPP by the two Zn centers, the coordination geometry was modulated as the attacking water was removed to recover the fluorescent xanthene conjugated ring (Fig. 12).

It should be noted that this novel reversible mechanism is a result of the good cooperation between the chemical reaction and coordination chemistry which demonstrates a high potential for several bioanalytical applications, including fluorescent staining of intracellular ATP storage in live Jurkat cells [101], fluorescent assay of the hydrolysis pathway of diadenosine tetraphosphate [102], organelle-localized multicolor fluorescence probes useful for imaging of NPP dynamics in living cells (Fig. 13) [103], detection of nucleic acid amplification reaction useful for viral infection diagnosis [104], detection of pathogenic spore-

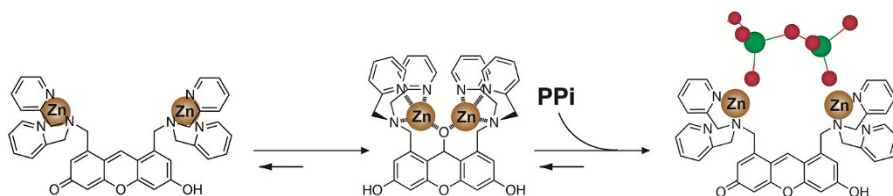


Fig. 12 Schematic illustration of a combination between chemical reaction and coordination chemistry as a reversible destruction of the covalent bond between bridging oxygen and xanthene ring results in a unique sensing mechanism of inorganic pyrophosphate (PPI)/nucleoside polyphosphate achieved by compound **32**

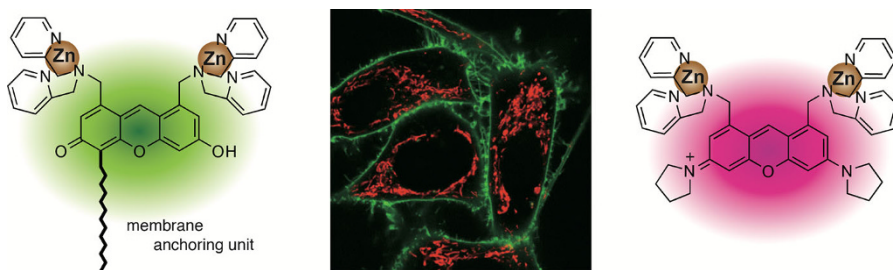


Fig. 13 Confocal fluorescence analysis of HeLa cells stained with *green 32* bearing a biocompatible anchor for membrane unit and *pink 32* with a positively charged pyronin ring. Reprinted with permission from Kurishita et al. [103]. Copyright 2012 American Chemical Society

forming bacteria through the intracellular ATP pool [105], and fluorescent visualization of red blood cell complement receptor 1-mediated ATP release [106].

4 Conclusion

In this review we have summarized how small molecular sensors turn their fluorescence ON or OFF in response to organophosphates, including Pi and PPi as well as potential applications. Hydrogen bonding interaction is quite useful for sensing in organic solvent where the hydration energy is weaker than the self-assemble capability between the sensor and the target molecule. Coordination chemistry provides a great advantage for selective binding of an oxoanion in a complete aqueous solution or the mixture. Several new sensing mechanisms, such as displacement assay, aggregation-induced emission, and chemical reactions enable the high sensitivity and selectivity for Pi/PPi detection. Towards this goal, new molecular design and sensing mechanisms which realize specific sensing of each phosphate species, including Pi/PPi, are urgently required. The development of a new sensing chemistry applicable to practical use is a worthy challenge for many chemists.

Acknowledgements JW is grateful to the Faculty of Science, Mahidol University and the Thailand Research Fund (IRG5980001).

References

1. Lee GJ, Marks J (2015) Intestinal phosphate transport: a therapeutic target in chronic kidney disease and beyond? *Pediatr Nephrol* 30:363–371
2. Hansen NM, Felix R, Bisaz S, Fleisch H (1976) Aggregation of hydroxyapatite crystals. *Biochim Biophys Acta* 451:549–559
3. Crook M, Swaminathan R (1996) Disorders of plasma phosphate and indications for its measurement. *Ann Clin Biochem* 33:376–396
4. Prie D, Beck L, Friedlander G, Silve C (2004) Sodium-phosphate cotransporters, nephrolithiasis and bone demineralization. *Curr Opin Nephrol Hypertens* 13:675–681
5. Knochel JP, Barcenas C, Cotton JR, Fuller TJ, Haller R, Carter NW (1978) Hypophosphatemia and rhabdomyolysis. *J Clin Invest* 62:1240–1246
6. Knochel JP (1977) The pathophysiology and clinical characteristics of severe hypophosphatemia. *Arch Intern Med* 137:203–220
7. Khoshniat S, Bourguin A, Julien M, Weiss P, Guicheux J, Beck L (2011) The emergence of phosphate as a specific signaling molecule in bone and other cell types in mammals. *Cell Mol Life Sci* 68:205–218
8. Rutecki GW, Cugino A, Jarjoura D, Kilner JF, Whittier FC (1997) Nephrologists' subjective attitudes towards end-of-life issues and the conduct of terminal care. *Clin Nephrol* 48:173–180
9. Weisinger JR, Bellorin-Font E (1998) Magnesium and phosphorus. *Lancet* 352:391–396
10. Shiber JR, Mattu A (2002) Serum phosphate abnormalities in the emergency department. *J Emerg Med* 23:395–400
11. Hruska KA, Mathew S, Lund RJ, Memon I, Saab G (2009) The pathogenesis of vascular calcification in the chronic kidney disease mineral bone disorder: the links between bone and the vasculature. *Semin Nephrol* 29:156–165
12. Kanbay M, Goldsmith D, Akcay A, Covic A (2009) Phosphate - the silent stealthy cardiorenal culprit in all stages of chronic kidney disease: a systematic review. *Blood Purif* 27:220–230
13. Hruska KA, Mathew S, Lund R, Qiu P, Pratt R (2008) Hyperphosphatemia of chronic kidney disease. *Kidney Int* 74:148–157

14. Devlin TM (2010) Textbook of biochemistry with clinical correlations, 7th edn. Wiley, New York
15. Heinonen JK (2001) Biological role of inorganic pyrophosphate, 1st edn. Kluwer, Dordrecht
16. Florence WLT (2012) Genetics and mechanisms of crystal deposition in calcium pyrophosphate deposition disease. *Curr Rheumatol Rep* 14:155–160
17. Costello JC, Rosenthal AK, Kurup IV, Masuda I, Medhora M, Ryan LM (2011) Parallel regulation of extracellular ATP and inorganic pyrophosphate: roles of growth factors, transduction modulators, and ANK. *Connect Tissue Res* 52:139–146
18. Rosenthal AK, Gohr CM, Mitton-Fitzgerald E, Lutz MK, Dubyak GR, Ryan LM (2013) The progressive ankylosis gene product ANK regulates extracellular ATP levels in primary articular chondrocytes. *Arthritis Res Ther* 15:R154
19. Rosenthal AK, Ryan LM (2016) Calcium pyrophosphate deposition disease. *N Engl J Med* 374:2575–2584
20. Chen M, Graedel TE (2016) A half-century of global phosphorus flows, stocks, production, consumption, recycling, and environmental impacts. *Glob Environ Change* 36:139–152
21. Mason CF (1991) Biology of freshwater pollution. Longman, New York
22. Hargrove AE, Nieto S, Zhang T, Sessler JL, Anslyn EV (2011) Artificial receptors for the recognition of phosphorylated molecules. *Chem Rev* 111:6603–6782
23. Hirsch AKH, Fischer FR, Diederich F (2007) Phosphate recognition in structural biology. *Angew Chem Int Ed* 46:338–352
24. Kim SK, Lee DH, Hong J-I, Yoon J (2009) Chemosensors for pyrophosphate. *Acc Chem Res* 42:23–31
25. Lee S, Yuen KKY, Jolliffe KA, Yoon J (2015) Fluorescent and colorimetric chemosensors for pyrophosphate. *Chem Soc Rev* 44:1749–1762
26. Busschaert N, Caltagirone C, Rossom WV, Gale PA (2015) Applications of supramolecular anion recognition. *Chem Rev* 115:8038–8155
27. Wu J, Liu W, Ge J, Zhang H, Wang P (2011) New sensing mechanisms for design of fluorescent chemosensors emerging in recent years. *Chem Soc Rev* 40:3483–3495
28. Ngo HT, Liu X, Jolliffe KA (2012) Anion recognition and sensing with Zn(II)-dipicolylamine complexes. *Chem Soc Rev* 41:4928–4965
29. Pak YL, Swamy KMK, Yoon J (2015) Recent progress in fluorescent imaging probes. *Sensors* 15:24374–24396
30. Yoon J, Kim SK, Singh NJ, Kim KS (2006) Imidazolium receptors for the recognition of anions. *Chem Soc Rev* 35:355–360
31. Gale PA (2006) Structural and molecular recognition studies with acyclic anion receptors. *Acc Chem Res* 39:465–475
32. Schmidtchen FP, Berger M (1997) Artificial organic host molecules for anions. *Chem Rev* 97:1609–1646
33. Marcus Y, Rashin A (1994) A simple empirical model describing the thermodynamics of hydration of ions of widely varying charges, sizes, and shapes. *Biophys Chem* 51:111–127
34. Czarnik AW (1994) Chemical communication in water using fluorescent chemosensors. *Acc Chem Res* 27:302–308
35. Chen K-H, Yang J-S, Hwang C-Y, Fang J-M (2008) Phospholipid-induced aggregation and anthracene excimer formation. *Org Lett* 10:4401–4404
36. Lee M, Moon JH, Jun EJ, Kim G, Kwon Y-U, Lee JY, Yoon J (2014) A tetranaphthoimidazolium receptor as a fluorescent chemosensor for phytate. *Chem Commun* 50:5851–5853
37. Zheng F, Guo S, Zeng F, Li J, Wu S (2014) Ratiometric fluorescent probe for alkaline phosphatase based on betaine-modified polyethylenimine via excimer/monomer conversion. *Anal Chem* 86:9873–9879
38. Kumar R, Srivastava A (2016) Anion binding-induced white light emission using a water-tolerant fluorescent molecular tweezer. *Chem Eur J* 22:3224–3229
39. Nishizawa S, Kato Y, Teramae N (1999) Fluorescence sensing of anions via intramolecular excimer formation in a pyrophosphate-induced self-assembly of a pyrene-functionalized guanidinium receptor. *J Am Chem Soc* 121:9463–9464
40. Qian X, Xiao Y, Xu Y, Guo X, Qian J, Zhu W (2010) “Alive” dyes as fluorescent sensors: fluorophore, mechanism, receptor and images in living cells. *Chem Commun* 46:6418–6436
41. Wu FY, Li Z, Guo L, Wang X, Lin MH, Zhao YF, Jiang YB (2006) A unique NH-spacer for *N*-benzamidothiourea based anion sensors. Substituent effect on anion sensing of the ICT dual fluorescent *N*-(*p*-dimethylaminobenzamido)-*N*-arylothioureas. *Org Biomol Chem* 4:624–630
42. Steed JW (2009) Coordination and organometallic compounds as anion receptors and sensors. *Chem Soc Rev* 38:506–519

43. Ojida A, Mito-oka Y, Inoue M, Hamachi I (2002) First artificial receptors and chemosensors toward phosphorylated peptide in aqueous solution. *J Am Chem Soc* 124:6256–6258
44. Ojida A, Mito-oka Y, Sada K, Hamachi I (2004) Molecular recognition and fluorescence sensing of monophosphorylated peptides in aqueous solution by bis(Zn(II)-dipicolylamine)-based artificial receptors. *J Am Chem Soc* 126:2454–2463
45. Elangannan A, Desiraju GR, Klein RA, Sadlej J, Scheiner S, Alkorta I, Clary DC, Crabtree RH, Dannenberg JJ, Hobza P, Kjaergaard HG, Legon AC, Mennucci B, Nesbitt DJ (2011) Definition of the hydrogen bond. *Pure Appl Chem* 83:1637–1641
46. S-i Kondo, Takai R (2013) Selective detection of dihydrogen phosphate anion by fluorescence change with tetraamide-based receptors bearing isoquinolyl and quinolyl moieties. *Org Lett* 15:538–541
47. Zhang D, Jiang X, Yang H, Su Z, Gao E, Martinez A, Gao G (2013) Novel benzimidazolium-urea-based macrocyclic fluorescent sensors: synthesis, ratiometric sensing of H_2PO_4^- and improvement of the anion binding performance via a synergistic binding strategy. *Chem Commun* 49:6149–6151
48. Zhang D, Jiang X, Yang H, Martinez A, Feng M, Donga Z, Gao G (2013) Acridine-based macrocyclic fluorescent sensors: self-assembly behavior characterized by crystal structures and a tunable bathochromic-shift in emission induced by H_2PO_4^- via adjusting the ring size and rigidity. *Org Biomol Chem* 11:3375–3381
49. Martí-Centelles V, Burguete MI, Galindo F, Izquierdo MA, Kumar DK, White AJP, Luis SV, Vilar R (2012) Fluorescent acridine-based receptors for H_2PO_4^- . *J Org Chem* 77:490–500
50. Gong W, Zhang Q, Wang F, Gao B, Lin Y, Ning G (2012) Selective sensing of H_2PO_4^- (Pi) driven by the assembly of anthryl pyridinium ligands. *Org Biomol Chem* 10:7578–7583
51. Caltagirone C, Bazzicalupi C, Isaia F, Light ME, Lippolis V, Montis R, Murgia S, Olivari M, Picci G (2013) A new family for bis-ureidic receptors for pyrophosphate optical sensing. *Org Biomol Chem* 11:2445–2451
52. Casula A, Bazzicalupi C, Bettoschi A, Cadoni E, Coles SJ, Horton PN, Isaia F, Lippolis V, Mapp LK, Marini GM, Montis R, Scorciapino MA, Caltagirone C (2016) Fluorescent asymmetric bis-ureas for pyrophosphate recognition in pure water. *Dalton Trans* 45:3078–3085
53. Sanchez G, Espinosa A, Curiel D, Tarraga A, Molina P (2013) Bis(carbazolyl)ureas as selective receptors for the recognition of hydrogen pyrophosphate in aqueous media. *J Org Chem* 78:9725–9737
54. Yuan Y, Gao G, Jiang ZL, You JS, Zhou ZY, Yuan DQ, Xie RG (2002) Synthesis and selective anion recognition of imidazolium cyclophanes. *Tetrahedron* 58:8993–8999
55. Bhalla V, Vij V, Kumar M, Sharma PR, Kaur T (2012) Recognition of adenosine monophosphate and H_2PO_4^- using zinc ensemble of new hexaphenylbenzene derivative: potential bioprobe and multichannel keypad system. *Org Lett* 14:1012–1015
56. Ni XL, Zeng X, Redshaw C, Yamato T (2011) Ratiometric fluorescent receptors for both Zn^{2+} and H_2PO_4^- ions based on a pyrenyl-linked triazole-modified homooxacalix[3]arene: a potential molecular traffic signal with an R-S latch logic circuit. *J Org Chem* 76:5696–5702
57. Lee HN, Swamy KMK, Kim SK, Kwon J-Y, Kim Y, Kim S-J, Yoon YJ, Yoon J (2007) Simple but effective way to sense pyrophosphate and inorganic Phosphate by fluorescence changes. *Org Lett* 9:243–246
58. Nadella S, Selvakumar PM, Suresh E, Subramanian PS, Albrecht M, Giese M, Fröhlich R (2012) Lanthanide(III) complexes of bis-semicarbazone and bis-imine-substituted phenanthroline ligands: solid-state structures, photophysical properties, and anion sensing. *Chem Eur J* 18:16784–16792
59. Ganjali MR, Hosseini M, Memari Z, Faridbod F, Norouzi P, Goldoos H, Badiei A (2011) Selective recognition of monohydrogen phosphate by fluorescence enhancement of a new cerium complex. *Anal Chim Acta* 708:107–110
60. Nadella S, Sahoo J, Subramanian PS, Sahu A, Mishra S, Albrecht M (2014) Sensing of phosphates by using luminescent Eu^{III} and Tb^{III} complexes: application to the microalgal cell *Chlorella vulgaris*. *Chem Eur J* 20:6047–6053
61. Mahapatra AK, Ali SS, Maiti K, Manna SK, Maji R, Mondal S, Uddin MdR, Mandal S, Sahoo P (2015) Aminomethylpyrene-based imino-phenols as primary fluorescence switch-on sensors for Al^{3+} in solution and in Vero cells and their complexes as secondary recognition ensembles toward pyrophosphate. *RSC Adv* 5:81203–81211
62. Kim HJ, Lee JH, Hong J-I (2011) Highly sensitive chemosensor for detection of PPI with improved detection limit. *Tetrahedron Lett* 52:4944–4946
63. Kimura E, Shiota T, Koike T, Shiro M, Kodama M (1990) A zinc(II) complex of 1,5,9-triazacyclododecane ([12]ane N_3) as a model for carbonic anhydrase. *J Am Chem Soc* 112:5805–5811

64. Kimura E, Aoki S, Koike T, Shiro M (1997) A tris(Zn^{II} – 1,4,7,10-tetraazacyclododecane) complex as a new receptor for phosphate dianions in aqueous solution. *J Am Chem Soc* 119:3068–3076
65. Bobyr E, Lassila JK, Wiersma-Koch HI, Fenn TD, Lee JJ, Nikolic-Hughes I, Hodgson KO, Rees DC, Hedman B, Herschlag D (2012) High-resolution analysis of Zn^{2+} coordination in the alkaline phosphatase superfamily by EXAFS and X-ray crystallography. *J Mol Biol* 415:102–117
66. Kim JS, Quang DT (2007) Calixarene-derived fluorescent probes. *Chem Rev* 107:3780–3799
67. Mizukami S, Nagano T, Urano Y, Odani A, Kikuchi K (2002) A fluorescent anion sensor that works in neutral aqueous solution for bioanalytical application. *J Am Chem Soc* 124:3920–3925
68. Yang C, Fu L-M, Wang Y, Zhang J-P, Wong WT, Ai X-C, Qiao YF, Zou BS, Gui L-L (2004) A highly luminescent europium complex showing visible-light-sensitized red emission: direct observation of the singlet pathway. *Angew Chem Int Ed* 43:5010–5013
69. Zhang JF, Kim S, Han JH, Lee S-J, Pradhan T, Cao QY, Lee SJ, Kang C, Kim JS (2011) Pyrophosphate-selective fluorescent chemosensor based on 1, 8-naphthalimide-DPA-Zn(II) complex and its application for cell imaging. *Org Lett* 13:5294–5297
70. Lin J-R, Chu C-J, Venkatesan P, Wu S-P (2015) Zinc(II) and pyrophosphate selective fluorescence probe and its application to living cell imaging. *Sens Actuat B* 207:563–570
71. Jiao S-Y, Li K, Xin Wang X, Huang Z, Pu L, Yu X-Q (2015) Making pyrophosphate visible: the first precipitable and real-time fluorescent sensor for pyrophosphate in aqueous solution. *Analyst* 140:174–181
72. Ojida A, Miyahara Y, Wongkongkatep J, S-i Tamaru, Sada K, Hamachi I (2006) Design of dual-emission chemosensors for ratiometric detection of ATP derivatives. *Chem Asian J* 1:555–563
73. Bhuyan M, Katayev E, Stadlbauer S, Nonaka H, Ojida A, Hamachi I, König B (2011) Rigid luminescent bis-Zinc(II)-bis-cyclen complexes for the detection of phosphate anions and non-covalent protein labeling in aqueous solution. *Eur J Org Chem* 2011 (15):2807–2817
74. Mesquita LM, André V, Esteves CV, Palmeira T, Berberan-Santos MN, Mateus P, Delgado R (2016) Dinuclear Zinc(II) macrocyclic complex as receptor for selective fluorescence sensing of pyrophosphate. *Inorg Chem* 55:2212–2219
75. Bhowmik S, Ghosh BN, Marjomäki V, Rissanen K (2014) Nanomolar pyrophosphate detection in water and in a self-assembled hydrogel of a simple terpyridine- Zn^{2+} complex. *J Am Chem Soc* 136:5543–5546
76. Kittiloespaisan E, Takashima I, Kiatpathomchai W, Wongkongkatep J, Ojida A (2014) Coordination ligand exchange of a xanthene probe-Ce(III) complex for selective fluorescence sensing of inorganic pyrophosphate. *Chem Commun* 50:2126–2128
77. Svane S, Kjeldsen F, McKee V, McKenzie CJ (2015) The selectivity of water-based pyrophosphate recognition is tuned by metal substitution in dimetallic receptors. *Dalton Trans* 44:11877–11886
78. Nguyen BT, Anslyn EV (2006) Indicator-displacement assays. *Coordin Chem Rev* 250:3118–3127
79. You L, Zha D, Anslyn EV (2015) Recent advances in supramolecular analytical chemistry using optical sensing. *Chem Rev* 115:7840–7892
80. Meng Q, Wang Y, Yang M, Zhang R, Wang R, Zhang Z (2015) A new fluorescent chemosensor for highly selective and sensitive detection of inorganic phosphate (Pi) in aqueous solution and living cells. *RSC Adv* 5:53189–53197
81. Wu J, Gao Y, Lu J, Hu J, Ju Y (2015) A steroid-coumarin conjugate for cascade recognition of copper ion and dihydrogen phosphate: microstructural features and IMPLICATION logic gate properties. *Sens Actuat B* 206:516–523
82. Wu J, Zhao X, Gao Y, Hu J, Ju Y (2015) A steroid-salen conjugate for zinc ion recognition and its applications in test-strips, living cells imaging, and cascade recognition for dihydrogen phosphate. *Sens Actuat B* 221:334–340
83. Jiao S-Y, Li K, Zhang W, Liu Y-H, Huang Z, Yu X-Q (2015) Cd(II)-terpyridine-based complex as a ratiometric fluorescent probe for pyrophosphate detection in solution and as an imaging agent in living cells. *Dalton Trans* 44:1358–1365
84. Zhu W, Huang X, Guo Z, Wu X, Yu H, Tian H (2012) A novel NIR fluorescent turn-on sensor for the detection of pyrophosphate anion in complete water system. *Chem Commun* 48:1784–1786
85. Que EL, Chang CJ (2006) A smart magnetic resonance contrast agent for selective copper sensing. *J Am Chem Soc* 128:15942–15943
86. Datta BK, Mukherjee S, Kar C, Ramesh A, Das G (2013) Zn^{2+} and pyrophosphate sensing: selective detection in physiological conditions and application in DNA-based estimation of bacterial cell numbers. *Anal Chem* 85:8369–8375

87. Qiang J, Chang C, Zhu Z, Wei T, Yu W, Wang F, Yin J, Wang Y, Zhang W, Xie J, Chen X (2016) A dinuclear-copper(II) complex-based sensor for pyrophosphate and its applications to detecting pyrophosphatase activity and monitoring polymerase chain reaction. *Sens Actuat B* 233:591–598
88. Butler SJ, Jolliffe KA (2012) Selective pyrophosphate recognition by cyclic peptide receptors in physiological saline. *Chem Asian J* 7:2621–2628
89. Hanshaw RG, Hilkert SM, Jiang H, Smith BD (2004) An indicator displacement system for fluorescent detection of phosphate oxyanions under physiological conditions. *Tetrahedron Lett* 45:8721–8724
90. McDonough MJ, Reynolds AJ, Lee WYG, Jolliffe KA (2006) Selective recognition of pyrophosphate in water using a backbone modified cyclic peptide receptor. *Chem Commun* 2006: 2971–2973. doi: [10.1039/B606917G](https://doi.org/10.1039/B606917G)
91. Luo J, Xie Z, Lam JWY, Cheng L, Chen H, Qiu C, Kwok HS, Zhan X, Liu Y, Zhu D, Tang BZ (2001) Aggregation-induced emission of 1-methyl-1,2,3,4,5-pentaphenylsilole. *Chem Commun* 2001:1740–1741. doi: [10.1039/B105159H](https://doi.org/10.1039/B105159H)
92. Chao D, Ni S (2016) Nanomolar pyrophosphate detection and nucleus staining in living cells with simple terpyridine–Zn(II) complexes. *Sci Rep* 6:26477
93. Wang J-H, Xiong J-B, Zhang X, Song S, Zhu Z-H, Zheng Y-S (2015) Tetraphenylethylene imidazolium macrocycle: synthesis and selective fluorescence turn-on sensing of pyrophosphate anions. *RSC Adv* 5:60096–60100
94. Yan X, Wang M, Cook TR, Zhang M, Saha ML, Zhou Z, Li X, Huang F, Stang PJ (2016) Light-emitting superstructures with anion effect: coordination-driven self-assembly of pure tetraphenylethylene metallacycles and metallacages. *J Am Chem Soc* 138:4580–4588
95. Liu H, Wei R, Xiang Y, Tong A (2015) Fluorescence turn-on detection of pyrophosphate based on aggregation-induced emission property of 5-chlorosalicylaldehyde azine. *Anal Methods* 7:753–758
96. Tang W, Xiang Y, Tong A (2009) Salicylaldehyde azines as fluorophores of aggregation-induced emission enhancement characteristics. *J Org Chem* 74:2163–2166
97. Chen X, Yamaguchi A, Namekawa M, Kamijo T, Teramae N, Tong A (2011) Functionalization of mesoporous silica membrane with a Schiff base fluorophore for Cu(II) ion sensing. *Anal Chim Acta* 696:94–100
98. Yang Y, Zhao Q, Feng W, Li F (2013) Luminescent chemodosimeters for bioimaging. *Chem Rev* 113:192–270
99. Guo LE, Zhang JF, Liu XY, Zhang LM, Zhang HL, Chen JH, Xie XG, Zhou Y, Luo K, Yoon J (2015) Phosphate ion targeted colorimetric and fluorescent probe and its use to monitor endogenous phosphate ion in a hemichannel-closed cell. *Anal Chem* 87:1196–1201
100. Kumari N, Huang H, Chao H, Gasser G, Zelder F (2016) A disassembly strategy for imaging endogenous pyrophosphate in mitochondria by using an Fe^{III}–salen complex. *ChemBioChem* 17:1211–1215
101. Ojida A, Takashima I, Kohira T, Nonaka H, Hamachi I (2008) Turn-on fluorescence sensing of nucleoside polyphosphates using a xanthene-based Zn(II) complex chemosensor. *J Am Chem Soc* 130:12095–12101
102. Kohira T, Takashima I, Nonaka H, Ojida A, Hamachi I (2008) Real-time off/on-mode fluorescence assay for enzyme reactions involving nucleoside polyphosphates by use of a xanthene Zn^{II}-Dpa chemosensor. *Chem Lett* 37:1164–1165
103. Kurishita Y, Kohira T, Ojida A, Hamachi I (2012) Organelle-localizable fluorescent chemosensors for site-specific multicolor imaging of nucleoside polyphosphate dynamics in living cells. *J Am Chem Soc* 134:18779–18789
104. Kittiloepsaisan E, Ojida A, Hamachi I, Seetang-Nun Y, Kiatpathomchai W, Wongkongkatep J (2012) Label-free fluorescent detection of loop-mediated isothermal amplification of nucleic acid using pyrophosphate-selective xanthene-based Zn(II)-coordination chemosensor. *Chem Lett* 41:1666–1668
105. Tiposoth P, Khamsakhon S, Ketsub N, Pongtharangkul T, Takashima I, Ojida A, Hamachi I, Wongkongkatep J (2015) Rapid and quantitative fluorescence detection of pathogenic spore-forming bacteria using a xanthene-Zn(II) complex chemosensor. *Sens Actuat B* 209:606–612
106. Melhorn MI, Brodsky AS, Estanislau J, Khoory JA, Illigens B, Hamachi I, Kurishita Y, Fraser AD, Nicholson-Weller A, Dolmatova E, Duffy HS, Ghiran IC (2013) CR1-mediated ATP release by human red blood cells promotes CR1 clustering and modulates the immune transfer process. *J Biol Chem* 288:31139–31153

Metal Fluorides: Tools for Structural and Computational Analysis of Phosphoryl Transfer Enzymes

Yi Jin^{1,2} · Robert W. Molt Jr.^{3,4,5} ·
G. Michael Blackburn²

Received: 2 December 2016 / Accepted: 1 March 2017 / Published online: 15 March 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract The phosphoryl group, PO_3^- , is the dynamic structural unit in the biological chemistry of phosphorus. Its transfer from a donor to an acceptor atom, with oxygen much more prevalent than nitrogen, carbon, or sulfur, is at the core of a great majority of enzyme-catalyzed reactions involving phosphate esters, anhydrides, amidates, and phosphorothioates. The serendipitous discovery that the phosphoryl group could be labeled by “nuclear mutation,” by substitution of PO_3^- by MgF_3^- or AlF_4^- , has underpinned the application of metal fluoride (MF_x) complexes to mimic transition states for enzymatic phosphoryl transfer reactions, with sufficient stability for experimental analysis. Protein crystallography in the solid state and ^{19}F NMR in solution have enabled direct observation of ternary and quaternary protein complexes embracing MF_x transition state models with precision. These studies have underpinned a radically new mechanistic approach to enzyme catalysis for a huge range of phosphoryl transfer processes, as varied as kinases, phosphatases, phosphomutases, and phosphohydrolases. The results, without

This article is part of the Topical Collection “Phosphate Labeling in Chemical Biology”; edited by Henning Jessen.

✉ G. Michael Blackburn
g.m.blackburn@sheffield.ac.uk

- ¹ Structural Biology Laboratory, Department of Chemistry, University of York, York YO31 7YD, UK
- ² Department of Molecular Biology and Biotechnology, Krebs Institute, University of Sheffield, Sheffield S10 2TN, UK
- ³ ENSCO, Inc., 4849 North Wickham Road, Melbourne, FL 32940, USA
- ⁴ Department of Chemistry and Chemical Biology, Indiana University-Purdue University, Indianapolis, IN 46202, USA
- ⁵ Department of Biochemistry and Molecular Biology, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

exception, have endorsed trigonal bipyramidal geometry (tbp) for concerted, “in-line” stereochemistry of phosphoryl transfer. QM computations have established the validity of tbp MF_x complexes as reliable models for true transition states, delivering similar bond lengths, coordination to essential metal ions, and virtually identical hydrogen bond networks. The emergence of protein control of reactant orbital overlap between bond-forming species within enzyme transition states is a new challenging theme for wider exploration.

Keywords MF_x · Phosphoryl group surrogates · Enzyme mechanisms · Transition state analogs · QM/MM computation · KS-DFT analysis

1 Background

Alexander Todd¹ and Frank Westheimer² held complementary, and sometime overlapping, views on the centrality of phosphates for life. Todd’s pronouncement: “Where there’s Life, there’s Phosphorus”, encapsulated his conviction that enzymes that manipulate phosphates have been at the heart of biology from the dawn of life anywhere in the universe [1]. Westheimer identified the evolutionary centrality of phosphate [2]. The cellular behavior of phosphate esters and anhydrides provides one of the most remarkable chemical paradoxes: phosphate monoesters hydrolyze spontaneously under physiological conditions with $t_{1/2}$ 10^{12} years, yet simple phosphatase enzymes have k_{cat} ca. 30 s^{-1} . The enormous difference corresponds to a remarkable catalytic rate enhancement of 10^{21} [3]. How do enzymes achieve this? This article focuses on the use of aluminum and magnesium fluoride complexes to mimic structures of transition states of enzymatic reactions that involve the phosphoryl group, PO_3^- , and to provide a structural base for quantum chemical computations to describe them in detail.

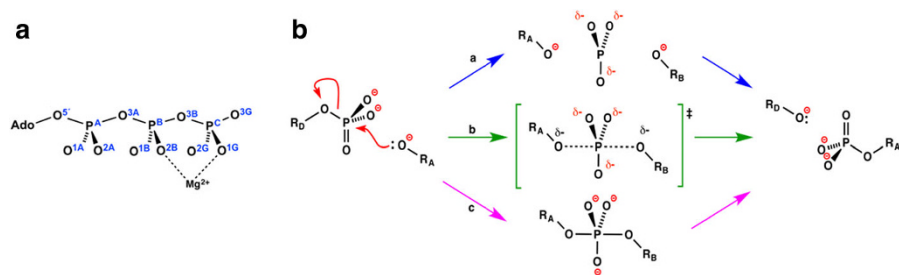
1.1 Basics of Phosphoryl Transfer

Studies on phosphoryl transfer reactions were greatly advanced by the use of oxygen isotopes to show that they generally involved P–O cleavage, with transfer of the phosphoryl group (PO_3^-) between a donor oxygen (O_D) and an acceptor oxygen (O_A) or, less commonly, nitrogen or sulfur atoms [4]. Polyphosphates, such as ATP, react by attack of water (or an alcohol) on the terminal γ -phosphorus, breaking the P–O bond to the $\text{O}^{3\text{B}}$ atom (usually oxygen and infrequently nitrogen) (Scheme 1a). More advanced isotope work, deploying ^{16}O , ^{17}O , and ^{18}O , established that the near-universal stereochemistry for such processes, for both chemical and enzymatic reactions, involves inversion of stereochemistry at the transferring phosphorus (Scheme 1) [5–7].

An accurate description of the technical aspects of the varieties and uses of MF_x models calls for a brief explanation of the terminology of phosphoryl transfer. The

¹ Lord Todd of Trumpington, Nobel Laureate 1957, 1702 Professor of Chemistry 1944–71, Cambridge, UK.

² Professor of Chemistry, 1953–2007, Harvard, Cambridge, USA.

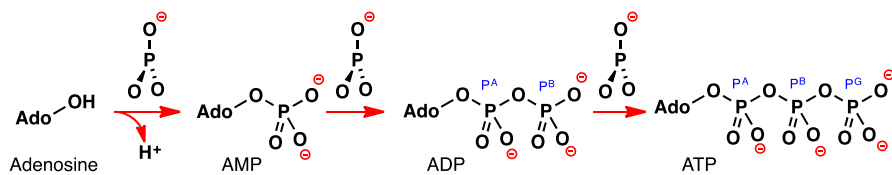


Scheme 1 **a** The atomic identities in ATP employ the new IUPAC nomenclature [8]. **b** Three mechanisms for transfer of a phosphoryl group (PO_3^- , *top center*) between a donor (*left*) and an acceptor (*right*) species. Sequential process (**a**, *blue arrows*) involves formation of a trigonal planar metaphosphate anion as an intermediate. Concerted process (**b**, *green arrows*) shows a trigonal bipyramidal (tbp) transition state with phosphorus fully bonded to three equatorial oxygens and partially bonded to the axial donor (O_D) and acceptor (O_A) oxygens. An alternative, sequential process (**c**, *magenta arrows*), largely discarded, shows formation of a stable pentacoordinate phosphorane intermediate having full bonds to all five oxygens

phosphoryl group, PO_3^- is identified throughout organic chemistry and biology as the anionic, trigonal planar assembly of a phosphorus and three oxygen atoms. It is usually drawn without $\text{P}=\text{O}$ double bonds, is highly electrophilic, and has not been identified in any condensed phase (Scheme 2). It is helpful to perceive its combination with an alcohol, such as adenosyl-5'-OH, to generate a phosphate monoester, illustrated for adenosine 5'-phosphate, AMP. The addition of a second phosphoryl group to a terminal oxygen generates a pyrophosphate monoester, illustrated for adenosine 5'-diphosphate, ADP; and capture of a third phosphoryl group gives adenosine 5'-triphosphate, ATP (Scheme 2). Strings of phosphorus atoms in such chains have conventionally been labeled P^α , P^β , P^γ , etc. but, with new IUPAC nomenclature, are now better identified as P^A , P^B , P^C , P^D , etc [8].

1.2 Historic Development of Mechanisms

Todd [9] and Westheimer [2] thought that phosphoryl transfer reactions should be stepwise, involving a monomeric metaphosphate intermediate species (Scheme 1a). That concept, unproven after extended but fruitless effort, has now been discarded in favor of concerted phosphoryl transfer reactions for phosphate monoesters and anhydrides. These have “in-line” geometry for $\text{O}_D\text{-P-O}_A$ in the transition state (TS), with variable associative or dissociative character (Scheme 1b) [10–12]. Isotope labeling studies have contributed historically to studies on phosphoryl transfer in biological systems [13] and ^{31}P NMR has been applied effectively for investigations on ATP [14] and phosphoarginine [15], but protein crystallography before the mid-1990s was restricted to binary complexes with stable substrate and bisubstrate analogs that gave limited information about reaction mechanisms [16, 17]. In 1994, that situation changed dramatically with the crystallization of ternary complexes of guanosine diphosphate (GDP) coordinated to tetrafluoroaluminate, AlF_4^- , and the small G protein, $\text{G}_{i\alpha 1}$ (PDB: **1gfi**) [18] and with transducin α (PDB: **1tad**) [19]. Although these complexes had octahedral geometry for the



Scheme 2 Successive capture of three phosphoryl groups, PO_3^- , converts an alcohol first into a phosphate monoester, then into an alkyl diphosphate, and finally into alkyl triphosphate, illustrated for adenosine 5'-triphosphate

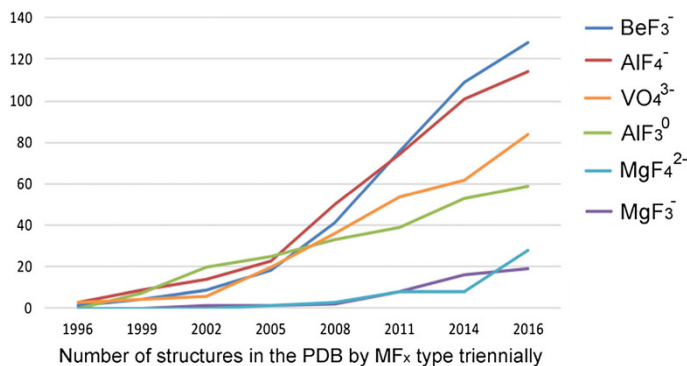


Fig. 1 MF_x structures deposited in the Protein Data Bank since their inception in 1996. Vanadate (orange) is included for comparison

aluminum tetrafluoride moiety, they were immediately described as transition state analogs (TSAs) for phosphoryl transfer. They were soon followed by further MF_x species, notably BeF_3^- , AlF_3^0 , and MgF_3^- . The number of such complexes has grown steadily, now exceeding 350 (Fig. 1).

2 Development of Metal Fluorides as Phosphate Analogs

2.1 The ‘Burst Phase’ of Analog Discovery

Exciting developments in the field of signal switch mechanisms based on hydrolysis of GTP by small G proteins and on the molecular biology discoveries depending on ATP hydrolysis stimulated the development of three analog systems in the mid-1990s. Using tetrafluoroaluminate, work on $\text{G}_{i\alpha 1}$, from the University of Texas, Dallas [18], narrowly edged out a closely related publication on Transducin α from Yale [19], while both focused on the role of the essential glutamine and arginine residues, of the catalytic magnesium, and on the positioning of water for attack on the terminal phosphate, P^{G} (Fig. 2a). Shortly after, a complex of ADP with AlF_4^- was described to represent the TS for ATP hydrolysis in myosin, alongside the first structure of a trifluoroberyllate (BeF_3^-) complex with ADP, recognized as a ground state analog (GSA; Fig. 2b) [20].

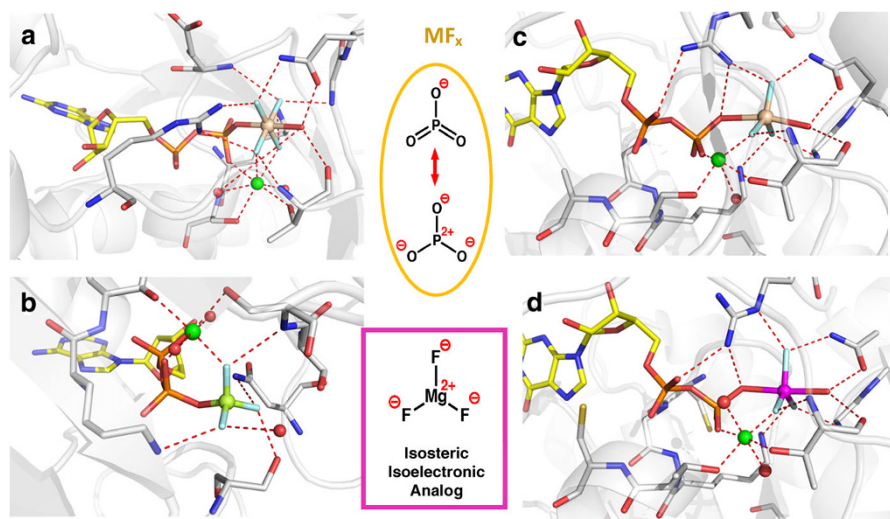


Fig. 2 Landmark protein structures with a fourfold range of MF_x octahedral tetrafluoroaluminate, *tbp* aluminum trifluoride and trifluoromagnesate, and tetrahedral trifluoroberyllate complexes. **a** Transducin α with GDP coordinated to AlF_4^- and catalytic magnesium (PDB: **1tad** at 1.7 Å). **b** Myosin with ADP coordinated to tetrahedral BeF_3^- and Mg_{cat}^{2+} (PDB: **1mmd** at 2.0 Å). **c** Ras-RasGAP with GDP coordinated to *tbp* AlF_3^0 and Mg_{cat}^{2+} (PDB: **1wq1** at 2.5 Å). **d** RhoA-RhoGAP with GDP coordinated to *tbp* MgF_3^- and Mg_{cat}^{2+} (PDB: **1ow3** at 1.7 Å) (colors: aluminum, grey; beryllium, lime; surrogate magnesium, magenta; Mg_{cat}^{2+} , green; fluorine, light blue; nucleotides, yellow; key amino acids, silver; nitrogen, blue; oxygen, red)

After a short interval, a third class of MF_x analog was reported: an aluminum trifluoride complex of magnesium ADP for a dinucleotide kinase, described alongside the corresponding trifluoroberyllate tetrahedral complex. Its great advantage was *tbp* geometry for the TSA complex that, for the first time, accurately mimicked the TS geometry of the γ -phosphate of ATP undergoing transfer [21]. This was quickly followed by a $GDP \cdot AlF_3^0$ complex for the small G protein Ras-RasGAP (Fig. 2c) [22] and then by an $ADP \cdot AlF_3^0 \cdot GDP$ complex for a quaternary complex of a nucleoside diphosphate kinase from the slime mold, *Dictyostelium discoideum* [23]. These, and subsequent examples of *tbp* complexes, recognized that AlF_3^0 was a neutral MF_x species and therefore a Coulombic mismatch for an *anionic* phosphoryl group. It was 5 years before that feature was rectified with the first identification of trifluoromagnesate (MgF_3^-) bound to GDP in a complex with the small G protein, RhoA. A key component of that work was the rigorous use of proton-induced X-ray emission spectroscopy (PIXE) to identify magnesium as the atom at the core of the *tbp* complex (Fig. 2d) [24].

By this time, there were some 50 structures deposited in the PDB for MF_x complexes, usually with anionic oxygen as one axial ligand. Their importance has stimulated a rapid, ongoing growth in their use (Fig. 1). We shall now examine the relative qualities of these four classes and their offshoots on a systematic basis, organized by geometric considerations.

3 MF_x Ground State Analogs

3.1 BeF₃⁻ as a Ground State Phosphate Mimic

In aqueous solution, beryllium (II) forms stable fluorides as a mixture of tetrahedral species including BeF₂·2H₂O, BeF₃⁻·H₂O, and BeF₄⁼ [25]. ¹⁹F NMR studies on fluoroberyllate complexes with ADP identified mixed fluoroberyllate·ADP species for myosin (Fig. 2b). Nearly 130 trifluoroberyllate complexes have now been described, with three structures solved by NMR and 119 X-ray structures having resolutions of ≥1.2 Å, generally having tetrahedral trifluoroberyllate bonded to an anionic oxygen. These comprise two sub-groups: over 70 have Be coordinated to an aspartate carboxylate while some 50 have Be coordinated to a terminal phosphate oxygen of a nucleotide. Just two have Be coordinated to a histidine ring nitrogen, while one has BeF₂ bridging two phosphates.

3.1.1 Aspartyl Trifluoroberyllates

Aspartyl phosphates are intermediates in many enzyme reactions, with a half-life for spontaneous hydrolysis from 23 s to a few hours [26]. Aspartyl trifluoroberyllates are stable and available for analysis by ¹⁹F NMR and protein crystallography. They are tetrahedral, ground state mimics of an aspartyl phosphate. The 70 structures have a common core, with bidentate coordination to a divalent metal ion, generally Mg²⁺ and rarely Mn²⁺, from fluorine F₁ and the second carboxylate oxygen to give a near-planar six-membered ring (Fig. 3a; all MF_x structural data are tabulated in a recent review [27]). Because of its low electron density, the beryllium atom is difficult to locate by X-ray diffraction, resulting in uncertainty in its exact position, leading to considerable variation in attributed geometry (Fig. 3b): the 27 best resolved structures have a Be–O distance 1.72 Å with Be–F 1.53 Å.

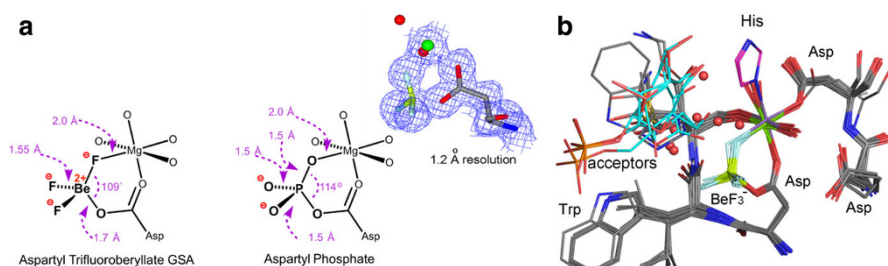


Fig. 3 **a** Typical aspartyl trifluoroberyllate structure with catalytic magnesium coordination (*left*). Aspartyl phosphate complex with catalytic magnesium from phosphoserine phosphatase (PSP) (PDB: **1j97**) for comparison of geometry (*center*). Electron density map for the 1.2-Å resolution structure for β -phosphoglucomutase (PDB: **2wf8**) (*center*). **b** Twenty aligned aspartyl-trifluoroberyllate structures with BeF₃⁻ locked in a six-membered ring. Catalytic Mg²⁺ and an Asp carboxylate fuse a 13-atom ring to the fluoroberyllate ring (*rear*). Octahedral coordination to Mg is completed by an additional aspartate (*right*), by 1–2 waters, and only twice by a histidine (*top, magenta*) (atom colors: fluorine, *light blue*; beryllium, *lime*; nitrogen, *blue*; oxygen, *red*; carbon, *grey*; 3-phosphoglycerate, *cyan*) (electron densities presented in CCP4MG from mtz data in EDS and contoured at 1 σ) (**a** adapted by the authors from [27])

3.1.2 BeF_3^- Nucleotide Structures

There are 42 X-ray structures of BeF_3^- complexes with ADP, and six with GDP. They are isosteric mimics of ATP and GTP (Fig. 4a) in kinases, F1 ATPase, hydrolases, mutases, helicases, and small G proteins. Twenty structures align very well (Fig. 4b) with Be bonded to a β -phosphate oxygen, while a catalytic Mg^{2+} is coordinated to F_1 and to another β -phosphate oxygen.

3.1.3 Histidine Trifluoroberyllates

Various approaches to analogs of η -phosphohistidine have been explored. Structural work on nicotinamide phosphoribosyltransferase (NAMPT) has mimicked phosphorylation of an active-site histidine using trifluoroberyllate. Crystal structures of NAMPT for reactant and product complexes (PDB: **3dhf**; Fig. 5b) have a covalent $\text{His247}\cdot\text{BeF}_3^-$, and in contrast to all other trifluoroberyllate structures, magnesium is coordinated to one fluorine without any direct linkage to His247 [28].

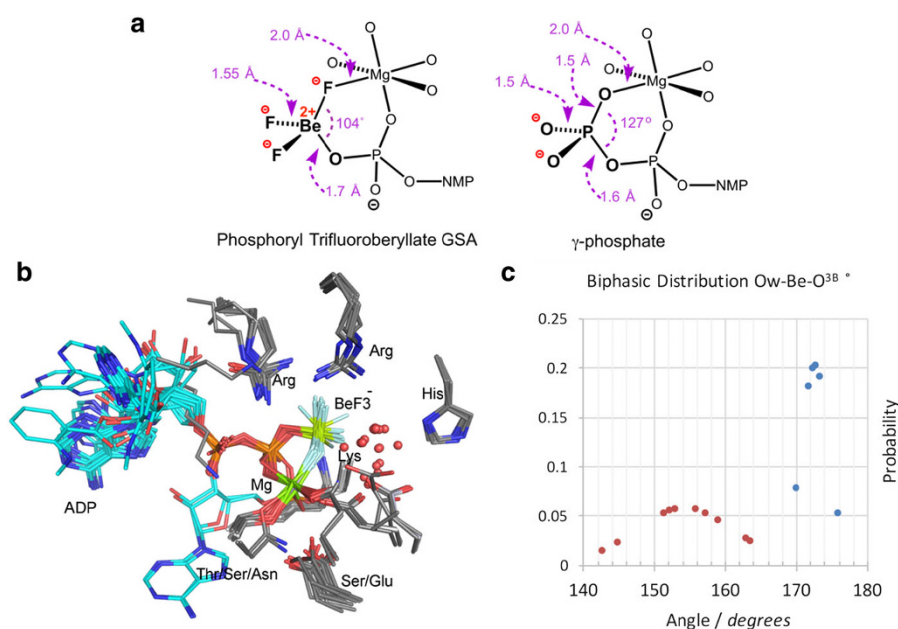


Fig. 4 **a** Typical nucleoside diphosphate trifluoroberyllate structure (left) with catalytic magnesium coordination for comparison of geometry with the nucleoside triphosphate (right). **b** The BeF_3^- moiety in 20 aligned ADP-trifluoroberyllate structures is in a six-membered ring (center) with Mg^{2+} coordinating F_1 and $\text{O}^{3\text{B}}$. γ -Phosphate coordination to an Arg and a Lys is also common. **c** Biphasic normal distribution of the location of the nucleophilic water, Ow, relative to the bond from $\text{ADP-O}^{3\text{B}}$ to beryllium in 16 ADP- BeF_3^- ground state complexes. Major group Ow-Be-O^{3B} angle $\geq 165^\circ$ (orange); minor group Ow-Be-O^{3B} angle $176^\circ \geq 170^\circ$ (blue)

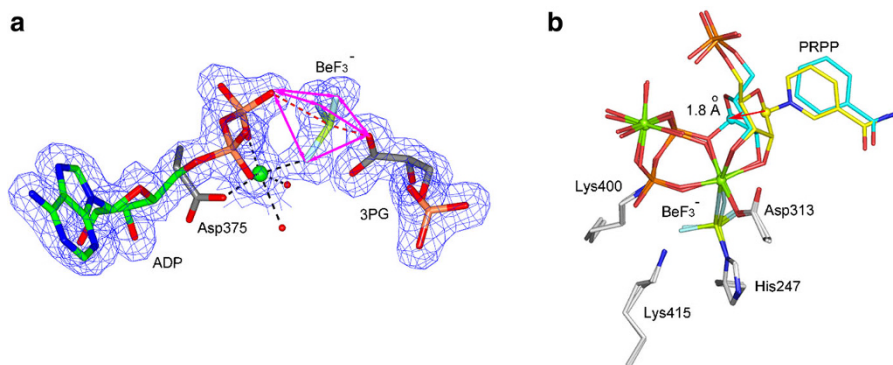


Fig. 5 **a** Structure of BeF_3^- complex for βPGK (PDB: **4axx**). Beryllium (*lime green*) is “in-line” between a $\text{O}^{3\text{B}}$ of ADP and 3PG. **b** Nicotinamide phosphoribosyl transferase (PDB: **3dhf**) catalyses displacement of pyrophosphate from C1 of ribose 5-phosphate. Structures of two overlaid complexes show BeF_3^- bound to $\text{N}\eta$ of His247 and one fluorine coordinating Mg^{2+} (*green sphere*). PRPP reactant C1' (*cyan sphere*) moves 1.8 Å to bond to nicotinamide N1 (atom colors: fluorine, *light blue*; beryllium, *lime*; nitrogen, *blue*; oxygen, *red*; protein residues are in *grey*; nucleotides in *cyan*) (**a** is reproduced from [27])

3.1.4 Structural Conclusions

The significant ability of beryllium (II) fluorides to complete tetrahedral coordination by binding to an anionic oxygen has made them good isosteric and electrostatic GSAs of phosphate for a wide range of uses [29]. Bond lengths for Be–F and Be–O are close to those for P–O ($1.6 \pm 0.5 \text{ \AA}$) and the strong ionic character of the Be–F bond means that its fluorines readily accept H-bonds from a range of donors and/or coordinate to Group 2 metal ions [30]. Thus, fluoroberyllates have been used beneficially to study changes in major conformations of proteins by crystallography, NMR, and EM, while studies on $\text{ADP}\cdot\text{BeF}_3^-$ have supported investigations on ATPases that drive various mechanical processes at a molecular level, particularly for myosin [31–36]. They have proved especially valuable for the identification of near attack conformations (NACs) in enzyme mechanisms, notably for β -phosphoglucomutase (βPGM) [37].

4 MF_x in Transition State Analog Complexes

4.1 Tetrafluoroaluminate TS Complexes— AlF_4^-

Aluminum (III) forms stable fluorides in water, the mixture of octahedral species including $\text{AlF}_2^+\cdot 4\text{H}_2\text{O}$, $\text{AlF}_3\cdot 3\text{H}_2\text{O}$, $\text{AlF}_4^-\cdot 2\text{H}_2\text{O}$, and $\text{AlF}_5^-\cdot \text{H}_2\text{O}$, depending on the concentration of fluoride [38, 39]. Crystal structures for octahedral $\text{GDP}\cdot\text{AlF}_4^-$ protein complexes [18–20] were prompted by the discovery that aluminum plus fluoride stimulates the activity of small G proteins in the presence of GDP [40], while ^{19}F NMR analysis of a $\text{GDP}\cdot\text{AlF}_x$ complex for $\text{G}\alpha$ [41] confirmed that they

could mimic bound GTP [42]. All 114 crystallographic AlF_4^- complex structures in the PDB (PDB ligand: **ALF**) are octahedral and have aluminum sandwiched between donor and acceptor atoms, predominantly oxygens. Unlike beryllium, the aluminum is well defined in the electron density map (Fig. 6a) and can accept a neutral oxygen as one axial ligand. However, aluminum forms insoluble $\text{Al}(\text{OH})_3$ above pH 7.5 [38, 39], which restricts the stability of aluminum fluoride complexes to pH <8.

4.1.1 Aspartyl Tetrafluoroaluminates

Fourteen PDB structures have tetrafluoroaluminate bonded to an aspartate with an essential Mg^{2+} in a six-membered ring. They align well on the best resolved complex, β -phosphoglucomutase (β PGM, PDB: **2wf7**; 1.05-Å resolution (Fig. 6b), with four equatorial oxygen ligands coordinating the catalytic Mg^{2+} . The structures fit into two subsets: six members of the first group have a second aspartate sub-adjacent to the first (Asp8 and Asp10 in β PGM). The $\text{O}_A\text{-Al-O}_D$ bonds are “in-line” ($167.5 \pm 7.0^\circ$) with the aluminum midway between the two oxygens (separation $3.9 \pm 0.1 \text{ \AA}$) and have a catalytic aspartate that accepts a short H-bond from the apical water/hydroxyl group ($2.59 \pm 0.05 \text{ \AA}$) to align this oxygen for nucleophilic attack on phosphorus [43].

The second subset has ATPases involved in pumping calcium, copper, and zinc ions. They use an aspartyl phosphate intermediate, whose TS for hydrolysis is mimicked by an octahedral AlF_4^- . An axial water oxygen forms short H-bonds to an invariant glutamate ($2.5 \pm 0.1 \text{ \AA}$) and to a threonine carbonyl ($2.57 \pm 0.05 \text{ \AA}$), which clearly orientate and polarize the water for “in-line” attack on the aspartyl phosphate [44].

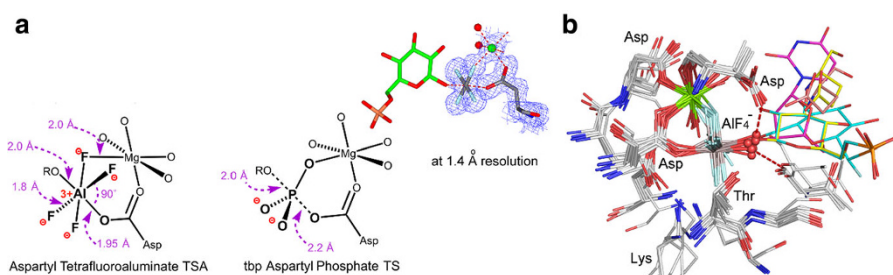


Fig. 6 **a** Typical aspartyl tetrafluoroaluminate structure with catalytic magnesium coordination (*left*). Aspartyl phosphate complex with catalytic magnesium from phosphoserine phosphatase (PDB: **1j97**) for comparison of geometry (*center*). Electron density map for the 1.2-Å resolution structure for β -phosphoglucomutase (PDB: **2wf8**) (*right*). **b** Structures of 13 aspartyl tetrafluoroaluminates aligned on C α . Octahedral aluminum is coordinated to Asp-O $^{\delta}$, in a six-membered ring with Mg_{cat} and “in-line” with a nucleophilic water oxygen (*red sphere*) or the OH group of a nucleoside or hexose reactant (*rainbow colors*) (atom colors: fluorine, *light blue*; aluminum, *grey*; nitrogen, *blue*; oxygen, *red*; magnesium, *green sphere*) (Figure reproduced from [27])

4.1.2 Nucleotide Guanosine Diphosphate (GDP) Tetrafluoroaluminates

GDP forms 50 AlF_4^- complexes that constitute isoelectronic but non-isosteric mimics of GTP in a broad range of proteins. The best resolved 21 align remarkably well (Fig. 7a), with aluminum bonded to $\text{O}^{3\text{B}}$ on GDP and the Mg_{cat} coordinated to F_1 and $\text{O}^{1\text{B}}$ in a six-membered ring. The guanosine base and ribose usually occupy a common conformation (Fig. 7a). The geometry of the AlF_4^- moiety is regularly octahedral, with “in-line” $\text{O}_\text{A}-\text{Al}-\text{O}_\text{D}$ angle $172.8 \pm 7.1^\circ$. All structures have an axial oxygen ligand (Fig. 7a, red spheres) coordinated to aluminum that is trigonal planar with respect to two H-bond acceptors: the backbone carbonyls of a threonine and a glutamine side-chain (occasionally a water) (Fig. 7a, lower right, red spheres).

4.1.3 Nucleotide Adenosine Diphosphate (ADP) Tetrafluoroaluminates

Forty-nine octahedral structures have AlF_4^- bonded to a terminal oxygen of ADP ($\text{O}^{3\text{B}}$) to mimic ATP in the TS. They are found in kinases, hydrolases, isomerases, ATPases, myosins, helicases, transporter pumps, and nitrogenase. The 31 that are best resolved have an axial $\text{O}_\text{A}-\text{Al}-\text{O}_\text{D}$ distance of $4.05 \pm 0.03 \text{ \AA}$ with an “in-line” angle of $170 \pm 8^\circ$ and most have water as the second oxygen ligand with a catalytic Mg^{2+} coordinating one of the fluorines. This is illustrated for F1 ATPase (PDB: 1h8e) (Fig. 8a). In contrast to the uniform conformation for complexes with GDP (Fig. 7), complexes with ADP show a great variety of conformations, as illustrated for 16 well-resolved structures (Fig. 8b).

4.2 Octahedral Aluminum Trifluoride Phosphate TS Mimics

An aluminum trifluoride moiety accepts three oxygens to give an octahedral, six-coordination TSA complex in three examples. In the small G protein Rab5a, the

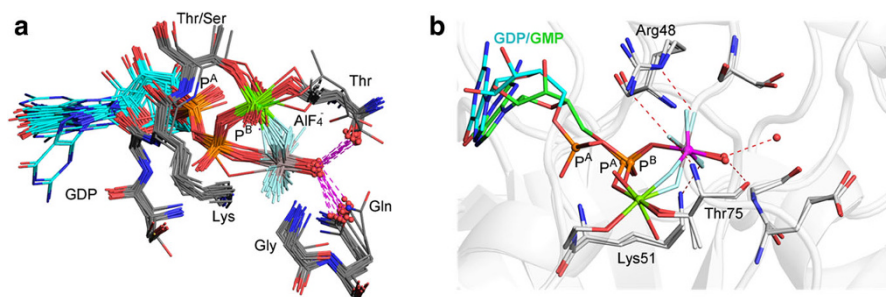


Fig. 7 **a** Twenty $\text{GDP}\cdot\text{AlF}_4^-$ structures aligned on α -carbon atoms of the invariant hexapeptide (in PDB: 2gj8). AlF_4^- is locked in a six-membered ring (center) with Mg_{cat} (green spheres) coordinating F_1 and a P^{B} oxygen. Octahedral coordination to Mg^{2+} is provided by a second P^{B} oxygen, two waters, a Thr hydroxyl (right), and a Ser/Thr hydroxyl (top). $\text{P}^{\text{B,G}}$ oxygens H-bond to a Lys (center). **b** Structures of hGBP1 with a $\text{GMP}\cdot\text{AlF}_3$ complex (cyan) aligned with a $\text{GDP}\cdot\text{AlF}_4^-$ complex (green) showing occupancy of the catalytic site by the AlF_3^0 mimic of P^{B} (magenta sphere) and by the AlF_4^- mimic of P^{G} (grey sphere) (atom colors: GDP, cyan; GMP, green; magnesium, green; fluorine, light blue; amino acids, silver; nitrogen, blue; oxygen, red)

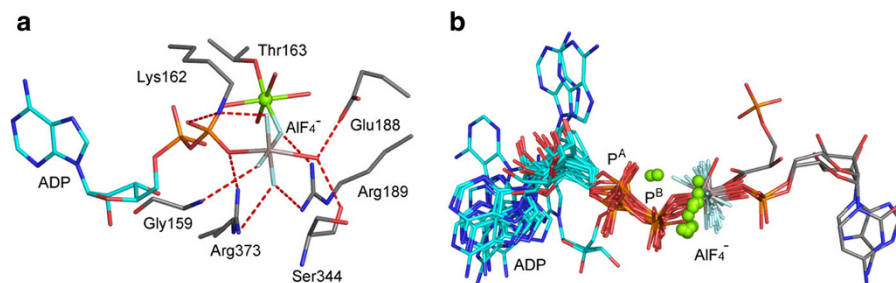


Fig. 8 **a** F1 ATPase TSA complex (PDB: 1h8e) with ADP·AlF₄⁻ showing local charge balance for five +ve and five -ve charges. **b** 16 ADP·AlF₄⁻ complexes aligned for C5', P^A, P^B, and Al show great variety in ATP analog conformations (atom colors: adenosines, cyan; magnesium, green spheres; fluorine, light blue; aluminum, gold; amino acids and second substrates, grey; nitrogen, blue; oxygen, red)

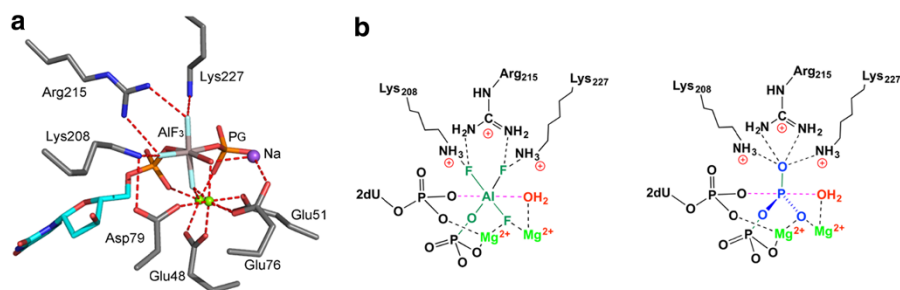


Fig. 9 **a** Aluminum trifluoride structure for dUTPase (PDB: 4di8). UMP (cyan) coordinates aluminum (grey) with in-line water (red) and with PO₄⁻ adjacent to the leaving O^{3A}. Two magnesiums (green spheres) are located by coordination to the reactants and to four carboxylate residues (amino acids in grey). **b** Cartoon showing octahedral aluminum trifluoride sharing the tbp coordination of the true TS for a phosphoryl group (colors: nucleoside, cyan; magnesium, green sphere; aluminum, grey; sodium, purple; amino acids, silver; nitrogen, blue; oxygen, red) (Figure adapted from [27])

mutation A30P enables addition of the side chain hydroxyl of Ser29 to aluminum trifluoride (PDB: 1n6k). In the case of hPGK, the K219A mutant has a water as the fourth ligand coordinated to the aluminum [45]. Thirdly, for a bacterial dUTPase, aluminum trifluoride takes the place of the P^B in dUTP with coordination to two oxygens from the β-phosphoryl group and to the water nucleophile to complete the octahedral array (Fig. 9a, b). This significant structure provides a unique example where nucleophilic attack is directed at a non-terminal nucleotide phosphorus [46].

5 MF₃ Improved Geometry Transition State Mimics

5.1 MgF₃⁻, Trifluoromagnesate

Magnesium does not form mixtures of stable fluorides in water at sub-molar concentration: only one resonance for magnesium fluoride is seen in ¹⁹F NMR

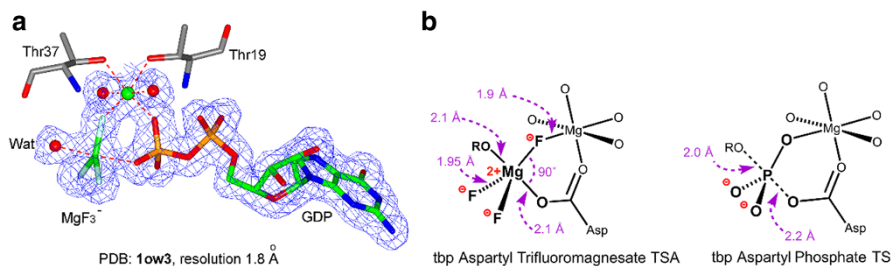


Fig. 10 **a** MgF_3^- complex with GDP for RhoA/GAP (PDB: **1ow3**) showing electron density. **b** Typical MgF_3^- complex with aspartate residues in a six-membered ring with the catalytic Mg^{2+} (*left*) compared to an aspartyl phosphate (*right*) (Figure adapted from [27])

solution spectra, that of MgF^+ . While MgF_2 is moderately soluble in water (2 mM), it has an estimated dissociation constant of 10^{-5} M [47]. Trifluoromagnesate protein complexes were first anticipated based on magnesium-dependent fluoride inhibition studies, and they led directly to the identification of MgF_3^- in a *tbp* crystalline TSA complex for the small G protein RhoA-RhoGAP (Fig. 10a) [24, 48]. The PDB now has 16 entries for trifluoromagnesate (PDB ligand: **MGF**) while a further three entries assigned as *tbp* AlF_3^0 have been shown by ^{19}F NMR to be MgF_3^- complexes [49–51]. Standard coordination chemistry identifies magnesium as being regularly octahedral, forming complexes with six (oxygen) ligands. By contrast, trifluoromagnesate in protein complexes is unexpectedly five-coordinate. This makes it ideal for mimicking *tbp* phosphoryl transfer and, moreover, MgF_3^- is isoelectronic with PO_3^- . Examples of its use include complexes of small and large molecule kinases, mutases, phosphatases, and hydrolases, which invariably involve fluoride coordination to a catalytic Mg^{2+} (two magnesiums in the case of some protein kinases). These are usually octahedral and built into a cyclic six-membered ring structure, as shown for aspartyl phosphate mimics (Fig. 10b). They have an axial $\text{O}_A\text{-Mg-O}_D$ distance of 4.19 ± 0.08 Å with an in-line angle $171.4 \pm 3.9^\circ$.

5.2 AlF_3^0 , Aluminum Trifluoride

There are now 56 examples of structures purported to have an AlF_3^0 core. Three of them are octahedral, while ^{19}F NMR has established that another three are MgF_3^- . For the remaining majority, only structures of two alkaline phosphatase complexes (AP) can be confidently identified as having a *tbp* AlF_3^0 core (Fig. 11). One is in mutant AP_{P300A} (PDB: **1kh5**), where two catalytic Zn^{2+} ions coordinate one fluoride while Ser102 and a zinc-coordinated water provide the axial ligands for the *tbp* aluminum (Fig. 11b). What about the remaining 48 “ AlF_3^0 complexes”?

The pH dependency of the transition between octahedral and *tbp* structures of AlF_x complexes in protein crystal structures was proposed to involve a switch from AlF_4^- to AlF_3^0 at elevated pH [52]. However, studies on the pH dependence of the solubility of aluminum ion [38, 39] provided an alternative interpretation. $\text{Al}(\text{OH})_3$ precipitates at $\text{pH} \geq 8$, which results in aluminum being superseded by magnesium in protein MF_x complexes at high pH, with a consequent change in geometry from

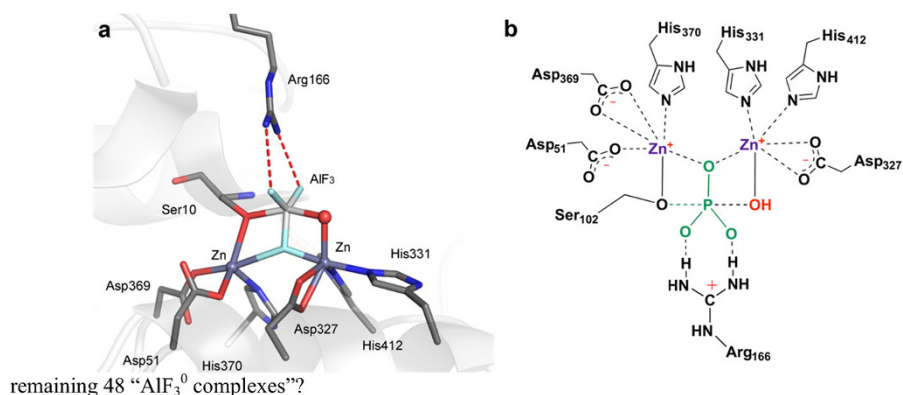


Fig. 11 **a** Structure of the catalytic center for alkaline phosphatase complexed to AlF_3 (PDB: **1kh5**). **b** Cartoon of the coordination organization in the active site with transferring phosphoryl group (blue) and nucleophilic water (red) (Figure adapted from [27])

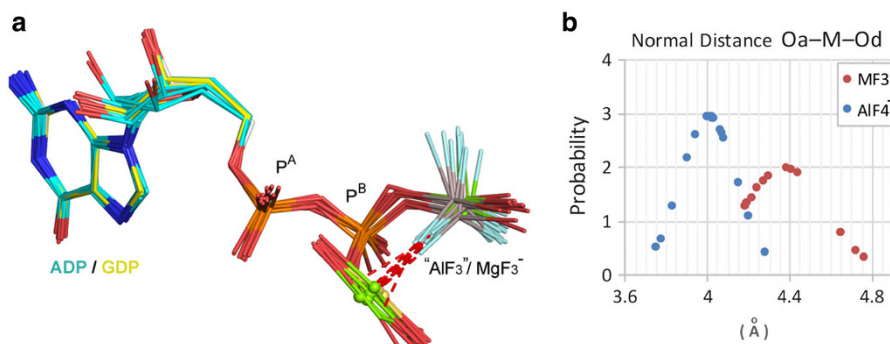


Fig. 12 **a** Overlay of five $\text{GDP}\cdot\text{MgF}_3^-$ (yellow) and eight $\text{GDP}\cdot\text{AlF}_3^0$ (cyan) complexes to show the geometric uniformity of the two sets of TSA structures. **b** Normal distribution plots for the $\text{O}_A\text{-M-O}_D$ distance for this set of 13 structures (red) and the corresponding $\text{O}_A\text{-Al-O}_D$ distance for 18 $\text{GDP}\cdot\text{AlF}_4^-$ TSA complexes (blue)

octahedral to tbp. That conclusion has now been validated by pH-dependent ^{19}F NMR analyses for several enzymes [50, 53]. In some marginal cases, e.g., protein kinase A (cAPK) and PSP, there is mixed occupancy of the active site by tbp and octahedral complexes in the crystal [43, 50, 51]. In geometric terms, “ AlF_3^0 ” tbp complexes closely map on those of trifluoromagnesates: axial $\text{O}_A\text{-M-O}_D$ bonds 4.29 ± 0.39 Å (Fig. 12b), and M-F bonds 1.75 ± 0.12 Å. It seems likely that some, or many, of these “ AlF_3^0 ” complexes are trifluoromagnesates: a conclusion supported by geometric analysis for both families of complex.

5.3 A Combined MgF_3^- - and AlF_3^0 Structural Analysis

A statistical analysis of the structures of AlF_3^0 and MgF_3^- complexes contributes to the resolution of this compositional uncertainty. The near-invariant geometry of

octahedral AlF_4^- complexes for GDP makes them a useful set for comparison with the corresponding set of tbp MF_3 complexes. Thus, eight GDP “ AlF_3^0 ” structures for small G proteins align very well with those for five MgF_3^- complexes (Fig. 12a). The axial separation for the donor and acceptor oxygens in these combined 13 GDP- MF_3 TSAs is $4.38 \pm 0.20 \text{ \AA}$, significantly distinct from the corresponding average for 19 GDP- AlF_4^- complexes, $4.02 \pm 0.14 \text{ \AA}$, and clearly supported by normal distribution analysis (Fig. 12b). The conclusion is: For “ AlF_3^0 ” read MgF_3^- !

Taking “ AlF_3^0 ” together with trifluoromagnesates, a common general pattern of axial ligands emerges. The MF_3 species requires at least one anionic oxygen. β -Oxygens from ADP (33 structures) and GDP (24 structures) provide the overwhelming majority of examples while aspartate (11 structures) is also significant. Water (27 structures) is the dominant neutral axial ligand while serine and threonine hydroxyls appear less frequently. Significantly, there is no example of both axial ligand positions being occupied by two neutral ROH groups.

5.4 MgF_4^- , Tetrafluoromagnesate

There are several structures for the Ca^{2+} pump ATPase that have been assigned as tetrahedral MgF_4^- moieties without objective experimental validation. Magnesium is only exceptionally four-coordinate and then it usually has sterically bulky ether oxygens as ligands [54]. The tetrahedral “ MgF_4^- ” moiety in all PDB examples is remote from ADP, is coordinated to a second magnesium, and has one or more of its four “fluorine” atoms in close contact with a backbone carbonyl oxygen, as shown for PDB: **1wpg** (Fig. 13a) [55]. Such “ MgF_4^- ” behavior closely resembles the six-membered ring tbp structures common for MgF_3^- complexes of aspartate (Fig. 10). Crystallographic re-refinement, with MgF_3^- replacing MgF_4^- for **1wpg**, can produce an equally valid structure. Thus, unless established by further measurements, a more consistent chemical interpretation for all such “ MgF_4^- ” situations is that they are trifluoromagnesates that mimic the TS for hydrolysis of an aspartyl phosphate. Subsequent work has described a similar tetrahedral moiety for the Na^+/K^+ pump ATPase (PDB: **2zxe**) [56].

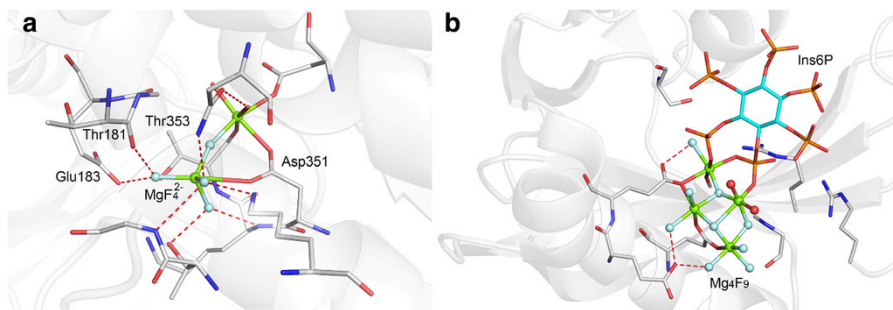


Fig. 13 **a** Structure of Ca^{2+} pump ATPase with MgF_4^- (PDB: **1wpg**). Coordination for MgF_4^- is typical of an aspartyl trifluoromagnesate complex (colors: fluorine, *light blue*; magnesium, *green*; nitrogen, *blue*, oxygen, *red*; carbons, *silver*). **b** Structure of hPPIP5K2 (PDB: **2q9p**) to show the “ Mg_4F_9 ” cluster adjacent to phosphates 4 and 5 of Ins6P

Finally, the most remarkable MF_x structure is that of a human diphosphoinositol phosphatase, co-crystallized with *myo*-inositol *hexakis*-phosphate and then soaked with sodium fluoride (PDB: **2q9p**) [57]. This complex has four octahedral magnesiums with nine ligands assigned as fluorines in a complex that embraces MgF_2 , MgF_3 , MgF_4 , and MgF_5 species in a single block. It also offers the first example of octahedral MgF_x (Fig. 13b). Its core appears related to the Rutile structure of MgF_2 , which is characterized by octahedral magnesium and trigonal planar fluorine [58].

6 ^{19}F NMR Studies of MF_x Complexes

The very high gyromagnetic ratio ($25.18 \times 10^7 \text{ T}^{-1} \text{ s}^{-1}$) of ^{19}F gives it very high sensitivity in NMR, which facilitates detection of fluorine-containing species at low concentration in large molecular weight complexes, as illustrated for AlF_4^- and MgF_3^- complexes with RhoA·GAP·GDP (Fig. 14a, b) [44, 50, 51, 59, 60]. In the context of TSAs and GSAs, the chemical shifts of ^{19}F resonances provide a key measure of interactions between MF_x moieties and their protein hosts, and report the electronic environment of the fluorine nuclei. When combined with NMR computations, they also act as indirect reporters of changes in electronic environment experienced by phosphoryl oxygen atoms in transfer reaction TSs [44, 61, 62]. ^{19}F NMR resonances display a high degree of dispersion and are calculable with good precision from QM analysis of electronic distribution [37, 51, 63], showing resonances strongly affected by neighboring H-bond donors. Reduction in the number of H-bond partners generally results in upfield shift of ^{19}F resonances, as shown clearly in a comparison of the G6P and the 2-deoxy G6P

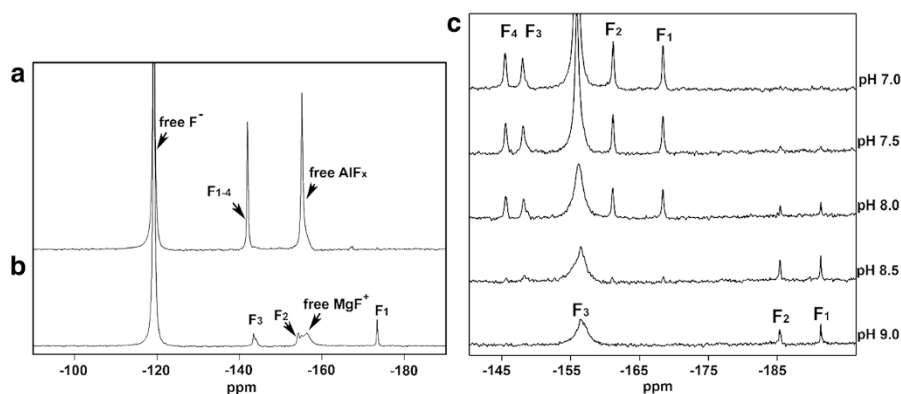


Fig. 14 **a** 1D ^{19}F NMR spectra of: **a** RhoA/GAP·GDP· AlF_4^- TSA complex with four fluorines rotationally averaged [44, 48]. **b** RhoA/GAP·GDP· MgF_3^- TSA complex with three fluorines resolved [44]. **c** Conversion of the cAPK·ADP· AlF_4^- TSA complex to the cAPK·ADP· MgF_3^- TSA complex in a pH titration from pH 7.0 to pH 9.0 [50]. The ^{19}F resonance at -119 ppm in each spectrum is from free F^- ion, while the broad peaks around -156 ppm are from unbound MgF^+ and AlF_x species, which may often overlap with protein bound fluorine resonances. In those cases, presaturation of the free fluoride resonance at -120 ppm can be applied to eliminate the unbound metal fluoride signals

complexes of β PGM (-18.1 ppm) [53]. In general, the resonance of the fluorine coordinated to a catalytic magnesium is always the most upfield, because of depletion of H-bond coordination [44, 45, 50, 51, 53, 60, 62]. Proton distribution near fluorine nuclei can be further assessed through the quantitation of ^{19}F - ^1H NOEs using perdeuterated enzyme in protonated buffer to suppress the ^1H - ^1H spin diffusion [49, 51], while resonance assignment of exchangeable ^1H nuclei in the protein enables unambiguous assignment of individual ^{19}F NMR resonances. The number of H-bond donors can also be assigned based on solvent induced hydrogen/deuterium primary isotope shifts (SIIS) of ^{19}F NMR resonances. For $\text{NH}\cdots\text{F}$ and $\text{OH}\cdots\text{F}$ H-bonds to MF_x moieties, the SIIS size reflects local proton densities [64], and this has been used to assign F_A , F_B , and F_C in a $\beta\text{PGM}\cdot\text{MgF}_3^-\cdot\text{G6P}$ TSA complex [62].

Scalar coupling between nuclei involved with $\text{N-H}\cdots\text{F}$ H-bonds is an additional parameter that shows details of the coordination of the MF_x moiety by the protein. $^1J_{\text{HF}}$ and $^2J_{\text{NF}}$ couplings have been reported for individual $\text{NH}\cdots\text{F}$ pairs, with values up to 59 and 36 Hz, respectively [62]. All the effects described above, SIIS, NOE, chemical shifts, and scalar couplings, correlate closely with H-bonding orientations and distances obtained from high resolution crystal structure analysis. ^{19}F chemical shifts are invariant over the pH range 6.5–9.5, they signal that there is no detectable change in protonation state of the enzyme in the environment of the TS complex, but the pH dependence of ^{19}F NMR resonances and multiplicity can identify a switch from AlF_4^- to MgF_3^- complexes above pH 8, as illustrated for cAPK (Fig. 14c) [57].

NMR measurements of ^{19}F nuclei in the active site of MF_x TSA complexes thus provide a picture of charge distribution between the phosphoryl group mimic and the protein. The good relationship between ^{19}F NMR chemical shifts and SIIS values illustrates the dominant influence that very localized H-bonds have on shaping charge density on MF_x moieties.

7 Computational Analyses of MF_x Complexes

7.1 Balancing Accuracy of Energy/Structure and Conformational Sampling

A computational simulation of the structure and bonding of a biochemical system at atomic resolution has two demanding features:

1. The solution of accurate molecular energies, ideally with as little parameterization as possible;
2. The exhaustive consideration of relevant conformations of macromolecules.

For the simulation of biomolecules, it is unavoidable that both criteria must be approximated to varying degrees. In practice, different computational methods put different emphasis on one or the other of these two features. Any useful calculation must meet both criteria adequately. Solutions of the energy of a macromolecule, and thence its structure, should be made for each conformer of the molecule. Hence, the

task of achieving reliable energies severely raises the cost of the computation. This constraint therefore drives down the number of conformers to be computed, with the risk that the program may fail to examine the specific conformation most relevant for the reaction under investigation.

To attain a compromise between these two features, the methodology used has to strike a balance between defining a central quantum mechanics (QM) zone and a molecular mechanics (MM) zone dealing with the major part of the macromolecule and environment. The combination of the two regions is called a QM/MM calculation. A QM description is necessary to describe bond-breaking-making processes or electronic excited states because molecular mechanics cannot describe these phenomena. Different balances between these two features are achieved by different choices in the apportionment of resource to the QM region. These include Kohn–Sham density functional theory (KS-DFT) [65–69] and empirical valence bond (EVB) [70, 71], while similar choices exist for the MM zone. However, the QM zone is the priority region.

7.2 Tradeoff in Accuracy of Energy/Structure: Parameterization Simplification vs. Mathematical Complexity

Accurate molecular energies can be obtained in an unbiased, systematically correctable manner [72–74] to get the desired accuracy. However, the computational resource required is very expensive, and is usually unacceptable because resource must be apportioned to adequate conformational sampling. In general, either an approximate QM method such as KS-DFT is used, or a heavily parameterized model is designed for a specific system such as EVB. Briefly, parameterization can tailor a QM method specifically to that molecule under analysis—and thereby eliminate many mathematical degrees of freedom. Hence, the calculation can be performed rapidly and can incorporate greater conformational sampling, but it must rely on the assumption that the reduced mathematical form faithfully represents the true quantum mechanics. By contrast, the various KS-DFT forms have parameters which are fixed by the design of the functions, and are completely independent of that particular biomolecule under investigation. Thus, the application of KS-DFT to a specific biomolecule has no freedom to change parameters to suit the target. Hence, KS-DFT deploys a more general mathematical framework, and more faithfully echoes exact quantum mechanics within budget.

7.3 Tradeoffs in Conformational Sampling: Dynamics vs. Statics

In order to balance the budget of the computation program, a choice has to be made between dynamics and statics. On the one hand, a *dynamics* description delivers an explicit femtosecond-by-femtosecond time evolution of the atoms, boosted by metadynamics [67]. On the other hand, a *statics* analysis of a few discrete critical points along the reaction identifies TSs and/or intermediates as maxima/minima along the reaction coordinate. Each has its strengths and weaknesses.

A dynamics computation shows the true time-evolution of the molecular system, especially how atoms re-arrange to move along all possible reaction paths, step-by-

step. All possible chemical reactions/conformations are sampled in due frequency with the Boltzmann distribution of states. The computation does not “target” a specific reaction path. TSs are rare-events, require long simulations or metadynamics [67], and so demand a smaller QM zone to allow an adequately fast calculation. This reduction of the QM zone, relative to that for statics described below, makes possible the conformational sampling needed to find the right state. A balance has to be struck between faithfully computing dynamics or prioritizing accurate energy calculations.

The choice for statics in following a reaction path, selected a priori, enables easy identification of the TSs for bond-breaking-making using standard quantum chemistry algorithms. Mathematical properties of energy maxima (TSs) and minima (intermediates) can be sought automatically. Users can seek out any desired pathway, but they have to sacrifice an understanding of the relative values of each path. This requires minimal computational resource compared to that required for a dynamics calculation, and so can accept a much larger QM region and/or a more accurate QM calculation. However, the a priori choice of the conformation is risky: it depends strongly on the accuracy of choice of the *true* TS conformation, which may or may not be found among existing crystal structures in the PDB.

Take for example the first mechanistic step of the hairpin ribozyme. This is cleavage of the bond from the 3'-phosphate of A-12 to the 5'-oxygen of G13 to form a 2',3'-cyclic phosphate which has been modeled as a pentacoordinate vanadate TSA structure (PDB: **1m5o**, 2.2-Å resolution; Fig. 15a). The two proximate nucleobases are G8' and A57' whose catalytic roles are controversial: there is good support for protonation of A57'-N¹ but some debate whether G8' is deprotonated on N¹ or not. The computation accepted formation of an intermediate pentacoordinated phosphorane and then posed the question: “How and when is the proton removed from A8-O2' and transferred to PO^{2A}?” A thorough benchmark study of

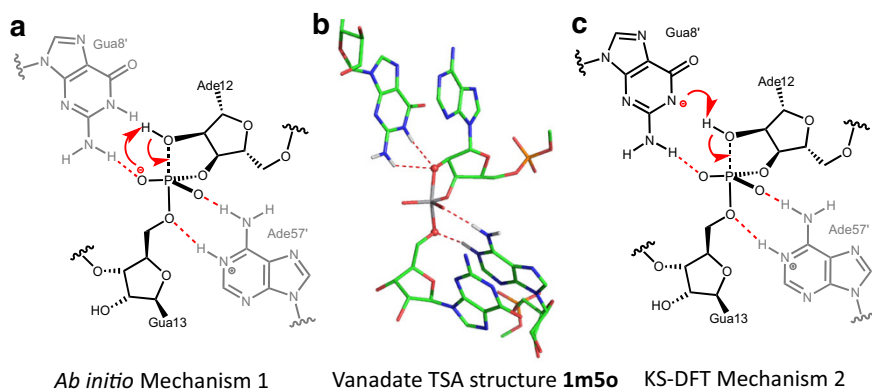


Fig. 15 **a** *Ab initio* mechanism for the first step of the hammerhead ribozyme reaction showing PT from Ade12-O2' to PO² in the formation of a transient pentaoxyphosphorane species. **b** Crystal structure of the hammerhead ribozyme as a tbp vanadate complex (PDB: **1m5o**). **c** DFT computed mechanism for PT to the anionic Gua8' preceding bond formation from P to Ade12-O2' (colors: carbon, *green*; nitrogen, *blue*; oxygen, *red*; vanadium, *grey*; H-bonds, *red dashes*)

comparative QM/MM methods has been applied to this mechanism [75], comparing an ab initio method (i.e., no parameterization) with a KS-DFT method (a small number of fixed parameters). As the energy differences for the two paths should be comparable, a careful QM analysis was necessary. Ab initio energy barrier prediction matched experimental estimates, giving good support to Mechanism 1 (Fig. 15b) with a direct four-center proton transfer (PT), not involving a neutral G8'.³ However, it was observed that the flow of atoms predicted by the parameterized method was inconsistent with benchmark calculations. The authors therefore employed umbrella sampling in a KS-DFT analysis to achieve dynamics convergence, and found Mechanism 2 to be preferred with the anionic G8' acting as a base to abstract the proton from Ado12-O2' (Fig. 15c).

7.4 KS-DFT as a QM Region Description

The most commonly chosen methodology for describing the QM portion is KS-DFT⁴ functionals. This makes a practical compromise between precision and cost of computation (Sect. 7.2). KS-DFT dispersion-corrected functionals can now describe molecular geometries to within 0.02 Å [76–78], and are particularly valuable because their description of energies and geometries is unbiased. They have been shown to describe basic bond-breaking behavior, H-bonding, and energetics [77–81]. In a QM/MM calculation, the boundary region at the interface between the QM zone and the MM zone can be problematic because the energies on the QM side need not be the same as those on the MM side. Any mismatch between kinetic and potential energies across the boundary leads to un-physical behavior. This boundary problem can be eliminated by depriving large portions of the macromolecule of an MM force field, which calls for a tradeoff between:

1. A more faithful representation of long-range chemical interactions and a potentially problematic boundary between zones introducing artifacts; and
2. Neglect of long-range chemical interactions altogether, with no un-physical artifacts introduced by the QM/MM boundary.

Either choice is problematic, and a case-by-case decision must be made. In some cases, to avoid boundary complexities exclusively QM calculations have been used, usually KS-DFT. They usually rely wholly on experimental data from the structure of a TS mimic, thereby obviating the need for a conformational search, and allowing full investment of the computational resource to maximize the size of the QM zone. For example, a recent study of GTP hydrolysis by RhoA/GAP to identify the reaction mechanism employed a large QM calculation [44]. The KS-DFT zone was large enough to embrace the reacting methyl triphosphate, its coordinating magnesium and nucleophilic water, and also residues from some

³ NB In other computational studies, such four-center PTs for phosphoryl reactions have been deemed to be very high energy.

⁴ In literature meant for using DFT in organic, biological, or inorganic applications, “KS-DFT” and “DFT” are used largely interchangeably. Theoreticians draw a distinction between these terms; KS-DFT is a subset of DFT for the given selection of expressing the kinetic energy in terms of orbitals.

18 additional amino acids that contribute to the stability of a network of 21 H-bonds which deliver the conformation of the TS for water attack on P^G . Successive rounds of DFT computing established that contributions from atoms in the third solvation shell of the transferring phosphoryl group were required to deliver stability. The result was a QM region of 91 heavy atoms (181 total atoms) (Fig. 16). Because the starting TSA structure (PDB: **1ow3**) was of sufficiently high resolution (1.8 Å) to give confidence that the study was based on a reliable model of the TS, the addition of an MM contribution was bypassed, obviating the need for a QM/MM boundary. However, this limited the computational output to geometric and spectroscopic features. The absence of conformational sampling, sacrificed because of the large and very expensive QM region, also limits comment on activation energies.

The iterative computational procedure delivered a mechanism in which the nucleophilic water is doubly protonated with H-bonds to carbonyl oxygens of both T37 and Q63 residues until after the TS for bond making/breaking, thereby orientating the nucleophilic water for good orbital overlap with the antibonding $O^{3B}-P^G \sigma^*$ orbital. PTs are not seen in the TS, but occur subsequently.

7.5 EVB as a QM Region Description

The Empirical Valence Bond method deploys a simplified mathematical framework to achieve the most rigorous possible conformational sampling. In essence, the EVB framework is largely a molecular mechanics based method, with the exception of its representation of a single “orbital” for each molecule, identified as involved in the bond-breaking-making reaction. No other electrons/orbitals are represented explicitly. This framework thus imposes the presumption that only a single orbital is involved in the bond-reorganization for a reaction. The EVB parameterization process is fundamentally chemistry-imposed: it identifies, a priori, what orbitals are involved and dictates chemistry-based molecular mechanics energy functions. This is in sharp contrast to a KS-DFT prescription of a QM region, which is fundamentally agnostic of chemistry, not defining bonds or selecting orbitals targeted for reaction, but merely defining a total number of electrons and nuclei involved, with no presumption of chemistry. As a result of the EVB simplifications, larger-scale changes in molecular conformation can be observed. In this way, the initial conditions of the experimental crystal structure are not a trap; the computational protocol allows the biomolecule to move freely.

A study of the mechanism of DNA polymerase β provides a good example of the application of the EVB methodology [82]. The questions under examination were (i) the destination and timing of PT from the nucleophilic 3'-OH, with three aspartates and water as potential acceptors, and (ii) the concerted or stepwise nature of phosphorus migration.⁵ The starting structure was native DNAPol β (PDB: **2fms**, 2.0 Å resolution) and some 70 heavy atoms were included in the QM zone (Fig. 17), linked to the assumption that the reaction takes place in the three steps: a PT from

⁵ A previous study favored a concerted reaction path (Lin et al. [83]).

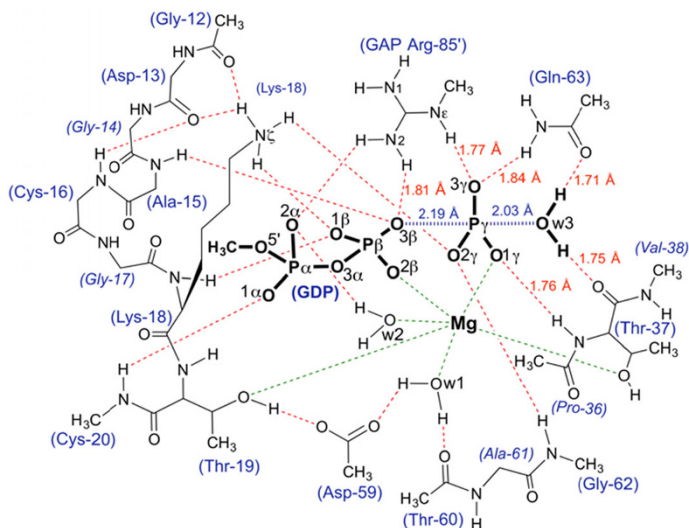
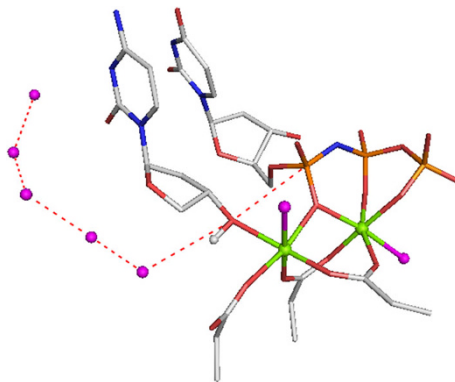


Fig. 16 Atoms in the QM zone for KS-DFT computation of the TS for GTP hydrolysis by RhoA/RhoGAP showing the 21 H-bonds in the catalytic network (*red dashes*) with ligands coordinated to Mg (*green dashes*). Amino acid residues are numbered according to RhoA sequence plus Arg85' from RhoGAP (Figure taken from Ref. [44])

Fig. 17 EVB analysis of DNAPol β showing atoms used in the QM region, augmented by additional water oxygens (*magenta*) and the focal hydrogen undergoing transfer (*white sphere*), and taken from PDB: 4fms



the primer 3'OH, followed by in-line nucleophilic attack of 3'O⁻ on the dUTP α -phosphate, with reaction completed by departure of pyrophosphate. Particular attention was paid to the electrostatic role of the two catalytic magnesium ions through the course of the reaction. The computations were guided by consideration of the pH-rate profile for wt and mutant polymerases and focused on seven EVB states. This led to the conclusion that PT from the nucleophilic hydroxyl group is to bulk water via a chain of water molecules that extends into the active site, rather than to any of the three neighboring aspartates (as favored in previous studies) and precedes P–O bond formation. The phosphorylation process was found to be

associative with a pentacovalent intermediate,⁶ though the investigation warned that extensive sampling is essential for EVB analysis of reaction mechanisms.

7.6 The Use of MF_x in Computational Studies of Enzyme Mechanisms

As has been described above, protein trifluoroberyllate complexes are fundamentally different from those involving metal fluorides of aluminum and magnesium. They have closely similar tetrahedral geometry and net charge to the parent phosphate, their macromolecular structures have folds and atomic organization that relate to the ground state structures they mimic. On the other hand, AlF_4^- , “ AlF_3^0 ”, and MgF_3^- complexes have geometries that bring two axial ligands into alignment and proximity typical of the TS for concerted phosphoryl transfer, and this results in protein folds and atom coordination that resembles the TS for reaction. Computational enterprises have taken up both of these opportunities for a plethora of purposes. The growth in such studies is illustrated in the chart (Fig. 18). In practice, the three MF_x categories converge for the majority of computer purposes as they are usually transposed into PO_3^- at an early stage in the computation. Therefore, the major part of the following description of computational studies on mechanisms of phosphoryl transfer will be focused on the target protein.

7.6.1 Validation of MF_x as a TSA for Phosphoryl Transfer

Relatively few computational studies have been directed at the structural identity of the MF_x complex per se. A contentious 1.8 Å resolution structure (PDB: **1o03**) focused on a six-atom *tbp* complex for β PGM, initially described as a pentaoxyphosphorane [84]. They have converged on identification of (a) the observed crystal structure as a five-coordinate trifluoromagnesate complex rather than a five-coordinate phosphorus [24], (b) an active site stabilized by an extensive H-bonding network, and (c) a concerted transfer of the phosphoryl group without a stable phosphorane or metaphosphate intermediate [85–87]. They concluded that MgF_3^- is a good TSA that can give insight into the geometry of the phosphoryl transfer TSs. A second example is a QM/MM analysis of the atomic nature of an MF_x moiety in a TSA complex for the key kinase, cAPK [61]. The structure of a *tbp* complex for cAPK·ADP· MF_x was originally described as AlF_3^0 (PDB: **113r**) but QM/MM simulations suggest that MgF_3^- is the correct description of the *tbp* moiety rather than AlF_3^0 , and that MgF_3^- is a near isosteric fit to PO_3^- in the computed TS for the hydrolysis of ATP [61, 88]. This result agrees with a ^{19}F NMR analysis, have been directed at MF_x complexes for cAPK [50]. The computations conclude that this kinase prefers a monoanionic analog (MgF_3^- or AlF_4^-) over a neutral analog (AlF_3^0) to match the –ve charge on the phosphoryl group.

⁶ It should be added that recent structural studies suggest that one of the catalytic magnesiums is lost from its pre-TS location during the bond-making-breaking process and then appears in a new location (PDB: **4klf** and **4klg**).

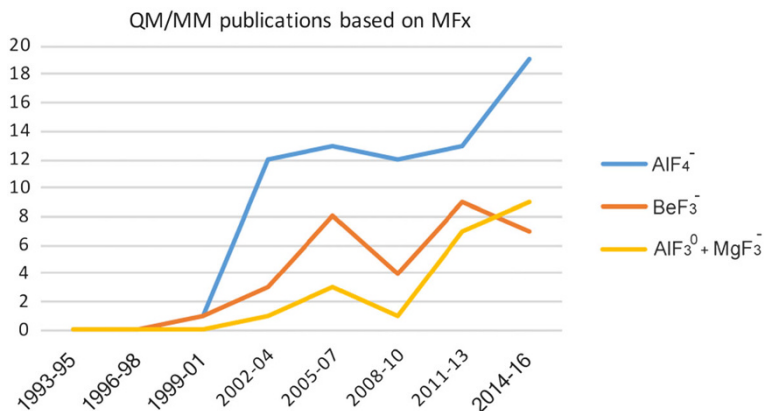


Fig. 18 Growth of computational publications since 1994 showing QM results on AlF_4^- leading with $[\text{AlF}_3^0 + \text{MgF}_3^-]$ comparable to studies based on BeF_3^- complexes

7.6.2 Studies Linking Reaction Mechanisms from Model Systems to MF_x Enzyme Complexes

Early QM studies on phosphoryl transfer analyzed the hydrolysis of methyl phosphate [89] and methyl pyrophosphate [90], added magnesium [91], and then transposed the results into the context of the Ras GTPase active site. The result does not match well to the MF_x structure for Ras-RasGAP (PDB: **1wq1**) because (i) the computed $\text{O}_A\text{—O}_D$ separation lies in the region 4.7–5.5 Å and in the MF_x structure is 4.4 Å. (ii) The computation calls for a second water to facilitate PT [92], however, in those (few) instances where a second water is seen in high-resolution MF_x structures for Ras, it occupies the site vacated by a displaced or missing Gln61 residue, and is in no position to deliver the proposed catalysis (Fig. 7a).

7.7 Computations Transposing $\text{GDP}\cdot\text{MF}_x$ into GTP Enzyme Complexes

7.7.1 Ras Family and GTP Hydrolysis

The use of MF_x TSA structures to identify the TS for hydrolysis of GTP by Ras proteins has been the basis of many computations. Several studies have used PDB: **1wq1** [22], the 2.5 Å-resolution structure of Ras-RasGAP-GDP· AlF_3^0 as starting point, and have employed both QM/MM [92–99] and EVB approaches [100, 101]. Some of these have aroused expert criticism of limitations inherent in the QM/MM approach [102]. The results have varied widely, from a two-step reaction mechanism with bond breaking preceding bond making (i.e. a dissociative process; Scheme 1a) [100], to exclusion of water by the arginine finger [98], tautomeric catalysis [17], electrostatic catalysis [101], a two-water mechanism [92], and sundry rationalizations of the adverse effects of mutations [97, 99, 101]. The QM zone has generally been limited to 30–40 heavy atoms and, in consequence, has not examined the role of the function of several amino acids in contact with the reactants, most

especially the extensive H-bonding network (as in Fig. 16). By contrast, an alternative computational approach using Kohn–Sham DFT analysis for RhoA-RhoGAP hydrolysis of GTP employed a QM zone of 91 heavy atoms, embracing a network of 21 H-bonds, and has attributed catalysis to orbital orientation determined by protein control of H-bonds donated by the nucleophilic water (Fig. 16) [44]. The same study validated the high relevance of MgF_3^- as a TSA by back-computing its structure from that of the calculated structure for the true TS complex for GTP hydrolysis.

7.7.2 Other GTPases and GTP Hydrolysis

A study on the structures of a $\text{GMP}\cdot\text{AlF}_3^0$ complex (PDB: **2b8w**) and a $\text{GDP}\cdot\text{AlF}_4^-$ complex (PDB: **2b92**) for hGBP1, has linked a mechanism for the hydrolysis of methyl triphosphate (MTP) to the two-step hydrolysis of GTP to GDP and thence to GMP by this interferon-activated human GTPase (Fig. 7b). The computation employed dated ab initio QM/MM molecular dynamics to simulate the hydrolysis of both GTP and of MTP as a reference system [103]. The study proposes that GTP hydrolysis involves an indirect, substrate-assisted catalysis mechanism, identifying the nearest general base as Glu99, which is 6.2 Å from the nucleophilic water in the TSA complex. This separation problem was resolved by invoking transmission of base catalysis via one water to Ser73, and thence via a second water to the nucleophilic water. These bridging waters are not present in the substantive (3.2 Å resolution) TSA complex but appear to be imported from a structure of hGBP1 with β , γ -imino-GTP that is clearly an NAC complex (PDB: **2bc9**; 2.8 Å resolution). This investigation merits a cautionary comment on the frailties of a computational analysis based on structures of poor resolution, under-informed by an adequate grasp of mechanisms of phosphoryl transfer.

7.8 Computations Transposing $\text{ADP}\cdot\text{MF}_x$ into ATP Enzyme Complexes

7.8.1 ATP Hydrolysis by Myosin

Myosins are a family of ATP-dependent motor proteins whose role in muscle contraction is driven by ATP hydrolysis. Multiple structures of $\text{Mg}\cdot\text{ADP}\cdot\text{MF}_x$ exist, including BeF_3^- (PDB: **1w9i** and **1mmd**) and AlF_4^- (PDB: **1w9l** and **1wj9**). These structures have been used to identify the catalytic amino acids and locate key water molecules, especially the nucleophilic water that attacks in-line at P^{G} (Fig. 8a). A recent QM/MM computation of the hydrolysis of ATP used a DFT method with B3LYP functional and a 6–31G(d,p) basis set to treat 84 atoms in the active site for the $\text{Mg}\cdot\text{ADP}\cdot\text{BeF}_3^-$ structure of the myosin II head group (PDB: **1mmd**), with ATP modeled by replacing BeF_3 with a phosphate (cf. Fig. 4a) [104]. Although the starting structure for the computation (Fig. 19a) has the nucleophilic water (W_a) in a NAC, as defined by its H-bond proximity to F^2 (2.7 Å) and out-of-line angle for the attack on Be (152°), the simulations delivered a H-bond network that lead to the final product, $\text{H}_2\text{P}^{\text{G}}\text{O}_4^-$. The proposed mechanism involves formation of a

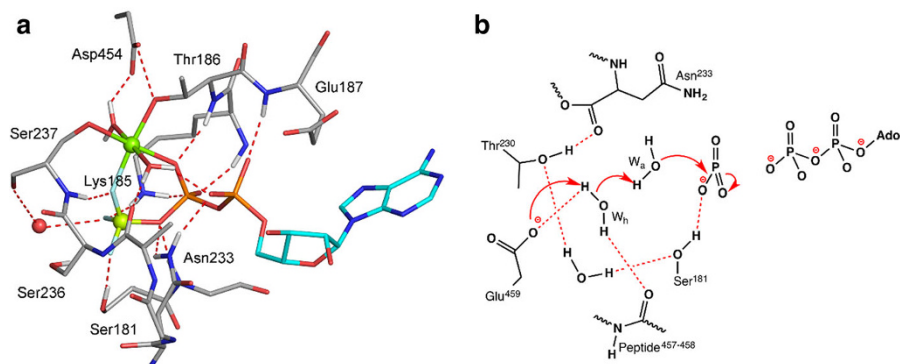


Fig. 19 Mechanism of ATP hydrolysis by myosin. **a** Myosin complex with Mg-ADP·BeF₃⁻ (PDB: **1mmD**) at 1.95 Å, with beryllium (lemon sphere) trifluoride bonded to O^{2B} of ADP (cyan). Amino acids in network (silver, blue, red) shown with H-bonds to TSA complex (red dashes). Nucleophilic water (red sphere) is in a NAC. **b** The key feature of the metaphosphate state is the extreme polarization of water W_a, due to two H-bonds with the Ser237–C=O and water. Attack of W_a on P^G involves attendant PTs via the helper water, W_h (red arrows) [104]

stable metaphosphate intermediate prior to the TS followed by a series of PTs (Fig. 19b).

7.8.2 ATP Hydrolysis by F1 ATPase

F1-ATPase (ATP synthase) is a membrane-bound protein that uses a proton gradient to drive ATP synthesis. There are three prime MF_x complexes for the α₃β₃ assembly at resolutions from ≥2.0 Å (PDB: **1w0j**, **1h8e**, and **1e1r**). These have been starting points for multiple computational studies, of which the majority are concerned with energetics of the chemical step, coupling between the subunits and the rotor, and rotational behavior of the synthetic complex [102, 105–107]. More recent studies involve the juxtaposition of several structures from the PDB. One of these builds a visual comparative structural approach that emphasizes pivotal roles for Mg²⁺ and protein P-loop residues in synthesizing ATP (159–163). It uses four structures, including the ATPase·Mg·ADP·AlF₄⁻ complex (PDB: **1h8e**; Fig. 20) [108].

7.8.3 Phosphoryl Transfer in Kinases

The catalytic subunit of cAPK is a serine/threonine kinase responsible for many of the effects of cAMP signaling. It is a prototype for the kinase family that uses two catalytic magnesiums, and has become the most widely studied of all kinases. Many computations have focused on the phosphorylation of a serine in the target peptide by ATP, but recent advances in high-resolution structures of an NAC complex with β,γ-imino-ATP and the products from its slow reaction during crystallization combined with an MgADP·AlF₃⁰ TSA, (PDB: **1r10**) [109] have given new opportunities for computational analysis. One of these, using MP2/aug-cc-pVTZ/CHARMM/B3LYP/6-31 + G(d)/CHARMM electronic structure calculations with

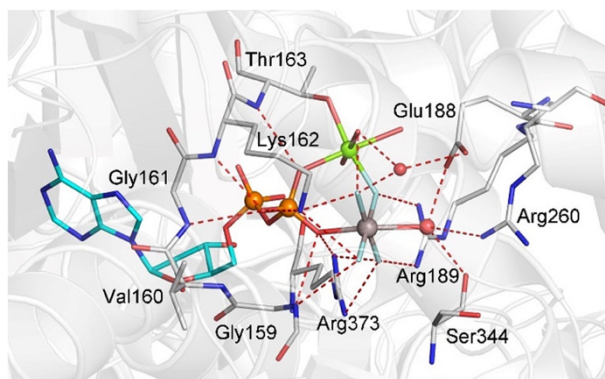


Fig. 20 FIATPase-Mg-ADP- AlF_4^- TSA complex for ATP hydrolysis (PDB: **1h8e**) showing in-line water attack on aluminum (*red sphere*) (color: phosphorus, *orange*; amino acids, *gray*; adenosine, *cyan*; waters, *red sphere*)

a completely solvated model of the $\text{cAPK}_{\text{cat}}\text{-ATPMg}_2\text{-SP20}$ system finds that a dissociative concerted mechanism involving two consecutive steps is more favorable than an associative mechanism [110, 111] or a concerted loose mechanism [112] (Scheme 1b). In step 1, phosphoryl transfer involves a dissociative TS with an $\text{O-P}^{\text{G}}\text{-O}$ distance of 4.7 Å. Then, step 2 follows with back-protonation of the serine phosphate.

The range of analyses to be found in such computations has to be set against recent structural work. The superposition of well-resolved complexes of cAPK with reactant, $\text{MgADP}\cdot\text{MgF}_3^-$ TSA analog (PDB: **113r**), and a product complex (PDB:**1rdq**) [113] shows structurally that the overall reaction is defined by the geometry of active site residues and involves migration of phosphorus only 1.1 Å from start to finish with $\text{O}_\text{A}\text{-O}_\text{D}$ separation ~ 4.5 Å in the TS (Fig. 21). Other kindred analyses have shown similar results [26].

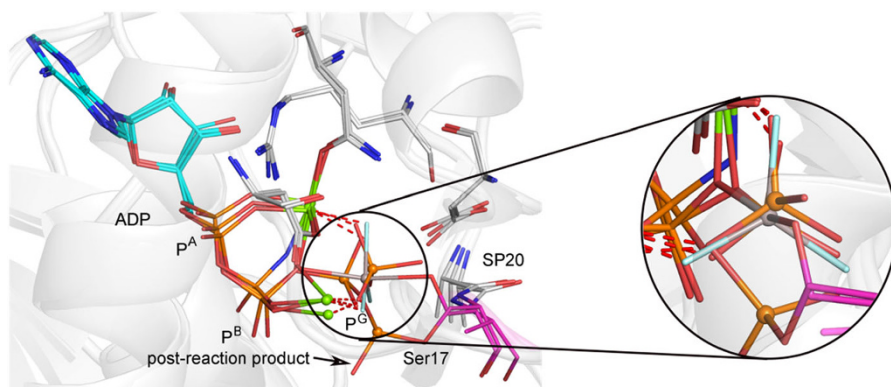


Fig. 21 Superposition of the active site of cAPK for 55% reactant and 45% product (PDB: **1rdq**, *red sticks*), TSA (PDB: **1r3l**, *gray sticks*), and product (PDB: **4hpt**) after conformational change post-reaction. Alignment of invariant amino acids (*silver*) and adenosine residues shows the high quality of fit (aligned for all $\text{C}\alpha$). The *inset* shows an orthogonal view of the threefold overlay of $\text{P}^{\text{G}}\text{O}_3$ and its mimic

7.9 Thoughts from Computations

The first phase of computational studies for phosphoryl transfer was largely focused on finding a close match between computed activation energies and the experimental ones. However, recent advances in computational methodology have cast a shadow on earlier methods, where the energy error might easily lie in the range of 2–10 kcal mol⁻¹, with error spreading as large as 30 kcal mol⁻¹ [78, 114, 115]. Current protocols enable a much larger number of heavy atoms to be embraced in the QM zone, leading to computer results that hinge on geometry of the TS and the H-bond network that it embraces [44, 108, 111]. Such analyses have identified the propensity of nucleotide analogs, particularly β,γ -iminoATP, β,γ -methyleneATP, and their GTP counterparts, to deliver NACs for phosphoryl transfer processes, which is now recognized in recent computational studies as capable of generating small but highly significant conformational changes in kinases and GTPases [44, 108]. Lastly, the belief that enzymes work by optimizing reaction mechanisms that work slowly in solution, as Knowles put it “Not different, Just better” [116], is proving to be wide of the mark for phosphate reactions. There is growing evidence that phosphoryl transfer takes place in a desolvated environment to enable full protein control of the catalytic region. Water is rigorously excluded to avoid disruption of H-bond networks that are *essential* for the organization of catalysis.

8 Conclusions

Trifluoroberyllate, tetrafluoroaluminate, and trifluoromagnesate are the primary anionic MF_x species that can mimic the phosphoryl group. Structural, spectroscopic, and computational methods have combined to validate their use as surrogates for PO₃⁻ in ground state and transition state analog complexes for many enzymes. Their use has delivered details of phosphoryl transfer at atomic resolution and supported investigations of protein folding and aggregation for tertiary structure problems. In particular, their analysis has confirmed existing concepts, introduced new ideas, and set new goals, of which the following comprise a brief summary:

- In-line stereochemistry and concertedness for S_N2(P) reactions has been established at atomic resolution;
- Relative priority of charge over geometry in transition state organization is well supported;
- Subtle conformational differences between NAC and TS conformations of amino acid functions are increasingly apparent;
- The role of H-bond networks to give structural coherence to proteins in transition states is burgeoning;
- The propensity of anionic phosphate oxygens to H-bond to ROH nucleophiles explains the need for solvent exclusion from the TS, while its impedance offers a new interpretation of general base catalysis.

Acknowledgments The authors thank Professors Nigel Richards (Cardiff University) and Jon Waltho (Manchester University) for critical comments on and many contributions to this work and Dr. Christian Roth (University of York) for advice on interpretation of some protein structures. We were supported by BBSRC Grant BB/M021637/1 and the Universities of York and Sheffield, UK. Y.J. is funded by ERC Advanced Grant AdG-322942. Multiple structural figures have used data taken from the Protein Data Bank. The use of illustrations from Refs. [27] and [44] has been appreciatively acknowledged where appropriate.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Todd AR (1981) Where there's life, there's phosphorus. In: Kageyama M, Nakamura K, Oshima T (eds) Japan Science Society Press, Tokyo, pp 275–279
2. Westheimer F (1987) *Science* 235:1173–1178
3. Lad C, Williams NH, Wolfenden R (2003) *Proc Natl Acad Sci USA* 100:5607–5610
4. Cohn M (1953) *J Biol Chem* 201:735–750
5. Knowles JR (1980) *Annu Rev Biochem* 49:877–919
6. Frey PA (1989) Chiral phosphorothioates: stereochemical analysis of enzymatic substitution at phosphorus. *Adv Enzymol Relat Areas Mol Biol* 62:119–201
7. Lowe G (1983) *Acc Chem Res* 16:244–251
8. Blackburn GM, Cherfils J, Moss PR, Nigel GJ, Waltho JP, Williams NH, Wittinghofer A (2017) *Pure App Chem* 89 (In press)
9. Todd SA (1959) *Proc Natl Acad Sci USA* 45:1389–1397
10. Mildvan AS (1997) *Proteins* 29:401–416
11. Cleland WW, Hengge AC (2006) *Chem Rev* 106:3252–3278
12. Lassila JK, Zalatan JG, Herschlag D (2011) *Annu Rev Biochem* 80:669–702
13. Smith JD, Traut RR, Blackburn GM, Monro RE (1965) *J Mol Biol* 13:617–628
14. Grisham CM, Mildvan AS (1974) *J Biol Chem* 249:3187–3197
15. Rao BD, Buttlair DH, Cohn M (1976) *J Biol Chem* 251:6981–6986
16. Milburn MV, Tong L, Devos AM, Brunger A, Yamaizumi Z, Nishimura S, Kim SH (1990) *Science* 247:939–945
17. Pai EF, Krenkel U, Petsko GA, Goody RS, Kabsch W, Wittinghofer A (1990) *EMBO J* 9:2351–2359
18. Coleman DE, Berghuis AM, Lee E, Linder ME, Gilman AG, Sprang SR (1994) *Science* 265:1405–1412
19. Sondek J, Lambricht DG, Noel JP, Hamm HE, Sigler PB (1994) *Nature* 372:276–279
20. Fisher AJ, Smith CA, Thoden JB, Smith R, Sutoh K, Holden HM, Rayment I (1995) *Biochemistry* 34:8960–8972
21. Xu Y-W, Morera S, Janin J, Cherfils J (1997) *Proc Natl Acad Sci USA* 94:3579–3583
22. Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, Schmitz F, Wittinghofer A (1997) *Science* 277:333–338
23. Schlichting I, Reinstein J (1997) *Biochemistry* 36:9290–9296
24. Graham DL, Lowe PN, Grime GW, Marsh M, Rittinger K, Smerdon SJ, Gamblin SJ, Eccleston JF (2002) *Chem Biol* 9:375–381
25. Mesmer RE, Baes CF (1969) *Inorg Chem* 8:618–626
26. Dai J, Finci L, Zhang C, Lahiri S, Zhang G, Peisach E, Allen KN, Dunaway-Mariano D (2009) *Biochemistry* 48:1984–1995
27. Blackburn GM, Jin Y, Richards NG, Waltho JP (2016) *Angew Chem Int Ed*. doi:10.1002/anie.201606474

28. Burgos ES, Ho MC, Almo SC, Schramm VL (2009) *Proc Natl Acad Sci USA* 106:13748–13753
29. Blackburn GM (1981) Phosphonates as analogues of biological phosphates. *Chem Ind (London)* 7:134–138
30. Pauling L (1960) *The nature of the chemical bond*, E3, vol E3. Cornell University Press, New York
31. Kowalinski E, Schuller A, Green R, Conti E (2015) *Structure* 23:1336–1343
32. Park AK, Lee JH, Chi YM, Park H (2016) *Biochem Biophys Res Commun* 473:625–629
33. Sheftic SR, White E, Gage DJ, Alexandrescu AT (2014) *Biochemistry* 53:311–322
34. Maruta S, Uyehara Y, Aihara T, Katayama E (2004) *J Biochem* 136:57–64
35. Hilbert BJ, Hayes JA, Stone NP, Duffy CM, Sankaran B, Kelch BA (2015) *Proc Natl Acad Sci USA* 112:E3792–E3799
36. Pylypenko O, Attanda W, Gauquelin C, Lahmani M, Coulibaly D, Baron B, Hoos S, Titus MA, England P, Houdusse AM (2013) *Proc Natl Acad Sci USA* 110:20443–20448
37. Griffin JL, Bowler MW, Baxter NJ, Leigh KN, Dannatt HRW, Hounslow AM, Blackburn GM, Webster CE, Cliff MJ, Waltho JP (2012) *Proc Natl Acad Sci USA* 109:6910–6915
38. Bruce Martin R (1988) *Biochem Biophys Res Commun* 155:1194–1200
39. Bruce Martin R (1996) *Coord Chem Rev* 149:23–32
40. Sternweis PC, Gilman AG (1982) *Proc Natl Acad Sci USA* 79:4888–4891
41. Higashijima T, Graziano MP, Suga H, Kainosho M, Gilman AG (1991) *J Biol Chem* 266:3396–3401
42. Bigay J, Deterre P, Pfister C, Chabre M (1987) *The EMBO J* 6:2907–2913
43. Wang W, Cho HS, Kim R, Jancarik J, Yokota H, Nguyen HH, Grigoriev IV, Wemmer DE, Kim SH (2002) *J Mol Biol* 319:421–431
44. Jin Y, Molt RW, Waltho JP, Richards NGJ, Blackburn GM (2016) *Angew Chem Int Ed* 55:3318–3322
45. Cliff MJ, Bowler MW, Varga A, Marston JP, Szabo J, Hounslow AM, Baxter NJ, Blackburn GM, Vas M, Waltho JP (2010) *J Am Chem Soc* 132:6507–6516
46. Hemsworth Glyn R, González-Pacanowska D, Wilson Keith S (2013) *Biochem J* 456:81–88
47. Fovet Y, Gal J-Y (2000) *Talanta* 53:617–626
48. Graham DL, Eccleston JF, Chung CW, Lowe PN (1999) *Biochemistry* 38:14981–14987
49. Baxter NJ, Olguin LF, Golcnik M, Feng G, Hounslow AM, Bermel W, Blackburn GM, Hollfelder F, Waltho JP, Williams NH (2006) *Proc Natl Acad Sci USA* 103:14732–14737
50. Jin Y, Cliff MJ, Baxter NJ, Dannatt HRW, Hounslow AM, Bowler MW, Blackburn GM, Waltho JP (2012) *Angew Chem Int Ed* 51:12242–12245
51. Baxter NJ, Blackburn GM, Marston JP, Hounslow AM, Cliff MJ, Bermel W, Williams NH, Hollfelder F, Wemmer DE, Waltho JP (2008) *J Am Chem Soc* 130:3952–3958
52. Schlichting I, Reinstein J (1999) *Nat Struct Biol* 6:721–723
53. Baxter NJ, Hounslow AM, Bowler MW, Williams NH, Blackburn GM, Waltho JP (2009) *J Am Chem Soc* 131:16334–16335
54. Bock CW, Kaufman A, Glusker JP (1994) *Inorg Chem* 33:419–427
55. Toyoshima C, Nomura H, Tsuda T (2004) *Nature* 432:361–368
56. Shinoda T, Ogawa H, Cornelius F, Toyoshima C (2009) *Nature* 459:446–450
57. Thorsell A-G, Persson C, Gräslund S, Hammarström M, Busam RD, Hallberg BM (2009) *Proteins. Struct Func Bioinfo* 77:242–246
58. Baur WH (1956) *Acta Crystallogr A* 9:515–520
59. Xiaoxia L, Marston JP, Baxter NJ, Hounslow AM, Yufen Z, Blackburn GM, Cliff MJ, Waltho JP (2011) *J Am Chem Soc* 133:3989–3994
60. Jin Y, Bhattasali D, Pellegrini E, Forget SM, Baxter NJ, Cliff MJ, Bowler MW, Jakeman DL, Blackburn GM, Waltho JP (2014) *Proc Natl Acad Sci USA* 111:12384–12389
61. Leigh KN, Webster CE (2014) *Dalton Trans* 43:3039–3043
62. Baxter NJ, Bowler MW, Alizadeh T, Cliff MJ, Hounslow AM, Wu B, Berkowitz DB, Williams NH, Blackburn GM, Waltho JP (2010) *Proc Natl Acad Sci USA* 107:4555–4560
63. Oldfield E (2005) *Phil Trans R Soc B* 360:1347–1361
64. Soñnicki JG, Langaard M, Hansen PE (2007) *J Org Chem* 72:4108–4116
65. Kohn W, Sham LJ (1965) *Phys Rev* 140:A1133–A1138
66. Hohenberg P, Kohn W (1964) *Phys Rev* 136:B864–B871
67. Cramer C (2008) *Essentials of computational chemistry*, 2nd edn. Wiley, West Sussex, pp 249–301
68. Martin RM (2004) *Electronic structure*, 1st edn. Cambridge University Press, Cambridge, pp 119–184

69. Parr RG, Tao W (1994) Density-functional theory of atoms and molecules, 1st edn. Oxford University Press, Oxford
70. Åqvist J, Warshel A (1993) *Chem Rev* 93:2523–2544
71. Hwang JK, King G, Creighton S, Warshel A (1988) *J Am Chem Soc* 110:5297–5311
72. Bennie SJ, van der Kamp MW, Pennifold RCR, Stella M, Manby FR, Mulholland AJ (2016) *J Chem Theory Comput* 12:2689–2697
73. Bartlett RJ, Shavitt I (2009) Many-body methods in chemistry and physics. Cambridge University Press, Cambridge, pp 251–340
74. iii GDP, Bartlett RJ (1982) *J Chem Phys* 76:1910–1918
75. Mlýnský V, Banáš P, Šponer J, van der Kamp MW, Mulholland AJ, Otyepka M (2014) *J Chem Theory Comput* 10:1608–1622
76. Peverati R, Truhlar DG (2011) *J Phys Chem Lett* 2:2810–2817
77. Zhao Y, Truhlar DG (2008) *Theor Chem Acc* 120:215–241
78. Goerigk L, Grimme S (2011) *Phys Chem Chem Phys* 13:6670–6688
79. Grimme S (2004) *J Comput Chem* 25:1463–1473
80. Grimme S, Antony J, Ehrlich S, Krieg H (2010) *J Chem Phys* 132:154104
81. Moellmann J, Grimme S (2014) *J Phys Chem C* 118:7615–7621
82. Matute RA, Yoon H, Warshel A (2016) *Proteins. Struct Funct Bioinf* 84:1644–1657
83. Lin P, Batra VK, Pedersen LC, Beard WA, Wilson SH, Pedersen LG (2007) *Proc Natl Acad Sci USA* 105:5670–5674
84. Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN (2003) *Science* 299:2067–2071
85. Webster CE (2004) *J Am Chem Soc* 126:6840–6841
86. Marcos E, Field MJ (2010) *Crehuet R* 78:2405–2411
87. Berente I, Beke T, Náray-Szabó G (2007) *Theor Chem Acc* 118:129–134
88. Ribeiro AJM, Ramos MJ, Fernandes PA, Russo N (2013) *Chem Phys Lett* 571:66–70
89. Florián J, Warshel A (1998) *J Phys Chem B* 102:719–734
90. Kamerlin SCL, Florián J, Warshel A (2008) *Chem Phys Chem* 9:1767–1773
91. Klähn M, Rosta E, Warshel A (2006) *J Am Chem Soc* 128:15310–15323
92. Rp B (2013) Plotnikov NV, Lameira J, Warshel A. *Proc Natl Acad Sci USA* 110:20509–20514
93. Grigorenko BL, Nemukhin AV, Topol IA, Cachau RE, Burt SK (2005) *Proteins. Struct Funct Bioinf* 60:495–503
94. Khrenova MG, Grigorenko BL, Kolomeisky AB, Nemukhin AV (2015) *J Phys Chem B* 119:12838–12845
95. Mironov VA, Khrenova MG, Lychko LA, Nemukhin AV (2015) *Proteins. Struct Funct Bioinf* 83:1046–1053
96. te Heesen H, Gerwert K, Schlitter J (2007) *FEBS Lett* 581:5677–5684
97. Khrenova MG, Grigorenko BL, Mironov VA, Nemukhin AV (2015) *Proteins. Struct Funct Bioinf* 83:2091–2099
98. Rudack T, Xia F, Schlitter J, Köttling C, Gerwert K (2012) *Proc Natl Acad Sci USA* 109:15295–15300
99. Khrenova MG, Mironov VA, Grigorenko BL, Nemukhin AV (2014) *Biochemistry* 53:7093–7099
100. Topol IA, Cachau RE, Nemukhin AV, Grigorenko BL, Burt SK (2004) *Biochim Biophys Acta* 1700:125–136
101. Shurki A, Warshel A (2004) *Proteins. Struct Funct Bioinf* 55:1–10
102. Kamerlin SC, Sharma PK, Prasad RB, Warshel A (2013) *Q Rev Biophys* 46:1–132
103. Tripathi R, Graves R, Marx D (2017) *Chem Sci* 8:371–380
104. Kiani FA, Fischer S (2014) *Proc Natl Acad Sci USA* 111:E2947–E2956
105. Martín-García F, Mendieta-Moreno JI, Marcos-Alcalde Í, Gómez-Puertas P, Mendieta J (2013) *Biochemistry* 52:959–966
106. Ito Y, Ikeguchi M (2014) *Biophys J* 108:85–97
107. Kleinekathöfer U, Isralewitz B, Dittrich M, Schulten K (2011) *J Phys Chem A* 115:7267–7274
108. Blum DJ, Ko YH, Pedersen PL (2012) *Biochemistry* 51:1532–1546
109. Bastidas AC, Wu J, Taylor SS (2015) *Biochemistry* 54:2–10
110. Pérez-Gallegos A, Garcia-Viloca M, González-Lafont À, Lluch JM (2014) *J Comput Aided Mol Des* 28:1077–1091
111. Pérez-Gallegos A, Garcia-Viloca M, González-Lafont À, Lluch JM (2015) *ACS Catal* 5:4897–4912
112. Perez-Gallegos A, Garcia-Viloca M, Gonzalez-Lafont A, Lluch JM (2015) *Phys Chem Chem Phys* 17:3497–3511

113. Gerlits O, Tian J, Das A, Langan P, Heller WT, Kovalevsky A (2015) *J Biol Chem* 290:15538–15548
114. Grimme S (2006) *J Comput Chem* 27:1787–1799
115. Molt RW, Watson T, Bazante AP, Bartlett RJ, Richards NGJ (2016) *Phys Chem Chem Phys* 18:26069–26077
116. Knowles JR (1991) *Nature* 350:121–124

Importance of Radioactive Labelling to Elucidate Inositol Polyphosphate Signalling

Miranda S. C. Wilson¹ · Adolfo Saiardi¹

Received: 26 October 2016 / Accepted: 21 December 2016 / Published online: 18 January 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract Inositol polyphosphates, in their water-soluble or lipid-bound forms, represent a large and multifaceted family of signalling molecules. Some inositol polyphosphates are well recognised as defining important signal transduction pathways, as in the case of the calcium release factor $\text{Ins}(1,4,5)\text{P}_3$, generated by receptor activation-induced hydrolysis of the lipid $\text{PtdIns}(4,5)\text{P}_2$ by phospholipase C. The birth of inositol polyphosphate research would not have occurred without the use of radioactive phosphate tracers that enabled the discovery of the “PI response”. Radioactive labels, mainly of phosphorus but also carbon and hydrogen (tritium), have been instrumental in the development of this research field and the establishment of the inositol polyphosphates as one of the most important networks of regulatory molecules present in eukaryotic cells. Advancements in microscopy and mass spectrometry and the development of colorimetric assays have facilitated inositol polyphosphate research, but have not eliminated the need for radioactive experimental approaches. In fact, such experiments have become easier with the cloning of the inositol polyphosphate kinases, enabling the systematic labelling of specific positions of the inositol ring with radioactive phosphate. This approach has been valuable for elucidating their metabolic pathways and identifying specific and novel functions for inositol polyphosphates. For example, the synthesis of radio-labelled inositol pyrophosphates has allowed the discovery of a new protein post-translational modification. Therefore, radioactive tracers have played and will continue to play an important role in dissecting the many complex aspects of

This article is part of the Topical Collection “Phosphate Labelling in Chemical Biology”; edited by Henning Jessen.

✉ Adolfo Saiardi
dmcbedo@ucl.ac.uk

¹ Medical Research Council Laboratory for Molecular Cell Biology, University College London, Gower Street, London WC1E 6BT, UK

inositol polyphosphate physiology. In this review we aim to highlight the historical importance of radioactivity in inositol polyphosphate research, as well as its modern usage.

Keywords Radioactivity · Inositol · Pyrophosphates · Metabolism · Phosphate

1 Introduction

Inositol polyphosphates comprise a vast and multifaceted family of cellular metabolites. The size of the family is explained by the ability to combinatorially substitute the six hydroxyls of the *myo*-inositol ring with phosphate moieties: mathematically, 64 such combinations are possible [1]. This number is in fact an underestimate, as diphosphate (or pyrophosphate) moieties also exist [2–4]. Among these myriad inositol polyphosphates, without doubt the most famous is the calcium release factor $\text{Ins}(1,4,5)\text{P}_3$, the prototypical second messenger. Hydrolysis of the lipid bond of $\text{PtdIns}(4,5)\text{P}_2$ by phospholipase C (PLC) following receptor activation to release the water-soluble $\text{Ins}(1,4,5)\text{P}_3$ and the lipid diacylglycerol (DAG) is a textbook example of signal transduction (Fig. 1) [5, 6]. It results in $\text{Ins}(1,4,5)\text{P}_3$ binding to the InsP_3 receptor and the consequent release of calcium from intracellular stores [7].

This PLC activity was indirectly assayed during the first discovery of receptor-stimulated inositol polyphosphate metabolism. In the early 1950s, the Hokin husband and wife team were studying RNA metabolism using radioactive orthophosphate [^{32}P] metabolic labelling, when they discovered an increase in cellular incorporation of radioactivity when pancreatic slices were stimulated with acetylcholine. Surprisingly, however, the large majority of radioactivity was not incorporated into nucleic acids but into the inositol-containing lipids called phosphoinositides (PI or PtdInsP , phosphatidylinositols); this acetylcholine-stimulated [^{32}P] incorporation was therefore termed the “PI response” [8, 9]. It was not until three decades later that the PI response was determined to be part of PLC activation [6, 10] (Fig. 1). The [^{32}P] taken up is initially incorporated into ATP and other nucleotides. Radioactive phosphorylation of DAG to phosphatidic acid (PA) allows its reattachment to inositol, creating PtdIns that is further phosphorylated, generating the radioactive PtdInsP / “PI” that the Hokins observed, and completing what is now known as the inositol cycle (Fig. 1).

This short historical background highlights the fundamental importance of radioactive phosphate labelling in the birth of the phosphoinositide and inositol polyphosphate signalling research fields. Equally, everyone is aware of the historical—and indeed, current—importance of radioactive labelling in nucleic acid research [11]. Without radioactivity-based methods, molecular biology would not have emerged and we would not be in the post-genomic era of biomedical research. It is not an overstatement to say that without radioactivity, the advances in biomedical science and consequent improvements in human health of the past century would not have been achieved. The ongoing importance of radioactive labelling to inositol polyphosphate research should also not be underestimated.

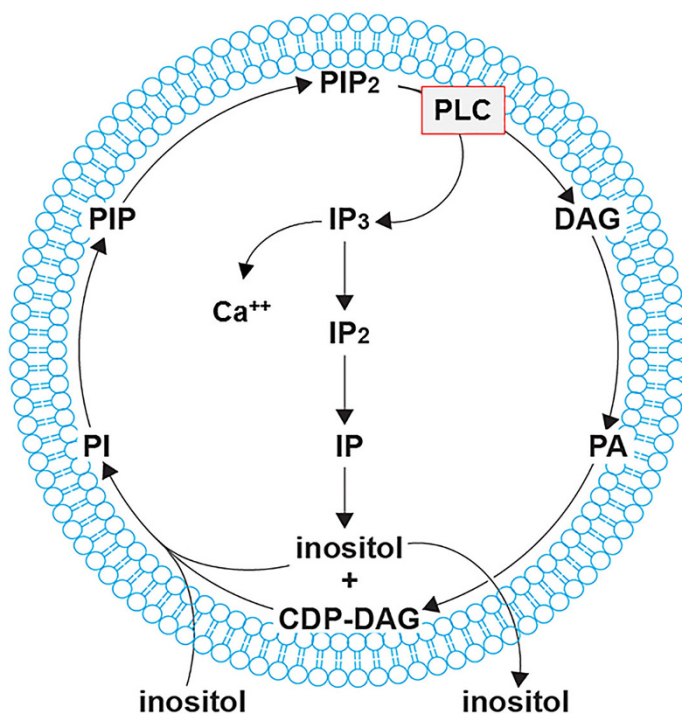


Fig. 1 Schematic representation of the inositol cycle. Inositol acquired from the extracellular space is incorporated into lipids by the action of the phosphatidylinositol synthase (PI-synthase). The conversion of PtdIns (PI), first to PtdIns(4)P (PIP) and then to PtdIns(4,5)P₂ (PIP₂), generates the substrate for phospholipase C (PLC, boxed in red). Once receptor activation occurs, PLC generates two second messengers: the (plasma) membrane-resident diacylglycerol (DAG) and the calcium (Ca²⁺) release factor Ins(1,4,5)P₃ (IP₃). The latter is converted back to inositol via two dephosphorisation steps, closing the cycle. The inositol cycle is particularly active in stimulated mammalian cells. The Hokin “PI response” [8, 9] measures the [³²P] taken up by the cell and its conversion to [³²P] γATP, with the subsequent radioactive phosphorylation of DAG to phosphatidic acid (PA). PA is then reattached to inositol, creating radioactive PI. For graphical reasons, inositol is abbreviated here as “Ins”, and phosphatidylinositol as “PI” instead of “PtdIns”

The general perception of “radioactivity” is one of fear, and understandably so, considering famous disasters such as those at Chernobyl or Fukushima [12]. However, using very low trace levels of radioactivity in a highly controlled research environment is safe: many protection and monitoring measures are available [13, 14]. Young researchers should therefore embrace radioactivity-based techniques, as they offer unique research opportunities, even in this twenty-first century, and especially in the inositol polyphosphate research field. Inositol polyphosphates, being unable to absorb UV/visible wavelength light, cannot be detected by absorbance or fluorescence methods. Thus the study of their metabolism and physiological functions has only been possible through radioactivity experiments. The current essay will focus primarily on the importance of radioactive phosphate labelling ([³²P] and [³³P]) in inositol polyphosphate biology. We will discuss the use of radioactive orthophosphate tracers to study the cellular metabolism of inositol

polyphosphates, as well as the biochemical synthesis and purification of inositol polyphosphates radiolabelled at specific positions on the inositol ring. To provide a complete view of relevant radioactive methods, we will also concisely describe the use of radioactive hydrogen (tritium [^3H]) and carbon (carbon-14 or [^{14}C]) labelling techniques.

2 Phosphoinositides and Inositol Phosphates

Before discussing radioactive labelling methods in inositol polyphosphate research, we must briefly introduce these molecules and the metabolic pathways connecting them. As it is not the main scope of this essay, discussion of inositol polyphosphate metabolism will be highly simplified; the interested person is encouraged to read the following more comprehensive reviews [15–18]. The carbon backbone of *myo*-inositol (hereafter simply called “inositol”) is by far the most common and biologically relevant of the naturally occurring stereoisomers. It is also the structural building block for the inositol polyphosphates. In its favoured chair conformation, inositol has five equatorial and one axial hydroxyl group (Fig. 2a) [19]. This axial hydroxyl is found at the carbon in position 2, using the D-numbering convention for cyclitols.

The simplest function of inositol is as an osmolyte, whose cellular concentration is regulated in response to hyperosmolarity. However, more interesting functions are achieved through phosphorylation to create inositol polyphosphates. These water-soluble molecules have a complicated biosynthetic pathway in yeast and, presumably, in mammalian cells: not just sequential phosphorylation or dephosphorylation, but synthesis that is intimately linked to the metabolism of the related PtdIns lipids (Figs. 1, 3) [20, 21]. Cells can synthesise inositol *de novo* from glucose-6-phosphate or acquire it from the extracellular environment, allowing it to enter the inositol cycle (Fig. 1). Inositol lipid synthesis starts with the activation of PA with CTP, becoming GDP-DAG that is subsequently attached to the 1-hydroxyl of the inositol ring, forming PtdIns. This can be phosphorylated to PtdIns(4)P and then to PtdIns(4,5)P₂, the substrate for phospholipase C (PLC), in the calcium release signalling paradigm described above. It is important to remember that the universal production of Ins(1,4,5)P₃ by PLC does not necessarily translate into calcium signalling, as many eukaryote clades including yeast and plants do not possess InsP₃ receptors [22]. The Ins(1,4,5)P₃ generated by PLC activity can be dephosphorylated back to inositol and reused for PtdIns synthesis. Conversely, Ins(1,4,5)P₃ can act as precursor for a large, diverse family of higher phosphorylated inositol polyphosphates (Fig. 3). For example, Ins(1,4,5)P₃ is a substrate of the inositol polyphosphate multikinase (IPMK; yeast Arg82) [23–26], which is able to phosphorylate both positions 3 and 6, creating Ins(1,3,4,5,6)P₅. This is acted on by inositol pentakisphosphate 2-kinase (IP₅-2K or IPPK; Ipk1 in yeast) [27, 28] to create the fully phosphorylated inositol hexakisphosphate (InsP₆ or phytic acid). Another metabolic route can also lead to Ins(1,3,4,5,6)P₅ synthesis. Phosphorylation of Ins(1,4,5)P₃ by ITPKA,B,C, the IP₃-3Ks, generates Ins(1,3,4,5)P₄ [29, 30] that is dephosphorylated by 5-phosphatases such as SHIP1 into a different InsP₃ isomer,

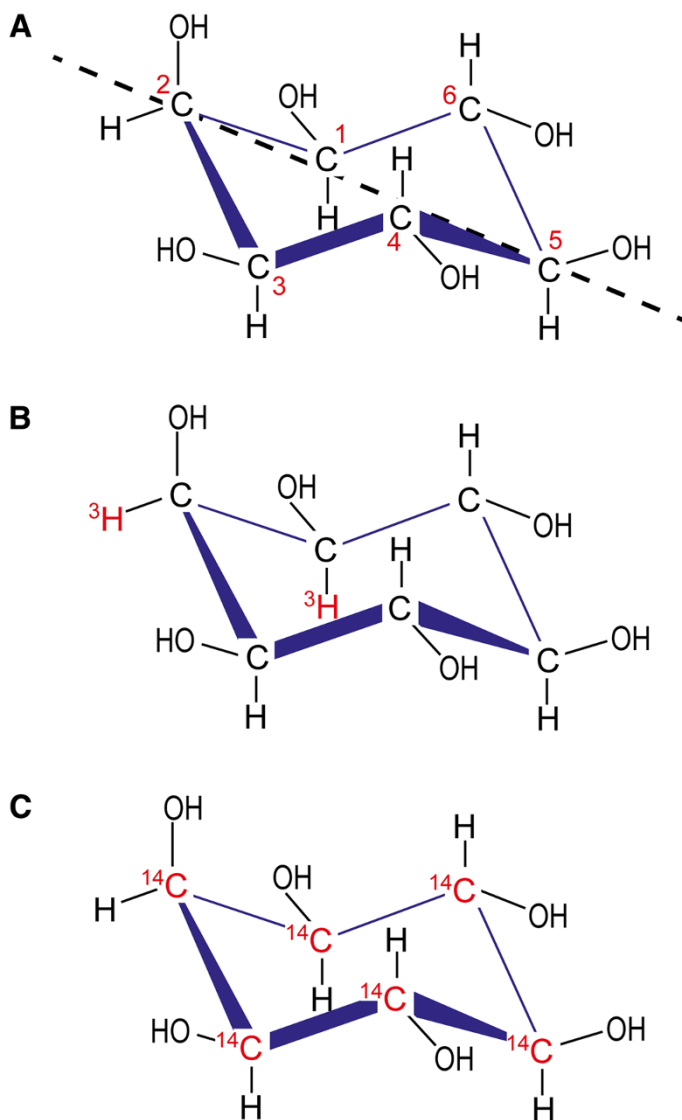
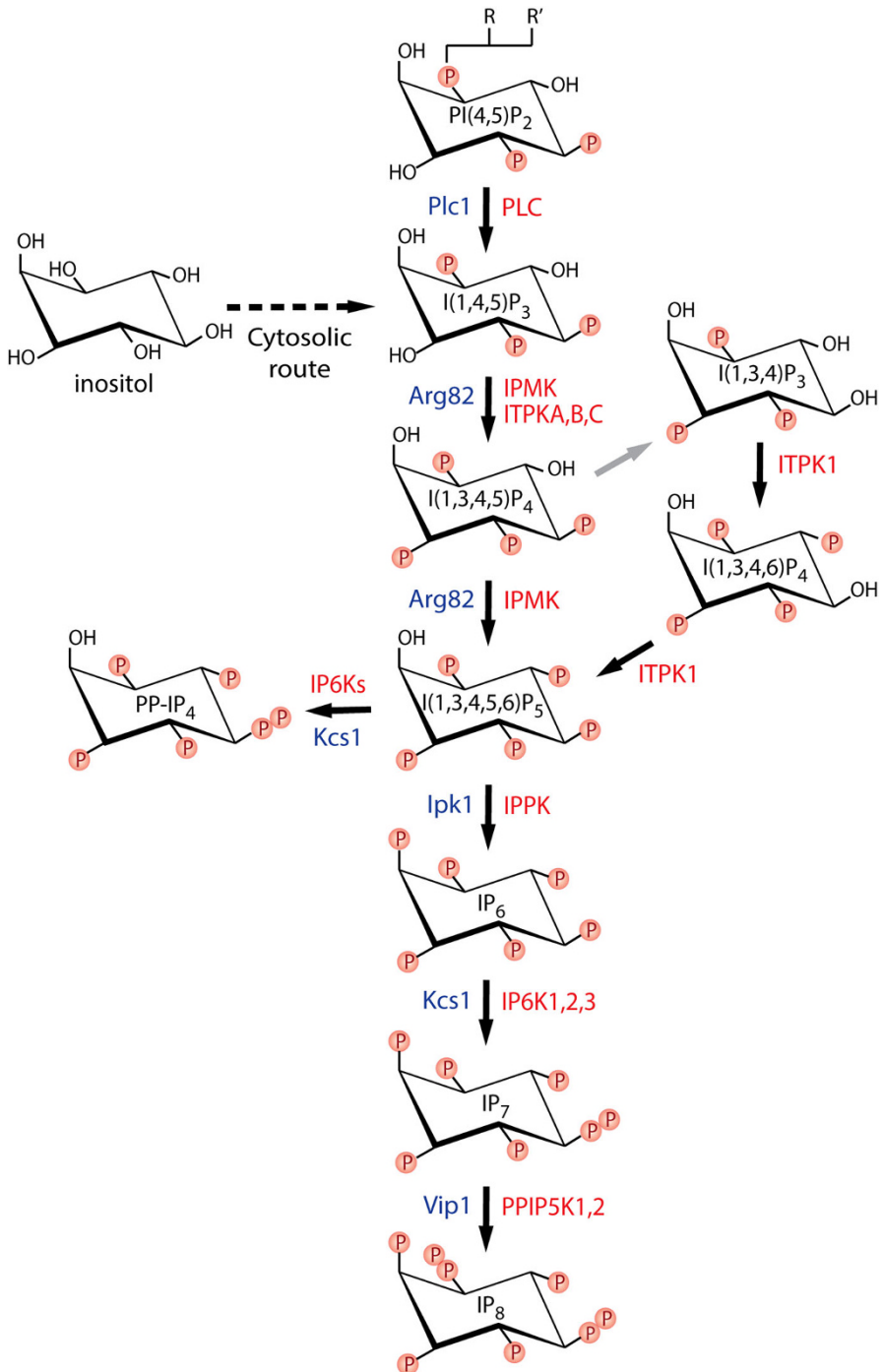


Fig. 2 *myo*-Inositol structure and its radiolabelled derivatives. While nine stereoisomeric configurations of inositol are possible, the structure of *myo*-inositol is depicted in (a), referred to in the review simply as inositol, since it is by far the most common and biologically relevant form of inositol. The modern D-numbering system for inositols is counterclockwise as viewed from above and assigns the single axial hydroxyl group of *myo*-inositol to the carbon in position 2, while the other five hydroxyls are equatorial. *myo*-Inositol possesses an axis of symmetry through carbons 2 and 5 (dashed line), making positions 1,3 and 4,6 enantiomeric. The most common commercially available tritium-labelled inositol (b) possess the [^3H] radiolabelled in position 1 and/or 2, while in [^{14}C]inositol the radiolabel is uniformly distributed (c)



◀ **Fig. 3** Inositol polyphosphates synthetic pathway. The synthesis of higher phosphorylated inositol polyphosphates begins with the synthesis of $\text{Ins}(1,4,5)\text{P}_3$. *Saccharomyces cerevisiae* uses only phospholipase C (PLC) hydrolysis of the lipid $\text{PI}(4,5)\text{P}_2$ to synthesise $\text{Ins}(1,4,5)\text{P}_3$ [28], whereas *Dictyostelium discoideum* utilises the cytosolic route, of which the enzymology is not fully elucidated (dashed line) [110]. $\text{Ins}(1,4,5)\text{P}_3$ is metabolised by ITPKA, B, or C to synthesise $\text{Ins}(1,3,4,5)\text{P}_3$, which is acted on by the 5-phosphatase (grey line) to generate the $\text{Ins}(1,3,4)\text{P}_3$ converted by ITPK1 to $\text{Ins}(1,3,4,5,6)\text{P}_5$. However, this isomer of InsP_5 can also be directly generated by IPMK from $\text{Ins}(1,4,5)\text{P}_3$. InsP_5 is converted to InsP_6 by the IP_5 -2Kinase IPPK. Phosphorylation of InsP_6 by the IP6Ks generates the inositol pyrophosphate InsP_7 , specifically the depicted isomer 5PP- InsP_5 , which is further acted on by PPIP5K1,2 to IP_8 , specifically to 1,5(PP) $_2$ - IP_4 . The IP6K enzymes can also use $\text{Ins}(1,3,4,5,6)\text{P}_5$ as a substrate, generating the inositol pyrophosphate PP- IP_4 . In this figure, for visual reasons, inositol is abbreviated as “I” instead of “Ins”, and phosphatidylinositol as “PI” instead of “PtdIns”. Kinases catalysing each step are indicated in red (human) and blue (*S. cerevisiae*)

$\text{Ins}(1,3,4)\text{P}_3$. This isomer is a substrate for ITPK1, another multikinase, which adds a phosphate group at positions 5 and 6, again resulting in an InsP_5 species with the remaining hydroxyl group at the 2 position [31–33] (Fig. 3). Furthermore ITPK1 can also phosphorylate position 1 of the inositol ring [34, 35].

Together, $\text{Ins}(1,3,4,5,6)\text{P}_5$ and InsP_6 are the major forms of inositol polyphosphates present in mammalian cells. Despite their association with numerous cellular functions, and unlike $\text{Ins}(1,4,5)\text{P}_3$, no dramatic or rapid changes in the cellular amount of either molecule are seen on receptor activation, although their relative level is modulated by neurotrophin signals [36]. However, they are not metabolically inert: $\text{Ins}(1,3,4,5,6)\text{P}_5$ and InsP_6 are substrates for the synthesis of a subfamily of inositol polyphosphates called inositol pyrophosphates [37–39]. These contain one or more high-energy phosphoanhydride (pyro) bonds as well as the phosphoesters. Two classes of enzyme synthesise inositol pyrophosphates in mammalian cells. The IP6Ks (InsP_6 kinase) are able to pyrophosphorylate position 5 of the inositol ring, generating 5PP- InsP_5 (diphosphoinositol pentakisphosphate, InsP_7) from InsP_6 or 5PP- InsP_4 from $\text{Ins}(1,3,4,5,6)\text{P}_5$ [25, 38–43]. Alternatively, PPIP5K (VIP1 in yeast) enzymes can pyrophosphorylate position 1, generating the 1PP- InsP_5 isomer of InsP_7 from InsP_6 in vitro [44–47], and InsP_8 ([PP] $_2$ - InsP_4 ; bis-diphosphoinositol-tetrakisphosphate) from 5PP- InsP_5 in vivo [48, 49].

The characterised inositol phosphate kinases (Table 1) offer the opportunity to label the inositol ring with radioactive phosphate simply by performing in vitro enzymatic reactions using recombinant enzyme with radioactive [^{32}P] γ ATP. We will first introduce the radioactive isotopes available, before describing their use.

3 Radioactive Labels for Inositol

Inositol polyphosphates consist of four different atomic species: for example, the chemical formula of the fully phosphorylated inositol ring of InsP_6 is $\text{C}_6\text{H}_{18}\text{O}_{24}\text{P}_6$. Of these elements, oxygen does not have relevant long-lived radioisotopes and therefore is not used for radiolabelling. Conversely, it is possible to generate useful unstable, and thus radioactive, isotopic species for carbon, hydrogen and phosphorus. We will briefly describe the biophysical characteristics of the most commonly

Table 1 Inositol phosphate kinases identified in the human and yeast (*S. cerevisiae*) genomes

Inositol polyphosphate kinase	Human	Yeast	Main enzymatic activities	References
Inositol-trisphosphate 3-kinase	ITPKA,B,C	–	Ins(1,4,5)P ₃ => Ins(1,3,4,5)P ₄	[74, 105–107]
Inositol polyphosphate multikinase	IPMK	Arg82	Ins(1,4,5)P ₃ => Ins(1,3,4,5)P ₄ Ins(1,3,4,5)P ₄ => Ins(1,3,4,5,6)P ₅ PtdIns(4,5)P ₂ => PtdIns(3,4,5)P ₂	[23–26]
Inositol-tetrakisphosphate 1-kinase	ITPK1	–	Ins(1,3,4)P ₃ => Ins(1,3,4,5)P ₄ Ins(1,3,4,5)P ₄ => Ins(1,3,4,5,6)P ₅ Ins(3,4,5,6)P ₄ => Ins(1,3,4,5,6)P ₅	[32, 33, 35, 85, 108]
Inositol pentakisphosphate 2-kinase	IPPK	Ipk1	Ins(1,3,4,5,6)P ₅ => InsP ₆	[27, 28]
Inositol hexakisphosphate kinase	IP6K1,2,3	Kcs1	InsP ₆ => 5PP-InsP ₅ 1PP-InsP ₅ => (1,5)PP ₂ -InsP ₄ Ins(1,3,4,5,6)P ₅ => PP-InsP ₃	[25, 42, 83]
Diphosphoinositol pentakisphosphate kinase	PPIP5K1,2	Vip1	InsP ₆ => 1PP-InsP ₅ 5PP-InsP ₅ => (1,5)PP ₂ -InsP ₄	[44–47, 109]

Table 2 Isotopic labels available to generate inositol polyphosphate radioactive tracers

	Symbol	β energy	Range in air	Half-life
Tritium, hydrogen-3	[³ H]	5.7 keV	<1 cm	12.3 years
Carbon-14	[¹⁴ C]	156 keV	24 cm	5730 years
Phosphorus-32	[³² P]	1709 keV	610 cm	13.5 days
Phosphorus-33	[³³ P]	249 keV	89 cm	25.4 days

used radioisotopes: tritium [³H], carbon [¹⁴C], and phosphorus [³²P] and [³³P] (Table 2). All are radioactive due to β decay.

Tritium contains two neutrons as well as the usual hydrogen proton. It is a weak β-emitter with a half-life of over 12 years. The emitted electron has very low energy, and can therefore travel only a few millimetres in air, and cannot penetrate the skin. This makes it particularly easy to work with, as special protective equipment is not required: gloves, goggles and lab coat are sufficient. On the minus side, the emitted radiation is too low to be detected by Geiger counter. Testing for tritium contamination before and after an experiment must instead be performed by swipe testing and liquid scintillation counting. Carbon-14 has a very long half-life, enabling its use in carbon dating in archaeology. The emitted electrons are still reasonably low-energy, although they can be detected using a Geiger counter, and

can travel less than 0.3 mm into the skin. Therefore, shielding is also not necessary when working with [^{14}C].

Phosphorus-32 is the most energetic radioisotope commonly used in biochemical laboratories. The electrons released are very high-energy and can travel over 6 m through air and 0.76 cm into human tissues. Consequently, extra precautions are required when working with [^{32}P]: a 1 cm Plexiglas shield is required for the working area, and dosimeter monitoring for the experimenter. Contamination is easily detected with a Geiger counter. Experiments requiring [^{32}P] can also be performed using the superior phosphorus-33 isotope. The emitted radiation from [^{33}P] is of lower energy and is therefore less dangerous, although the same safety measures are required. The half-life of [^{33}P] is also longer than that of [^{32}P]: 25.4 days compared to 14.3 days. This can be extremely beneficial in maximising the value of any radiolabelled compounds synthesised. Only very rare experimental circumstances require the high energy of [^{32}P] radiation, but this isotope is still widely used, since there is a huge price difference between the two radioactive isotopes. Phosphorus-33 is much more expensive.

4 Inositol Polyphosphate in Vivo Studies Using Radioactive Metabolic Labelling

As mentioned above, the main problem for inositol polyphosphate research is that there is no easy way to detect the inositol ring using spectrophotometry. A technique was developed in the 1980s that combined chromatography, post-column derivatisation and spectrophotometry to visualise the phosphate groups, and thus indirectly the inositol polyphosphate, to try to solve this problem [50]. This method is not sensitive enough to be routinely used with mammalian cells, although there are a few reports of its use [46]; it may be more appropriate for organisms with high levels of inositol polyphosphates such as the amoeba *Dictyostelium discoideum* [3, 51]. In general, for effective, sensitive and reliable methods for detecting and studying the metabolism and many functions of inositol polyphosphates, we must turn to radioactivity.

The standard procedures for investigating inositol polyphosphate metabolism in vivo require the use of radiolabelled inositol tracers [52, 53]. These are usually based on tritium labels, with carbon-14 rarely used, as it is more expensive (Fig. 2b, c). These [^3H]inositol or [^{14}C]inositol tracers (Fig. 2b, c) are added to the extracellular growth medium, where they are taken up by cells and enter the inositol cycle (Figs. 1, 3). Different inositol polyphosphates and phosphoinositides then begin to be radiolabelled. The labelling must be given sufficient time to reach isotopic equilibrium, where all the inositol polyphosphates species are in equilibrium with [^3H]inositol or [^{14}C]inositol. Given the presence of more than 30 inositol polyphosphate species in eukaryote cells [15, 17], and that, for example, seven sequential enzymatic reactions are required to generate InsP_7 , it is clear that allowing time to reach isotopic equilibrium is fundamental to generating reliable experimental data. For yeast, it is normally sufficient to label the cells overnight, which corresponds to 7–8 cell divisions. Unsurprisingly, mammalian cells must be

labelled for much longer to reach equilibrium: 4 or 5 days may be enough, depending on cell type. Before starting labelling experiments in a new cell type, a pilot study must be performed to determine the time needed to reach metabolic equilibrium [36]. This is calculated by dividing the radioactivity accumulated in InsP_6 by the radioactivity of the lipid phosphoinositide pool. This InsP_6 /phosphoinositide ratio increases over time; when it remains constant, the labelling has reached equilibrium. In the modern literature, too often the isotopic equilibrium is not properly calculated or even considered, casting doubt on the reliability of the data generated.

Once metabolic equilibrium has been reached, radiolabelled inositol polyphosphates can be acid-extracted and resolved by strong anion exchange high-performance liquid chromatography (saxHPLC). Two options exist for analysing the samples. Firstly, an in-line radioactivity detector can be used, greatly speeding up analysis and reducing handling at the cost of sensitivity. The alternative is to collect fractions for manual counting with a scintillation β -counter; while more labour-intensive, this method significantly increases the sensitivity. The inositol pyrophosphate species InsP_7 and InsP_8 can be easily detected in labelled yeast extracts using manual counting [52, 54], while only the more abundant precursor InsP_6 has been identified using in-line detectors [28].

It should also be mentioned that dual isotopic labelling is possible. The difference in the energy of electrons emitted by $[^3\text{H}]$ inositol and $[^{14}\text{C}]$ inositol enables scintillation counters to distinguish these two inositol species. By labelling cells to isotopic equilibrium using $[^{14}\text{C}]$ inositol, and then briefly with $[^3\text{H}]$ inositol, it is possible to study the possible heterogeneity within pools of phosphoinositides or inositol polyphosphates [55]. This approach was employed successfully in the study of inositol polyphosphates generated after vasopressin or prostaglandin stimulation of vascular tissue. Their source was found to be rapidly labelled phosphoinositides, while the bulk of the highly phosphorylated inositol polyphosphates InsP_5 and InsP_6 were not created from the rapid phosphoinositide turnover, and were therefore deemed metabolically inert [56]. Dual labels can also have more technical uses. Chiefly, a second isotopic label is widely used to determine chromatographic saxHPLC peak identity: spiking a $[^3\text{H}]$ inositol labelled extract with a $[^{14}\text{C}]$ inositol polyphosphate standard enables conclusive identification of the nature of the eluted peaks [24].

The use of commercially available radiolabelled standards would of course be ideal. Unfortunately, many inositol polyphosphates, including InsP_7 , are not available commercially in their radiolabelled form or often even in unlabelled format. The next best option for peak identification is to use the well-characterised yeast inositol kinase mutant strains. Separation of radiolabelled wild-type yeast extracts by saxHPLC reveals a simple elution profile with one major peak of InsP_6 and two smaller, more polar and therefore later-eluting peaks of InsP_7 and InsP_8 . These two peaks are absent in extracts from *kcs1 Δ* (IP6K mutant) yeast. The *vip1 Δ* mutant instead accumulates InsP_7 [48, 49]. Other mutants show increased peaks for inositol polyphosphates: $\text{Ins}(1,4,5)\text{P}_3$ accumulates in *arg82 Δ* multikinase mutants [23], while $\text{Ins}(1,3,4,5,6)\text{P}_5$ accumulates and is converted by Kcs1 into PP-IP₄ in an *ipk1 Δ* mutant. The *ipk1 Δ kcs1 Δ* double mutant has only the increased InsP_5 peak

[54]. Radiolabelled inositol polyphosphate standards can also be generated in vitro enzymatically, whether using radioactive inositol polyphosphate precursors or cold inositol polyphosphates with [^{32}P] γ ATP (as described below). Laboratories with a serious interest in inositol polyphosphate analysis would do well to create standards with a long half-life: extracting and purifying specific inositol polyphosphates from [^3H]inositol- or [^{14}C]inositol-labelled cells gives defined standards that can be used over several years [24, 57].

While radioactive orthophosphate labelling has been instrumental in the early development [58–60] of the inositol phosphate research field, it must be admitted that [^{32}P] or [^{33}P] orthophosphate labelling currently has limited use for in vivo analysis of inositol polyphosphate metabolism. Many phosphorylated molecules in eukaryotic cells, primarily the abundant nucleotides, are co-purified during the acidic extraction normally employed to purify inositol polyphosphates. Therefore, any chromatogram is essentially undecipherable until the nucleotides have eluted off. The interfering nucleotides can be removed first by charcoal treatment, but this complicates the extraction procedure [61]. For this reason, only a handful of papers published after the 1990s have reported the use of phosphate labelling to study inositol polyphosphate metabolism in vivo. One such use was the original identification of inositol pyrophosphates in *D. discoideum* extracts [3]: the saxHPLC elution region for these high-polarity molecules in this organism is conveniently free from interference from contaminating phosphorylated molecules. This is not the case for extracts from cell types rich in the linear polymer of phosphates [62, 63] inorganic polyphosphate (polyP), such as yeast or trypanosomes, where polyP is particularly abundant [64–66]. In these organisms, orthophosphate labelling experiments will reveal the continuous presence throughout the chromatogram of radiolabelled polyP peaks that cover the InsP_6 , InsP_7 and InsP_8 signals (A. Saiardi, unpublished observation).

Orthophosphate labelling using [^{32}P] or [^{33}P] is more useful in studies of PtdIns lipids. While several types of potentially contaminating phospholipids exist, their number is not as great as the water-soluble (i.e., acid-extracted) phosphorylated molecules present in eukaryotic cytosol. Furthermore, nucleotides are eliminated by the organic solvent extraction procedure required to purify phosphoinositides. Thus it is possible to identify and study radiolabelled phosphoinositides directly, using thin-layer chromatography (TLC) or by resolution by saxHPLC after deacylation [67]. The highly energetic phosphate isotopes also allow the use of in-line radioactive detection methods to measure the saxHPLC-eluted deacylated lipids. It is important to note that orthophosphate [^{32}P] or [^{33}P] labelling is usually performed over a short period of time, from a few minutes to a few hours only. The phosphorus isotopes are toxic to the cell, and incubation over a few days induces cell stress or even death. Furthermore, orthophosphate labelling does not require reaching any isotopic equilibrium. Tritium or [^{14}C] labels are components of the inositol ring itself (Fig. 1b, c), while phosphate labels are not: they are added and removed by the dynamic action of specific phosphatases and kinases, and thus isotopic equilibrium labelling is not necessary.

5 Biochemical Synthesis of $^{32/33}\text{P}$ Phosphate Radiolabelled Inositol Polyphosphates

Even before the cloning of inositol polyphosphate kinases, allowing the synthesis and purification of recombinant enzymes from bacteria, partially purified enzymatic activities or simple cell extracts were used to synthesise radiolabelled inositol polyphosphates [68, 69]. The synthesis of these molecules, labelled with radioactive phosphate at specific positions of the ring, has had numerous applications, including as standards, for monitoring enzymatic activity, or even the discovery of the new post-translational modification protein pyrophosphorylation. Most of the metabolic pathways and functions of lower phosphorylated inositol polyphosphates were elucidated before the cloning of the kinases responsible for their synthesis, often using elegant [^{32}P] radiolabelled biochemical assays [70–73]. Studies of the functions of the higher phosphorylated inositol polyphosphates, such as InsP_5 , InsP_6 , and their inositol pyrophosphate derivatives have benefitted from this previous knowledge and by the cloning of their kinases. Using one of the six now known inositol polyphosphate kinases (Table 1) and [$^{32/33}\text{P}$] γATP , it is possible to label almost any position of the inositol ring with radioactive phosphate. We will start by discussing the use of recombinant ITPKA, an IP_3 -3 kinase, to generate InsP_4 labelled at position 3.

6 Synthesis of [^{32}P] $_i$ Radiolabelled InsP_4 and Its Use

The biosynthesis of [^{32}P] InsP_4 can be achieved by incubating ITPKA with $\text{Ins}(1,4,5)\text{P}_3$ and [^{32}P] γATP . This enzyme specifically phosphorylates position 3 of the ring, and thus generates 3[^{32}P] $\text{Ins}(1,3,4,5)\text{P}_4$ [74]. Several studies have used this method to investigate the metabolism of this isomer. Notably, its degradation by inositol polyphosphate 5-phosphatases (namely SHIP1/2) can be studied by following the formation of the radiolabelled and thus easily traceable 3[^{32}P] $\text{Ins}(1,3,4)\text{P}_5$ [75]. The anabolism of 3[^{32}P] $\text{Ins}(1,3,4,5)\text{P}_4$ and conversion to [^{32}P] InsP_6 has also been studied after incubation with nuclear extracts from *D. discoideum* [76].

SHIP2 is more famous for its ability to convert $\text{PtdIns}(3,4,5)\text{P}_3$ to $\text{PtdIns}(3,4)\text{P}_2$, thus regulating the signal from these two important lipids [77]. To study SHIP activity against lipids, radiolabelled [^{32}P] PtdInsP_3 substrate may be required. Previously, PI3Ks were employed to create this, but these are large proteins for which recombinant expression from bacteria is difficult. A current alternative is to use IPMK, which can act on not only the soluble $\text{InsP}(1,4,5)\text{P}_3$ but also the lipid $\text{PtdIns}(4,5)\text{P}_2$ [78], allowing straightforward synthesis of 3[^{32}P] $\text{PtdIns}(3,4,5)\text{P}_3$, specifically labelled in position 3 [79]. Human IPMK is easily produced from *Escherichia coli* [26, 78].

7 Preparing [^{32}P] $_i$ Radiolabelled InsP_5 and Its Use

Several inositol phosphate kinases are quite promiscuous. IPMK, as the name states, is a multikinase able to phosphorylate the inositol ring at positions 4 and 6, but also shows, at least in vitro, the ability to convert InsP_5 to the inositol pyrophosphate PP-

InsP₄ [26, 80]. As noted above, it is also able to phosphorylate the lipid PtdIns(4,5)P₂ to PtdIns(3,4,5)P₃ [78, 81]. The IP6Ks can metabolise several isomers of InsP₅ and InsP₆ to inositol pyrophosphates [82, 83]. But perhaps the inositol phosphate kinase most catalytically flexible is ITPK1, which while primarily characterised as a 5- and 6-kinase (Fig. 3) [33], in certain species also has the ability to phosphorylate position 1 of the inositol ring [35]. This activity enables the synthesis of 1[³²P]Ins(1,3,4,5,6)P₅ [84] by incubating Ins(3,4,5,6)P₄ with [³²P]γATP in the presence of *Entamoeba histolytica* ITPK1 produced in *E. coli* [85] (Fig. 4). Using 1[³²P]Ins(1,3,4,5,6)P₅, an intriguing intersubstrate phosphate transfer activity was also discovered for ITPK1. In the presence of 1[³²P]Ins(1,3,4,5,6)P₅ and ADP, human ITPK1 transfers the radioactive phosphate, generating [³²P]γATP and Ins(3,4,5,6)P₄. The addition of Ins(1,3,4)P₃ to this reaction augmented the rate of dephosphorylation of 1[³²P]Ins(1,3,4,5,6)P₅, as Ins(1,3,4)P₃ now became the acceptor of the radioactive phosphate group, forming radiolabelled [³²P]Ins(1,3,4,5/6)P₄ phosphorylated at the 5 or 6 position, plus again Ins(3,4,5,6)P₄ [84]. Thus the use of radioactive 1[³²P]Ins(1,3,4,5,6)P₅ enabled elucidation of how human ITPK1 regulates the synthesis of Ins(3,4,5,6)P₄, a signalling molecule fundamental to controlling chloride channel conductance [31, 32]. It was later demonstrated, again using 1[³²P]Ins(1,3,4,5,6)P₅, that ITPK1 from the plant *Solanum tuberosum* possesses similar intersubstrate phosphotransferase activity [86].

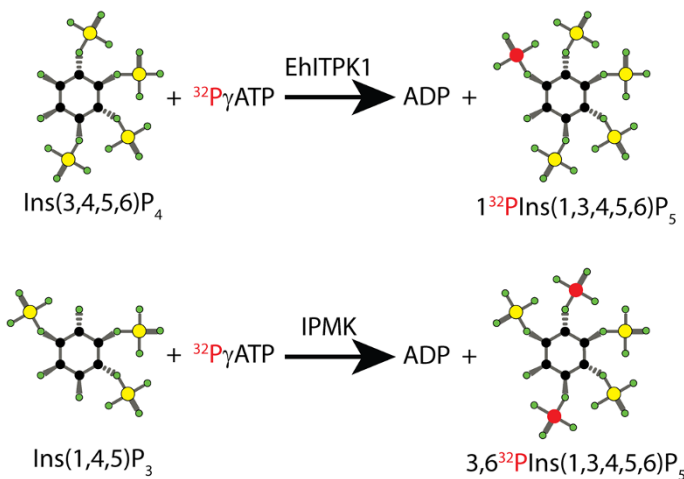


Fig. 4 Schematic synthesis of radiolabelled InsP₅. The most abundant inositol pentakisphosphate isomer, Ins(1,3,4,5,6)P₅, is the common end product of two multikinases, IPMK and ITPK1, using different starting inositol triphosphates (see Fig. 3). Therefore, using these two enzymes and different species of InsP₃ and InsP₄, it is possible to generate InsP₅ labelled in different positions of the inositol ring. The *top* reaction illustrates the ability of *Entamoeba histolytica* ITPK1 (EhITPK1) to phosphorylate position 1 [34, 35], enabling specific synthesis of 1[³²P]Ins(1,3,4,5,6)P₅. Conversely, the *bottom* reaction uses the mammalian IPMK, a 3,6 kinase. Recombinant IPMK can thus be used to generate radiolabelled 3,6[³²P]Ins(1,3,4,5,6)P₅ [88]. Different atoms are colour-coded as follows: carbon *black circle*; oxygen *green circle*; phosphate *yellow circle*; radioactive phosphate *red circle*

A different [^{32}P]Ins(1,3,4,5,6) P_5 radiolabelled isomer was initially purified from metabolically labelled cells [72, 87]. It can now be made using IPMK with Ins(1,4,5) P_3 and [^{32}P] γ ATP, generating 3,6[^{32}P]Ins(1,3,4,5,6) P_5 (Fig. 4). Similarly, [^3H]Ins(1,3,4,5,6) P_3 is generated by incubating recombinant IPMK with the commercially available [^3H]Ins(1,3,4,5) P_4 and ATP. Both radiolabelled reagents 3,6[^{32}P]Ins(1,3,4,5,6) P_5 and [^3H]Ins(1,3,4,5,6) P_5 have been employed to great effect in studying the metabolic stability of Ins P_5 and its anti-tumour capability [88].

8 Preparing [^{32}P] $_i$ Radiolabelled Ins P_6 and Its Use

The commercial availability of [^3H]Ins P_6 has been intermittent over the years. This is unfortunate, as it was indispensable, for example, in the cloning of the IP6K enzymes from rat brain homogenate: the conversion of [^3H]IP $_6$ to [^3H]Ins P_7 was followed using polyethylenimine cellulose thin-layer chromatography (PEI-TLC) [25, 68]. The custom synthesis of [^{32}P]Ins P_6 was similarly essential to the identification of the other class of enzymes, PPIP5Ks, able to synthesise inositol pyrophosphates. High specific activity [^{32}P]Ins P_6 can be generated enzymatically by incubating Ins(1,3,4,5,6) P_5 and [^{32}P] γ ATP with recombinant *Arabidopsis thaliana* IPK1 enzyme [89]. After HPLC purification, this enzymatic reaction generates [^{32}P]IP $_6$ specifically labelled in position 2, i.e., 2[^{32}P]Ins(1,2,3,4,5,6) P_6 . The use of this compound spurred the identification, cloning and characterisation of the yeast PPIP5K (Vip1), since its conversion to [^{32}P]Ins P_7 by *kcs1Δ* (IP6K deletion) yeast extracts revealed the presence of another inositol pyrophosphate synthase activity [47]. Radiolabelled [^{32}P]Ins P_6 has also been used to follow IP6K activity during the yeast cell cycle [90] by simply resolving radioactive [^{32}P]IP $_6$ and [^{32}P]IP $_7$ by PEI-TLC. In addition to eliminating the need for a sophisticated HPLC apparatus, this experimental approach is quantitative, since the radioactivity present in the [^{32}P]Ins P_6 and [^{32}P]Ins P_7 TLC spots can be measured by scraping and counting them in a scintillation counter.

9 Preparing [^{32}P] $_i$ Radiolabelled Ins P_7 and Its Use in Protein Pyrophosphorylation Reactions

The inositol pyrophosphates Ins P_7 and Ins P_8 have been linked to many cellular roles, but the mechanism is not clear. These are known to be dynamic molecules, unlike their “metabolically inert” precursor Ins P_6 [2]: in mammalian cells, up to 50% of the pool of Ins P_6 may be converted to Ins P_7 or Ins P_8 per hour [91]. One possible mode of action for Ins P_7 is protein pyrophosphorylation, in which the β -phosphate is donated to a pre-phosphorylated serine, becoming Ins P_6 and generating a pyrophosphoserine residue. The discovery of this post-translational modification followed the cloning of IP6K1 [25] and the subsequent ability to synthesise Ins P_7 radiolabelled at the β -position of the pyrophosphate moiety, 5[^{32}P] β Ins P_7 (Fig. 5), using Ins P_6 and [^{32}P] γ ATP [92]. Synthesis must be followed by a saxHPLC purification procedure to remove any leftover [^{32}P] γ ATP. Radiolabelled

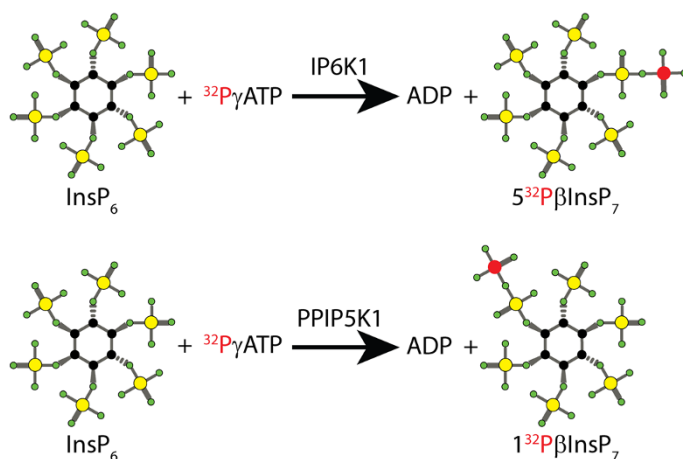


Fig. 5 Schematic synthesis of radiolabelled $5[^{32}\text{P}]\beta\text{InsP}_7$. Two different InsP_7 isomers can be easily and rapidly synthesised biochemically, using recombinant IP6K1 or the kinase domain of PPIP5K1 (or its yeast counterpart Vip1). Using radiolabelled $[^{32}\text{P}]\gamma\text{ATP}$ and InsP_6 as substrate, IP6K1 generates $5[^{32}\text{P}]\beta\text{InsP}_7$ (top), thus transferring the radioactive $[^{32}\text{P}]$ from ATP to the phosphorylated position 5 of InsP_6 . PPIP5K instead generates the isomer $1[^{32}\text{P}]\beta\text{InsP}_7$ (bottom). Different atoms are colour-coded as follows: carbon black circle; oxygen green circle; phosphate yellow circle; radioactive phosphate red circle

$5[^{32}\text{P}]\beta\text{InsP}_7$ fractions are then ready to use following desalting with a Sep-Pak QMA cartridge and concentration with a centrifugal evaporator [52]. Experiments demonstrated kinase-independent phosphorylation of multiple proteins *in vitro* [93]. In the absence of alternative detection methods, the use of the labelled $5[^{32}\text{P}]\beta\text{InsP}_7$ is still reported in all publications of serine pyrophosphorylation [94–96].

Performing the *in vitro* pyrophosphorylation/transphosphorylation experiment, once $5[^{32}\text{P}]\beta\text{InsP}_7$ is synthesised, is a straightforward process. However, it is important to remember the requirement for a pre-phosphorylated serine residue as substrate. Thus proteins of bacterial origin, such as recombinant mammalian proteins expressed in *E. coli*, require a priming event by a classical ATP kinase, usually casein kinase 2 (CK2) [93, 94]. The $5[^{32}\text{P}]\beta\text{InsP}_7$ pyrophosphorylation is a non-enzymatic, temperature-dependent event that, while it does occur at physiological temperatures, is enhanced by incubating the sample at higher temperatures [97]. Once pyrophosphorylated, proteins can be resolved by polyacrylamide gel electrophoresis (PAGE), and their radioactivity incorporation can be easily identified by autoradiography. Several substrates of InsP_7 -mediated pyrophosphorylation have been identified so far. A common feature to all is stretches of serines embedded in regions rich in acidic residues; as magnesium is required for pyrophosphorylation, it is possible that the acidic residues are required to coordinate these ions. Pyrophosphorylated proteins are more acid-labile but also more resistant to phosphatases than proteins phosphorylated by ATP alone [93]. If this is true *in vivo*, it could have huge implications for cellular signalling: InsP_7 -mediated protein pyrophosphorylation could act in a dominant manner to allow continued signalling even during phosphatase activation.

It is important to note that inositol pyrophosphates appear to be interchangeable in their ability to pyrophosphorylate proteins. For example, $1[^{32}\text{P}]\beta\text{InsP}_7$ synthesised using the Vip1 (PPIP5K) kinase domain and $[^{32}\text{P}]\gamma\text{ATP}$ (Fig. 5) is able to pyrophosphorylate proteins, as is $1,5[^{32}\text{P}]\text{InsP}_8$ synthesised by the double action of IP6K1 and Vip1. In fact, the incubation of *S. cerevisiae* extracts with $5[^{32}\text{P}]\beta\text{InsP}_7$, $1[^{32}\text{P}]\beta\text{InsP}_7$ or $1,5[^{32}\text{P}]\beta\text{InsP}_8$ revealed an identical pattern of pyrophosphorylated proteins [93]. This indicates that if pyrophosphorylation is the mechanism of action for inositol pyrophosphates, they can substitute for each other in performing their biological role. However, allosteric regulation of proteins has also been proposed as a mechanism [98, 99], and in this circumstance, different isomeric species of inositol pyrophosphates could be recognised by specific effector proteins.

The synthesis described here for $5[^{32}\text{P}]\beta\text{InsP}_7$ is a straightforward procedure, as is using it to identify pyrophosphorylated proteins. However, processing millicurie amounts of radioactivity to generate a reagent with a two week half-life requires a degree of dedication. Therefore, protein pyrophosphorylation, while considered an exciting post-translational modification, has received little attention, as only a handful of laboratories have invested in the synthesis of $5[^{32}\text{P}]\beta\text{InsP}_7$. The recent development of organic synthesised experimental tools to study protein pyrophosphorylation [100, 101] will likely lead to the further and very welcome development of non-radioactive detection methods, and to the demonstration of the existence of this modification in vivo. These developments will certainly increase interest in protein pyrophosphorylation, but it must be remembered that it was the innovative use of radioactive labelling methods that permitted the discovery of this modification in the first place. This is an excellent demonstration that original radioactive assays should not be avoided.

10 Perspective

In today's "omics" era, radioactive labelling appears anachronistic. Current proteomic, metabolomic or genomic studies give us plentiful information in an ever-growing number of databases. Ultimately, however, these huge amounts of data must not be just statistically annotated, but must drive biochemistry experiments, which as noted before, provide the only experimental basis for the understanding of biological mechanisms [102, 103]. In this context, radioactive molecular tracers have played and will continue to play a contributory role in our quest to understand the molecular mechanisms of life.

This review has highlighted the fundamental importance of radioactive labelling in the birth and the development of inositol polyphosphate research. Although recent technological development efforts such as PAGE analysis [82] and TiO_2 extraction of inositol polyphosphates [104] are facilitating the study of the metabolism and functions of these molecules without the need for radioactive precursors or metabolic labelling, we definitely foresee further need for phosphate-32/33-, carbon-14- or tritium-based experiments to fully appreciate the importance of these molecules in cell biology.

Acknowledgements We wish to thank the members of the Saiardi laboratory for reading of the manuscript. We have tried to review all the relevant literature, but we may have missed some specific work; therefore, we would like to apologise to colleagues whose research was not properly recognised. This work is supported by the Medical Research Council (MRC) core support to the MRC/UCL Laboratory for Molecular Cell Biology University Unit (MC_U122680443).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. York JD (2006) Regulation of nuclear processes by inositol polyphosphates. *Biochim Biophys Acta* 1761(5–6):552–559
2. Menniti FS et al (1993) Turnover of inositol polyphosphate pyrophosphates in pancreaticoma cells. *J Biol Chem* 268(6):3850–3856
3. Stephens L et al (1993) The detection, purification, structural characterization, and metabolism of diphosphoinositol pentakisphosphate(s) and bisdiphosphoinositol tetrakisphosphate(s). *J Biol Chem* 268(6):4009–4015
4. Bennett M et al (2006) Inositol pyrophosphates: metabolism and signaling. *Cell Mol Life Sci* 63(5):552–564
5. Irvine RF (2003) 20 years of Ins(1,4,5)P₃, and 40 years before. *Nat Rev Mol Cell Biol* 4(7):586–590
6. Michell RH (1975) Inositol phospholipids and cell surface receptor function. *Biochim Biophys Acta* 415(1):81–147
7. Berridge MJ, Lipp P, Bootman MD (2000) The versatility and universality of calcium signalling. *Nat Rev Mol Cell Biol* 1(1):11–21
8. Hokin LE, Hokin MR (1955) Effects of acetylcholine on phosphate turnover in phospholipides of brain cortex in vitro. *Biochim Biophys Acta* 16(2):229–237
9. Hokin MR, Hokin LE (1953) Enzyme secretion and the incorporation of P₃₂ into phospholipides of pancreas slices. *J Biol Chem* 203(2):967–977
10. Streb H et al (1983) Release of Ca²⁺ from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate. *Nature* 306(5938):67–69
11. Hershey AD, Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* 36(1):39–56
12. Ivanov VK (2012) Lessons from Chernobyl and prognosis for Fukushima: radiological consequences. *J Radiol Prot* 32(1):N55–N58
13. Engelbrecht R, Schwaiger M (2008) State of the art of standard methods used for environmental radioactivity monitoring. *Appl Radiat Isot* 66(11):1604–1610
14. Meisenhelder J, Semba K (2001) Safe use of radioisotopes. *Curr Protoc Cell Biol* Appendix 1: p. Appendix 1D
15. Irvine RF, Schell MJ (2001) Back in the water: the return of the inositol phosphates. *Nat Rev Mol Cell Biol* 2(5):327–338
16. Resnick AC, Saiardi A (2008) Inositol Polyphosphates. In: Begley TP (ed) *Wiley Encyclopedia of Chemical Biology*. Wiley, Hoboken
17. Shears SB et al (2012) Defining signal transduction by inositol phosphates. *Subcell Biochem* 59:389–412
18. York JD et al (2001) An expanded view of inositol signaling. *Adv Enzyme Regul* 41:57–71
19. Irvine RF (2005) Inositide evolution-towards turtle domination? *J Physiol* 566(Pt 2):295–300
20. Balla T (2013) Phosphoinositides: tiny lipids with giant impact on cell regulation. *Physiol Rev* 93(3):1019–1137
21. Di Paolo G, De Camilli P (2006) Phosphoinositides in cell regulation and membrane dynamics. *Nature* 443(7112):651–657

22. Livermore TM et al (2016) Phosphate, inositol and polyphosphates. *Biochem Soc Trans* 44(1):253–259
23. Odom AR et al (2000) A role for nuclear inositol 1,4,5-trisphosphate kinase in transcriptional control. *Science* 287(5460):2026–2029
24. Saiardi A et al (2000) Inositol polyphosphate multikinase (ArgRIII) determines nuclear mRNA export in *Saccharomyces cerevisiae*. *FEBS Lett* 468(1):28–32
25. Saiardi A et al (1999) Synthesis of diphosphoinositol pentakisphosphate by a newly identified family of higher inositol polyphosphate kinases. *Curr Biol* 9(22):1323–1326
26. Saiardi A et al (2001) Mammalian inositol polyphosphate multikinase synthesizes inositol 1,4,5-trisphosphate and an inositol pyrophosphate. *Proc Natl Acad Sci U S A* 98(5):2306–2311
27. Verbsky JW et al (2002) The synthesis of inositol hexakisphosphate. Characterization of human inositol 1,3,4,5,6-pentakisphosphate 2-kinase. *J Biol Chem* 277(35):31857–31862
28. York JD et al (1999) A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science* 285(5424):96–100
29. Takazawa K et al (1991) Molecular cloning and expression of a human brain inositol 1,4,5-trisphosphate 3-kinase. *Biochem Biophys Res Commun* 174(2):529–535
30. Togashi S et al (1997) Structural identification of the myo-inositol 1,4,5-trisphosphate-binding domain in rat brain inositol 1,4,5-trisphosphate 3-kinase. *Biochem J* 326(Pt 1):221–225
31. Mitchell J et al (2008) An expanded biological repertoire for Ins(3,4,5,6)P4 through its modulation of CIC-3 function. *Curr Biol* 18(20):1600–1605
32. Saiardi A, Cockcroft S (2008) Human ITPK1: a reversible inositol phosphate kinase/phosphatase that links receptor-dependent phospholipase C to Ca²⁺-activated chloride channels. *Sci Signal* 1(4):pe5
33. Wilson MP, Majerus PW (1996) Isolation of inositol 1,3,4-trisphosphate 5/6-kinase, cDNA cloning and expression of the recombinant enzyme. *J Biol Chem* 271(20):11904–11910
34. Liu C et al (2001) Synthesis and biological activity of D- and L-chiro-inositol 2,3,4,5-tetrakisphosphate: design of a novel and potent inhibitor of Ins(3,4,5,6)P4 1-kinase/Ins(1,3,4)P3 5/6-kinase. *J Med Chem* 44(18):2984–2989
35. Yang X, Shears SB (2000) Multitasking in signal transduction by a promiscuous human Ins(3,4,5,6)P4 1-kinase/Ins(1,3,4)P3 5/6-kinase. *Biochem J* 351(Pt 3):551–555
36. Loss O et al (2013) Modulation of inositol polyphosphate levels regulates neuronal differentiation. *Mol Biol Cell* 24(18):2981–2989
37. Shears SB (2009) Diphosphoinositol polyphosphates: metabolic messengers? *Mol Pharmacol* 76(2):236–252
38. Wilson MS, Livermore TM, Saiardi A (2013) Inositol pyrophosphates: between signalling and metabolism. *Biochem J* 452(3):369–379
39. Wundenberg T, Mayr GW (2012) Synthesis and biological actions of diphosphoinositol phosphates (inositol pyrophosphates), regulators of cell homeostasis. *Biol Chem* 393(9):979–998
40. Barker CJ et al (2009) Inositol pyrophosphates: structure, enzymology and function. *Cell Mol Life Sci* 66(24):3851–3871
41. Draskovic P et al (2008) Inositol hexakisphosphate kinase products contain diphosphate and triphosphate groups. *Chem Biol* 15(3):274–286
42. Saiardi A et al (2001) Identification and characterization of a novel inositol hexakisphosphate kinase. *J Biol Chem* 276(42):39179–39185
43. Shears SB (2015) Inositol pyrophosphates: why so many phosphates? *Adv Biol Regul* 57:203–216
44. Choi JH et al (2007) Purification, sequencing, and molecular identification of a mammalian PP-InsP5 kinase that is activated when cells are exposed to hyperosmotic stress. *J Biol Chem* 282(42):30763–30775
45. Fridy PC et al (2007) Cloning and characterization of two human VIP1-like inositol hexakisphosphate and diphosphoinositol pentakisphosphate kinases. *J Biol Chem* 282(42):30754–30762
46. Lin H et al (2009) Structural analysis and detection of biological inositol pyrophosphates reveal that the family of VIP/diphosphoinositol pentakisphosphate kinases are 1/3-kinases. *J Biol Chem* 284(3):1863–1872
47. Mulugu S et al (2007) A conserved family of enzymes that phosphorylate inositol hexakisphosphate. *Science* 316(5821):106–109
48. Laha D et al (2015) VIH2 regulates the synthesis of inositol pyrophosphate InsP8 and jasmonate-dependent defenses in *Arabidopsis*. *Plant Cell* 27(4):1082–1097

49. Onnebo SM, Saiardi A (2009) Inositol pyrophosphates modulate hydrogen peroxide signaling. *Biochem J* 423(1):109–118. doi:10.1042/BJ20090241
50. Mayr GW (1988) A novel metal-dye detection system permits picomolar-range h.p.l.c. analysis of inositol polyphosphates from non-radioactively labelled cell or tissue specimens. *Biochem J* 254(2):585–591
51. Albert C et al (1997) Biological variability in the structures of diphosphoinositol polyphosphates in *Dictyostelium discoideum* and mammalian cells. *Biochem J* 327(Pt 2):553–560
52. Azevedo C, Saiardi A (2006) Extraction and analysis of soluble inositol polyphosphates from yeast. *Nat Protoc* 1(5):2416–2422
53. Stevenson-Paulik J et al (2006) Inositol phosphate metabolomics: merging genetic perturbation with modernized radiolabeling methods. *Methods* 39(2):112–121
54. Saiardi A et al (2002) Inositol pyrophosphates regulate endocytic trafficking. *Proc Natl Acad Sci U S A* 99(22):14206–14211
55. Michell RH et al (1988) Inositol lipids: receptor-stimulated hydrolysis and cellular lipid pools. *Philos Trans R Soc Lond B Biol Sci* 320(1199):239–246
56. Maccallum SH et al (1989) The use of cells doubly labelled with [¹⁴C]inositol and [³H]inositol to search for a hormone-sensitive inositol lipid pool with atypically rapid metabolic turnover. *J Endocrinol* 122(1):379–389
57. Hughes PJ et al (1989) The regulation of the phosphorylation of inositol 1,3,4-trisphosphate in cell-free preparations and its relevance to the formation of inositol 1,3,4,6-tetrakisphosphate in agonist-stimulated rat parotid acinar cells. *J Biol Chem* 264(33):19871–19878
58. Downes CP, Hawkins PT, Irvine RF (1986) Inositol 1,3,4,5-tetrakisphosphate and not phosphatidylinositol 3,4-bisphosphate is the probable precursor of inositol 1,3,4-trisphosphate in agonist-stimulated parotid gland. *Biochem J* 238(2):501–506
59. King CE et al (1987) Multiple metabolic pools of phosphoinositides and phosphatidate in human erythrocytes incubated in a medium that permits rapid transmembrane exchange of phosphate. *Biochem J* 244(1):209–217
60. Palmer S et al (1986) The labelling of polyphosphoinositides with [³²P]Pi and the accumulation of inositol phosphates in vasopressin-stimulated hepatocytes. *Biochem J* 238(2):491–499
61. Sulpice JC et al (1989) The separation of [³²P]inositol phosphates by ion-pair chromatography: optimization of the method and biological applications. *Anal Biochem* 179(1):90–97
62. Azevedo C, Saiardi A (2014) Functions of inorganic polyphosphates in eukaryotic cells: a coat of many colours. *Biochem Soc Trans* 42(1):98–102
63. Docampo R, Ulrich P, Moreno SN (2010) Evolution of acidocalcisomes and their role in polyphosphate storage and osmoregulation in eukaryotic microbes. *Philos Trans R Soc Lond B Biol Sci* 365(1541):775–784
64. Gerasimaite R et al (2014) Coupled synthesis and translocation restrains polyphosphate to acidocalcisome-like vacuoles and prevents its toxicity. *J Cell Sci* 127(Pt 23):5093–5104
65. Lander N, Ulrich PN, Docampo R (2013) Trypanosoma brucei vacuolar transporter chaperone 4 (TbVtc4) is an acidocalcisome polyphosphate kinase required for in vivo infection. *J Biol Chem* 288(47):34205–34216
66. Lonetti A et al (2011) Identification of an evolutionarily conserved family of inorganic polyphosphate endopolyphosphatases. *J Biol Chem* 286(37):31966–31974
67. Jones DR et al (2013) Measurement of phosphoinositides in the zebrafish *Danio rerio*. *Nat Protoc* 8(6):1058–1072
68. Voglmaier SM et al (1996) Purified inositol hexakisphosphate kinase is an ATP synthase: diphosphoinositol pentakisphosphate as a high-energy phosphate donor. *Proc Natl Acad Sci U S A* 93(9):4305–4310
69. Ye W et al (1995) Inhibition of clathrin assembly by high affinity binding of specific inositol polyphosphates to the synapse-specific clathrin assembly protein AP-3. *J Biol Chem* 270(4):1564–1568
70. Downes CP, Michell RH (1981) The polyphosphoinositide phosphodiesterase of erythrocyte membranes. *Biochem J* 198(1):133–140
71. Shears SB et al (1987) Dephosphorylation of myo-inositol 1,4,5-trisphosphate and myo-inositol 1,3,4-trisphosphate. *Biochem J* 242(2):393–402
72. Stephens LR, Downes CP (1990) Product-precursor relationships amongst inositol polyphosphates. Incorporation of [³²P]Pi into myo-inositol 1,3,4,6-tetrakisphosphate, myo-inositol 1,3,4,5-

- tetrakisphosphate, myo-inositol 3,4,5,6-tetrakisphosphate and myo-inositol 1,3,4,5,6-pentakisphosphate in intact avian erythrocytes. *Biochem J* 265(2):435–452
73. Stephens LR et al (1988) L-myo-inositol 1,4,5,6-tetrakisphosphate (3-hydroxy)kinase. *Biochem J* 249(1):283–292
 74. Van Dijken P et al (1994) Phosphorylation of inositol 1,4,5-trisphosphate analogues by 3-kinase and dephosphorylation of inositol 1,3,4,5-tetrakisphosphate analogues by 5-phosphatase. *Eur J Biochem* 226(2):561–566
 75. Pesesse X et al (1998) The SH2 domain containing inositol 5-phosphatase SHIP2 displays phosphatidylinositol 3,4,5-trisphosphate and inositol 1,3,4,5-tetrakisphosphate 5-phosphatase activity. *FEBS Lett* 437(3):301–303
 76. Van der Kaay J, Wesseling J, Van Haastert PJ (1995) Nucleus-associated phosphorylation of Ins(1,4,5)P₃ to InsP₆ in *Dictyostelium*. *Biochem J* 312(Pt 3):911–917
 77. Erneux C et al (2011) SHIP2 multiple functions: a balance between a negative control of PtdIns(3,4,5)P₃ level, a positive control of PtdIns(3,4)P₂ production, and intrinsic docking properties. *J Cell Biochem* 112(9):2203–2209
 78. Resnick AC et al (2005) Inositol polyphosphate multikinase is a nuclear PI3-kinase with transcriptional regulatory activity. *Proc Natl Acad Sci U S A* 102(36):12783–12788
 79. Vandeput F et al (2006) The influence of anionic lipids on SHIP2 phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase activity. *Cell Signal* 18(12):2193–2199
 80. Zhang T, Caffrey JJ, Shears SB (2001) The transcriptional regulator, Arg82, is a hybrid kinase with both monophosphoinositol and diphosphoinositol polyphosphate synthase activity. *FEBS Lett* 494(3):208–212
 81. Maag D et al (2011) Inositol polyphosphate multikinase is a physiologic PI3-kinase that activates Akt/PKB. *Proc Natl Acad Sci U S A* 108(4):1391–1396
 82. Losito O et al (2009) Inositol pyrophosphates and their unique metabolic complexity: analysis by gel electrophoresis. *PLoS One* 4(5):e5580
 83. Saiardi A et al (2000) The inositol hexakisphosphate kinase family. Catalytic flexibility and function in yeast vacuole biogenesis. *J Biol Chem* 275(32):24686–24692
 84. Chamberlain PP et al (2007) Integration of inositol phosphate signaling pathways via human ITPK1. *J Biol Chem* 282(38):28117–28125
 85. Miller GJ et al (2005) Specificity determinants in inositol polyphosphate synthesis: crystal structure of inositol 1,3,4-trisphosphate 5/6-kinase. *Mol Cell* 18(2):201–212
 86. Caddick SE et al (2008) A *Solanum tuberosum* inositol phosphate kinase (SITPK1) displaying inositol phosphate-inositol phosphate and inositol phosphate-ADP phosphotransferase activities. *FEBS Lett* 582(12):1731–1737
 87. Caffrey JJ et al (2001) Expanding coincident signaling by PTEN through its inositol 1,3,4,5,6-pentakisphosphate 3-phosphatase activity. *FEBS Lett* 499(1–2):6–10
 88. Maffucci T et al (2005) Inhibition of the phosphatidylinositol 3-kinase/akt pathway by inositol pentakisphosphate results in antiangiogenic and antitumor effects. *Cancer Res* 65(18):8339–8349
 89. Stevenson-Paulik J et al (2005) Generation of phytate-free seeds in Arabidopsis through disruption of inositol polyphosphate kinases. *Proc Natl Acad Sci U S A* 102(35):12612–12617
 90. Banfic H et al (2013) Inositol pyrophosphates modulate S phase progression after pheromone-induced arrest in *Saccharomyces cerevisiae*. *J Biol Chem* 288(3):1717–1725
 91. Glennon MC, Shears SB (1993) Turnover of inositol pentakisphosphates, inositol hexakisphosphate and diphosphoinositol polyphosphates in primary cultured hepatocytes. *Biochem J* 293(Pt 2):583–590
 92. Saiardi A et al (2004) Phosphorylation of proteins by inositol pyrophosphates. *Science* 306(5704):2101–2105
 93. Bhandari R et al (2007) Protein pyrophosphorylation by inositol pyrophosphates is a posttranslational event. *Proc Natl Acad Sci U S A* 104(39):15305–15310
 94. Azevedo C et al (2009) Inositol pyrophosphate mediated pyrophosphorylation of AP3B1 regulates HIV-1 Gag release. *Proc Natl Acad Sci U S A* 106(50):21161–21166
 95. Chanduri M et al (2016) Inositol hexakisphosphate kinase 1 (IP6K1) activity is required for cytoplasmic dynein-driven transport. *Biochem J* 473(19):3031–3047
 96. Thota SG et al (2015) Inositol pyrophosphates regulate RNA polymerase I-mediated rRNA transcription in *Saccharomyces cerevisiae*. *Biochem J* 466(1):105–114
 97. Werner Jr JK, Speed T, Bhandari R (2010) Protein pyrophosphorylation by diphosphoinositol pentakisphosphate (InsP₇). *Methods Mol Biol* 645:87–102

98. Burton A, Hu X, Saiardi A (2009) Are inositol pyrophosphates signalling molecules? *J Cell Physiol* 220(1):8–15
99. Shears SB et al (2011) Diphosphoinositol polyphosphates: what are the mechanisms? *Adv Enzyme Regul* 51(1):13–25
100. Brown NW, Marmelstein AM, Fiedler D (2016) Chemical tools for interrogating inositol pyrophosphate structure and function. *Chem Soc Rev* 45(22):6311–6326
101. Wu M et al (2014) Elucidating diphosphoinositol polyphosphate function with nonhydrolyzable analogues. *Angew Chem Int Ed Engl* 53(28):7192–7197
102. Brenner S (2000) Biochemistry strikes back. *Trends Biochem Sci* 25(12):584
103. Kornberg A (2004) Biochemistry matters. *Nat Struct Mol Biol* 11(6):493
104. Wilson MS et al (2015) A novel method for the purification of inositol phosphates from biological samples reveals that no phytate is present in human plasma or urine. *Open Biol* 5(3):150014
105. Communi D, Vanweyenberg V, Erneux C (1995) Molecular study and regulation of D-myo-inositol 1,4,5-trisphosphate 3-kinase. *Cell Signal* 7(7):643–650
106. Dewaste V et al (2000) Cloning and expression of a cDNA encoding human inositol 1,4,5-trisphosphate 3-kinase C. *Biochem J* 352(Pt 2):343–351
107. Takazawa K et al (1990) Cloning and expression in *E. coli* of a rat brain cDNA encoding a Ca^{2+} /calmodulin-sensitive inositol 1,4,5-trisphosphate 3-kinase. *Biochem J* 272(1):107–112
108. Wilson MP et al (2001) Inositol 1,3,4-trisphosphate 5/6-kinase is a protein kinase that phosphorylates the transcription factors c-Jun and ATF-2. *J Biol Chem* 276(44):40998–41004
109. Wang H et al (2012) Structural basis for an inositol pyrophosphate kinase surmounting phosphate crowding. *Nat Chem Biol* 8(1):111–116
110. Stephens LR, Irvine RF (1990) Stepwise phosphorylation of myo-inositol leading to myo-inositol hexakisphosphate in *Dictyostelium*. *Nature* 346(6284):580–583

Applications and Advantages of Stable Isotope Phosphate Labeling of RNA in Mass Spectrometry

Kayla Borland¹ · Patrick A. Limbach¹

Received: 21 November 2016 / Accepted: 17 February 2017 / Published online: 11 March 2017
© Springer International Publishing Switzerland 2017

Abstract Mass spectrometry (MS) has become an enabling technology for the characterization of post-transcriptionally modified nucleosides within ribonucleic acids (RNAs). These modified RNAs tend to be more challenging to completely characterize using conventional genomic-based sequencing technologies. As with many biological molecules, information relating to the presence or absence of a particular compound (i.e., qualitative measurement) is only one step in sample characterization. Additional useful information is found by performing quantitative measurements on the levels of the compound of interest in the sample. Phosphate labeling of modified RNAs has been developed by our laboratory to enhance conventional mass spectrometry techniques. By taking advantage of the mechanism of action of many ribonucleases (RNases), digesting RNA samples in the presence of ¹⁸O-labeled water generates an ¹⁸O-labeled 3'-phosphate in each digestion product. We describe the historical development of this approach, contrast this stable isotope labeling strategy with others used in RNA mass spectrometry, and provide examples of new analytical mass spectrometry methods that are enabled by phosphate labeling in this fashion.

Keywords Stable isotope labeling · Modified nucleosides · RNA sequencing · tRNA · ¹⁸O-enriched water · LC-MS/MS

This article is part of the Topical Collection “Phosphate Labeling and Sensing in Chemical Biology”; edited by Henning Jessen.

✉ Patrick A. Limbach
Pat.Limbach@uc.edu

¹ Rieveschl Laboratories for Mass Spectrometry, Department of Chemistry, University of Cincinnati, PO Box 210172, Cincinnati, OH 45221-0172, USA

1 Mass Spectrometry of Modified Ribonucleic Acids

Mass spectrometry (MS) is a powerful and popular analytical platform for the characterization of biomolecules. However, the application of mass spectrometry for characterizing nucleic acids has lagged behind other classes, due to the simplicity, speed and sensitivity of amplification-driven technologies such as Sanger and Next-Gen sequencing. Where MS has proven most useful in nucleic acids is in the direct detection of modified nucleosides. In both deoxyribonucleic acids (DNA) and ribonucleic acids (RNA), many nucleosides can be enzymatically or chemically modified. These modifications have important biological functions or outcomes, including RNA editing, RNA stability and protein expression [1–3]. Not surprisingly, techniques and technologies that enable the rapid determination of modified nucleosides remain an ongoing interest. The effectiveness of MS as an enabling technology is that it can reveal the mass and the structure of the modified nucleoside, which many amplification-based approaches are unable to perform directly.

A primary focus of our laboratory has been developing MS approaches that enable the rapid and accurate identification of modified nucleosides from RNA. More specifically, one goal has been to create a platform that enables RNA modification mapping—placing identified modified nucleosides into the correct RNA sequence context. The basis for RNA modification mapping by MS is a hyphenated liquid chromatography tandem MS (LC–MS/MS) approach although an alternative method using matrix-assisted laser desorption/ionization MS (MALDI-MS) has been used by us and others.

RNA modification mapping by MS was initially developed by McCloskey and co-workers [4]. The general approach involves two separate experiments. The first experiment allows one to obtain a census of all the modified nucleosides in the RNA sample of interest by completely digesting the intact RNA into individual nucleosides, which are separated and identified by LC–MS/MS [5]. The second experiment requires that the intact RNA first be digested using a specific nuclease, which will generate a mixture of oligonucleotides of varying length. This mixture of digestion products is then analyzed by LC–MS/MS as well [6]. Here, the MS/MS step is used to fragment an oligonucleotide by collision-induced dissociation (CID) such that the original sequence can be reconstructed [7]. As noted above, a similar approach can be used with MALDI-MS, as demonstrated by Kirpekar and coworkers during the mapping of post-transcriptional modifications to ribosomal RNAs (rRNAs) [8].

RNA modification mapping by MS is facilitated these days by the availability of known RNA sequences, which arise due to advances in genomic sequencing technologies. These sequences reflect the status of the RNA lacking modification, thus one can readily calculate the molecular weights of unmodified RNAs and any subsequent RNase digestion products using a variety of online tools. Because nearly all RNA modifications result in an increase in the mass of the canonical nucleoside, digestion products matching the calculated value will not be modified. As such, experimental strategies now limit data analysis primarily to those RNase digestion

products whose masses do not match, suggesting the digestion product contains a modification. The interested reader is directed to a number of recent publications that describe RNA modification mapping by MS in more detail [9–12].

2 Stable Isotope Labeling Methods for RNA Mass Spectrometry

While methods that allow one to identify modified nucleosides and map those nucleosides onto specific sequence locations of an RNA sample are quite powerful, methods that allow for quantitative measurement of modification levels are needed to better inform and understand the biological significance of these molecules. As is commonly conducted in other areas of MS, the field of nucleic acid modifications has turned to stable isotope approaches to improve both the qualitative analysis of modified RNAs and the quantitative measurement of modified nucleosides/nucleic acids. Before discussing phosphate labeling of RNA by using ^{18}O -labeled water in detail, it is first of value to briefly summarize other approaches that have been used in the field as those will serve as a good entry point to the analytical advantages of using stable isotopes in the identification of modified nucleosides and mapping those modifications onto RNA sequences.

Two styles of labeling can be employed: *in vivo* or *in vitro*. An example of *in vivo* labeling is when a medium containing stable isotope labeled nutrients (e.g., essential amino acids) is used in the culturing of the organism of interest. The normal biochemistry pathways of the organism will result in the incorporation of the stable isotope, which provides a specific traceable marker for identification in MS. The most common form of *in vivo* labeling in MS is stable isotope labeling by amino acids in cell culture (SILAC) [13], which has found widespread application in proteomics. SILAC relies on the addition of Leu-D3- or ^{13}C -labeled arginine or lysine to the medium for incorporation in protein synthesis [13, 14].

The alternative approach is to use some *in vitro* method for labeling. Most often, these *in vitro* methods rely on chemical or enzymatic strategies to incorporate the stable isotope into the biomolecule(s) of interest. Due to ionization rates and the potential coupling to liquid chromatography, the use of labels as close to the original product is desired. The goal of isotopic labeling is to change the mass of the ion but minimize effects on ionization rates and chromatography retention.

Bruckl and co-workers demonstrated parallel isotope-based quantification of modified transfer RNA (tRNA) nucleosides [15]. A subset of modified tRNA nucleosides were generated using deuterium labeling of a methyl group, which were used as internal standards to quantify modified nucleoside levels in tumor cells versus healthy tissue. The area under the peak of the known concentration spike was compared to the calibration curves created for each isotopically labeled modified nucleoside. Kellner and co-workers have developed a method for the absolute quantification of modified ribonucleosides using biosynthetic isotopomers [16]. By feeding ^{13}C glucose to bacteria, stable isotope-labeled modified RNAs are synthesized by the organism. The stable isotope-labeled RNA can be hydrolyzed to nucleosides and those naturally occurring modified nucleosides, generated by the bacterium, are then used as internal standards.

A different stable isotope labeling strategy was developed by Dickman and co-workers [17]. By using ^{15}N -labeled medium, they could successfully map post-transcriptional modifications in bacterial 16S rRNA. The approach requires two samples—one cultured in ^{14}N -labeled medium and the other in ^{15}N -labeled medium. This labeling approach allows for the unambiguous identification of base composition in each digestion product, thereby improving the accuracy of RNA modification mapping experiments. While the quantitative applications of this approach were not explored in depth by Dickman and co-workers, such a strategy clearly pointed towards the additional utility of isotope labeling for relative quantification during RNA modification mapping experiments.

Williamson and co-workers adapted the Dickman methodology to quantify rRNA modification levels [18]. Cells were cultured in minimal media and minimal media supplemented with ^{15}N ammonium sulfate as the heavy nitrogen source. The heavy labeled culture was used as an internal standard. Known ratios of ^{15}N and ^{14}N cultured cells were combined for relative quantification. Methylated nucleosides in rRNA were quantified using CD_3 -methionine-supplemented medium while pseudouridine modifications were quantified by cultures supplemented with 5,6-D-uracil.

More recently, an alternative strategy has been developed by Taoka and co-workers for the absolute quantification of post-transcriptional modifications in rRNA [19]. This approach, deemed stable isotope-labeled ribonucleic acid internal standard (SILNAS), relies on an internal standard that is generated by *in vitro* transcription of RNA using ^{13}C -labeled nucleoside triphosphates, which generates an unmodified copy of the rRNA uniformly labeled. After RNase digestion of both the sample of interest and the *in vitro* transcript internal standard, any LC peaks that lack a co-eluting heavy transcript (internal standard) were indicative of a modification in that digestion product. The modified oligonucleotide could be quantified through the ion abundance ratio of spike to sample. Building on that approach, we have recently adapted *in vitro* internal standards for the comparative analysis of tRNA digests [20].

3 ^{18}O Labeling in Mass Spectrometry

The origins of ^{18}O labeling in MS first arose in the field of proteomics. In 1951, Sprinson and Rittenberg took advantage of ^{18}O to better understand enzyme activity of proteases [21]. This idea was adapted and used by Desiderio and Kai in protein sample preparation for MS [22]. By 1983, they were taking advantage of stable isotope-incorporated peptide internal standards for field desorption MS quantification of peptides in biological tissues [22]. They were preparing internal standards by digesting proteins with trypsin in ^{18}O -labeled water. Due to the mechanism of trypsin, digestion under these conditions can result in the C-terminus of the peptide being labeled with one or two ^{18}O molecules, which leads to a 2- or 4-Da mass increase in the tryptic peptide (Fig. 1). As such, unique heavy internal standards could be generated and applied to the MS-based analysis of peptides (and proteins) for identification and quantification.

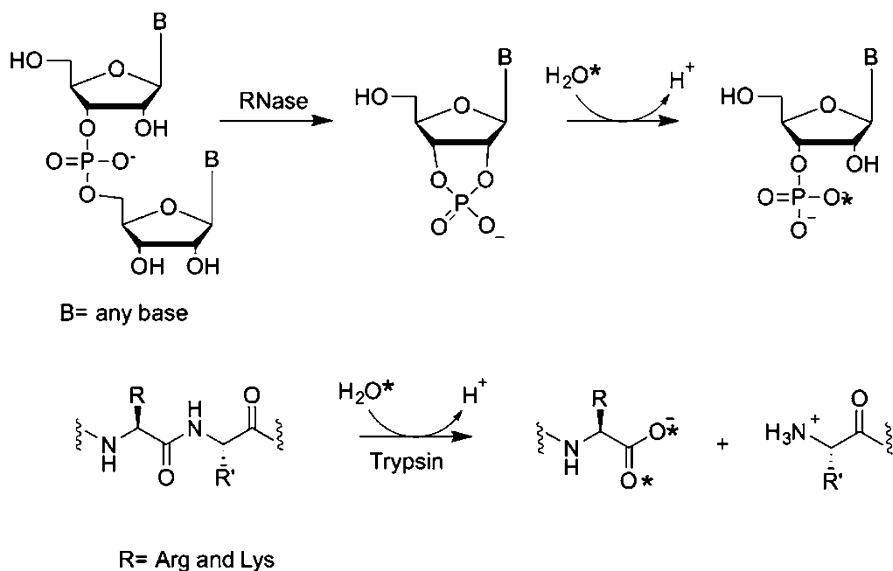


Fig. 1 Endonuclease method for RNase T1 and other RNases. The RNA oligonucleotide is cleaved and a 3' cyclic phosphate intermediate is formed. An ^{18}O atom from the reaction solvent can break the cyclic phosphate and allow for the addition of the stable isotope label to the final 3' linear phosphate product. Protease method for Trypsin. The protein is cleaved at the C-terminus after arginine or lysine residues. The ^{18}O from the labeled reaction solvent is incorporated into the newly formed carboxylic acid after the cleavage site

In 2000, Mirgorodskaya paired ^{18}O stable isotope labeling of proteins with MALDI-MS for quantification of protein samples [23]. The following year Yao introduced a shotgun comparative proteomics tool based on ^{16}O versus ^{18}O labeling of trypsin digested proteins [24]. This comparative proteomics labeling approach has even been applied to clinical samples when a pooled ^{18}O -labeled reference sample was spiked into patient samples [25]. Unlike *in vivo* approaches, ^{18}O labeling is cost-effective due to the minimization of “wasting” the stable isotope as labeling is only performed on extracted protein. On the other hand, ^{18}O labeling was found to have drawbacks as compared to *in vivo* methods including a poorer dynamic range in protein identification and a limit of detection—at best—in the high femtomole range [26].

The use of ^{18}O labeling in MS has not been limited to only proteins. In 2013, Hamasaki and coworkers used solid-phase synthesis to incorporate ^{18}O into the oligonucleotide to enable the MS-based study of oligonucleotide therapeutics [27]. Because these labeled standards are generated via solid-phase synthesis, they can be used as quantitative standards for a variety of pharmacokinetic and pharmacodynamics studies, in particular for oligonucleotide therapeutics like small interfering RNAs (siRNAs). The mechanism of action and drug clearance can be studied directly due to the mass label, which provides advantages over fluorescently tagged siRNAs that may not behave ideally due to the structural differences in the drug caused by the fluorescent tag.

4 ^{18}O Labeling of RNA—Early Applications for Mass Spectrometry

Learning from the field of proteomics, Beniam Berhane in our lab began investigating the applicability of enzyme-mediated labeling of nucleic acids using ^{18}O -labeled water. The initial studies focused on whether the similarity of enzyme mechanisms between proteases, such as trypsin, and nucleases, such as ribonuclease T1 (RNase T1), would enable a similar labeling method to be used for RNA (Fig. 1) [28]. Once it was found that RNase T1 could be used to incorporate ^{18}O onto the 3'-terminal phosphate of the oligonucleotide digestion product, this approach was exploited to simplify data interpretation in MALDI post-source decay (PSD) analysis of oligonucleotides [28].

Samples were digested in 50:50 light (^{16}O -labeled) and heavy (^{18}O -labeled) water to give the characteristic doublets for samples successfully digested bearing the 3'-phosphate group (Fig. 2). The doublet leads to simplified identification of products for further analysis. The only potential drawback was the need to use twice as much sample, because one was digested in light water at the same time as the other half of the sample was digested in the 50/50 mixture. This approach allowed for the spectra to be directly compared. Without the “normal” spectrum, it would be difficult to identify the +2 doublet of the ^{18}O -labeled digest.

Once we determined that certain RNases could be used to enzymatically label terminal phosphates with a single ^{18}O , Zhaojing Meng in the lab next turned to developing an approach for quantifying RNase digestion products [29]. This time samples were separately digested in ^{18}O -labeled and ^{16}O -labeled water. Method development was performed using commercially available *Escherichia coli* tRNA-Val to determine the effectiveness of this strategy for sample quantification (Fig. 3). The heavy and light digestion products were combined in ratios from 1:10 through 10:1 and analyzed using MALDI-MS. The averaged ion abundance ratio (heavy:-light) was plotted against the prepared sample ratio to generate a calibration curve.

From this study, it was determined that more accurate results were obtained when the ^{18}O -labeled sample was more abundant than the ^{16}O -labeled sample. When the ^{16}O -labeled sample is more abundant, other natural isotopes in the digestion product (e.g., ^{13}C , ^{15}N) can interfere with accurate determination of the ^{18}O -labeled peak abundance. By ensuring the more abundant sample is labeled with ^{18}O , those interfering isotope peaks from the ^{16}O -labeled sample are proportionally much less than the ^{18}O peak abundance, which minimizes errors in relative quantification. This information can guide the application of this approach in quantitative analysis. Accurate relative quantification required the generation of a calibration curve for each RNase digestion product of interest. To demonstrate the robust nature of this approach, a blinded analysis of heavy and light *E. coli* tRNA-Val mixtures was performed using the previously established calibration curve.

One of the more significant limitations of using enzyme-mediated labeling of RNA is that this approach requires complete enzymatic digestion of the RNA. As the mechanism involves a cyclic phosphate intermediate (Fig. 1), incomplete digestion products will not be labeled, significantly impacting the utility of this approach [28]. To circumvent this issue, higher amounts of RNases and a longer incubation time at an optimal temperature have been linked to a decrease in cyclic

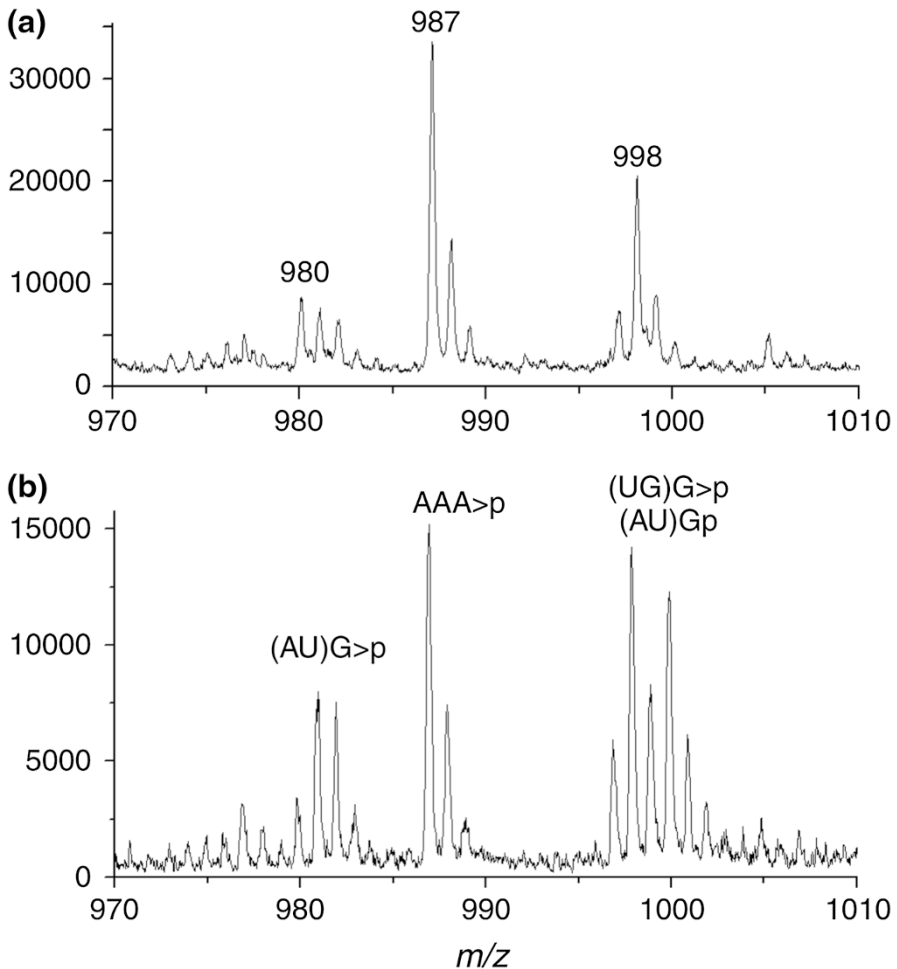


Fig. 2 Expanded view of MALDI mass spectral data obtained from the RNase T1 digestion products of *E. coli* 5S rRNA. **a** Digestion was done in unlabeled water. Three major ions are detected (m/z 980, 987 and 998). **b** After digestion in a 50:50 (v/v) mixture of unlabeled and ^{18}O -labeled water, only those oligonucleotide digestion products that contain a 3'-phosphate will exhibit the characteristic A + 2 doublet. From this, base compositions for the three ions can be made or confirmed. Reproduced with permission from Berhane et al. [28] Copyright 2003

phosphate digestion products [28, 30]. Another experimental challenge identified is the presence of sodium or potassium salt adducts to the RNase digestion products. These adducts can interfere with accurate detection and quantification, thus sample preparation and desalting are important to ensure accurate relative quantification when using MALDI-MS.

Having demonstrated the applicability of ^{18}O -based quantification of individual RNA samples, our interest next turned to using this approach to examine more complex mixtures of RNAs. Our specific interest was to characterize the total pool of tRNAs in a cell, which would obviate the need to individually purify tRNAs one

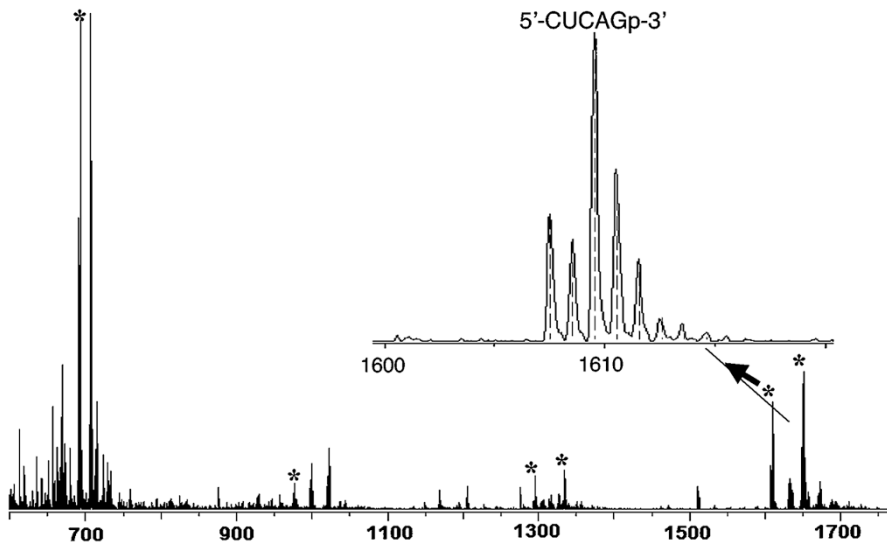


Fig. 3 Representative MALDI mass spectrum of RNase T1 digestion products obtained from a tRNA-Val mixture prepared at a heavy-to-light RNA ratio of 2:1. The asterisk denotes expected RNase digestion product pairs. *Inset:* expanded view of the RNase T1 digestion product 5'-CUCAGp-3' used for quantifying RNA levels for tRNA-Val with overlaid calculated isotopic distribution assuming a 2:1 ratio. Reproduced with permission from Meng et al. [29] Copyright 2005

by one from a sample. The analysis of total tRNA pools would not only decrease sample preparation time, it would also allow more information to be gained from a single MS experiment. When such a strategy is applied to the total cellular pool of tRNAs, information regarding codon usage and potential codon bias can be obtained in a more straightforward fashion [31].

The initial method developed to identify individual tRNAs within an unseparated mixture of total tRNAs was presented by Mahmud Hossain, who described the signature endonuclease digestion product (SDP) concept for tRNA identification using MALDI-MS [32]. In short, when one performs an *in silico* digest of known tRNA sequences (e.g., all *E. coli* tRNAs), the RNase digestion product masses that result will reveal that each individual tRNA will have at least one digestion product that is unique in both mass and sequence. Thus, these unique or *signature* digestion products can be used as a proxy to confirm the presence of any specific tRNA within the sample mixture.

Colette Castleberry built upon this SDP strategy by first demonstrating LC-MS/MS was just as effective at SDP identification as MALDI-MS [31]. In the same work, she then focused on combining RNase-mediated ^{18}O -labeling and the signature digestion product concept to create quantitative signature digestion products (qSDPs)—those SDPs that could simultaneously be used for tRNA identification and quantification. The criteria for defining a digestion product as a qSDP include incorporation of the ^{18}O label, a difference in mass by more than 2 Da from other known digestion products, and the labeled SDP must be able to provide a linear response spanning a fivefold change in SDP amount. By creating a set of

qSDPs, studies on how specific tRNA levels change as a result of culturing conditions were examined (Table 1).

5 18O Labeling of RNA for Modification Mapping by Mass Spectrometry

Our lab next turned the focus to how 3'-terminal phosphate labeling can enable alternative strategies to characterize RNA samples. Of particular interest to the lab is the discovery and characterization of post-transcriptionally modified nucleosides

Table 1 Analysis of RNase T1 quantitative signature digestion products from *E. coli*. Table reproduced with permission from Castleberry et al. [31]

tRNA	qSDP sequence	Experimental I_{18}/I_{16}	%CV
Decrease in relative abundance			
Cys	CA[ms ² i ⁶ A]AΨCCGp	0.75	19
Cys	U[s ⁴ U]AACAAAGp	0.67	25
Tyr 1, 2	ACUQUA[ms ² i ⁶ A]AΨCUGp	0.60	18
Increase in relative abundance			
Gly 1	AUCCCCUUCGp	1.44	26
Gly 2	CCU[Um]CCAAGp	1.28	24
Gly 3	AAUAGp	1.88	13
Ser 1, 4, 5	AAAAGp	2.43	18
Ser 1, 2	A[ms ² i ⁶ A]AACCGp	1.42	10
No change in relative abundance			
Ala 1, 2	[m ⁷ G]UCUGp	1.17	15
Arg 1, 2	[m ² A]ACCGp	1.04	20
Asn	UCCUCUGp	1.21	15
Glu 1, 2, 3	AAUCCCCUAGp	1.06	15
Glu 1, 2, 3	UCCCCUUCGp	1.20	15
Leu 1	UCCCCCCCUCGp	1.09	15
Phe	AA[ms ² i ⁶ A]AΨCCCCGp	1.31	13
Phe	A[s ⁴ U]AGp-3'	1.22	25
Phe	U[m ⁷ G][acp ³ U]CCUUGp	1.21	25
Ser 3	CUCCC[s ² C]UGp	1.00	17
Trp	UCUCUCCGp	1.21	27
Trp	U[Cm]UCCA[ms ² i ⁶ A]AACCGp	1.29	22
Val 1	AU[s ⁴ U]AGp	0.86	16
Indeterminate			
His	UU[m ⁷ G]UCGp	1.80	24
His	AAUCCCAUUAGp	1.36	21
His	[m ² A]ΨΨCCAGp	1.00	27
Ini 1, 2	TΨCAAAUCCGp	1.51	23
Ini 1, 2	[Cm]UCAUAACCCGp	0.97	41

in RNA samples. As noted earlier, RNA modification mapping by MS is an analytical approach that is used to identify the specific sequence location for modified nucleosides. Although several different strategies have been developed for modification mapping, it was thought that by using ^{18}O labeling, one could multiplex the analysis. This would reduce analysis time and cost as well as improving run-to-run reproducibility.

While ^{18}O labeling only enables duplex analysis (i.e., two different samples in a single analysis), it provides the template for even higher levels of multiplexing in the future. Our first ^{18}O labeled multiplexing investigation was described in 2012 by Siwei Li [33]. Comparative analysis of RNA digests (CARD) pairs a sample with known post-transcriptional modification with a sample of unknown post-transcriptional modifications. The idea is that any peaks appearing as doublets separated by 2 Da indicate that the digestion product from the “unknown” is identical to the digestion product of the “known” or reference sample. Thus, by examining all doublets one can quickly identify the similarities of RNA samples. By the same reasoning, digestion products appearing as singlets (either from the ^{16}O -labeled sample alone or the ^{18}O -labeled sample alone), inform one of differences between the two RNA samples. These differences could arise because the unknown sample is modified differently than the known sample or singlets could arise due to sequence differences in the two samples (Fig. 4).

If the known is truly “well characterized” in terms of post-transcriptional modification identity and pattern, the CARD approach would simply report sample equivalence through doublet identifications and sample differences through singlet identifications. Siwei first demonstrated the proof-of-concept studies by comparing a single purified tRNA from two bacteria [33], and then expanded this approach into an RNA modification mapping strategy for total tRNAs (Fig. 5) [34]. We found that this CARD approach is most effective when one uses reference and unknown samples (i.e., organisms) that share high sequence homology. By using phylogenetically related organisms, one minimizes the number of singlets that arise due simply to sequence differences so that mapping RNA modifications for the “unknown” sample is enhanced.

While CARD was found to significantly improve RNA modification mapping of total tRNA pools from organisms whose tRNA modification patterns were previously unknown, the minimal mass difference of the ^{18}O label (versus ^{16}O label) combined with interferences from naturally occurring stable isotopes (e.g., ^{13}C) limited our ability to generate automated methods for identifying singlets and doublets within the sample. To overcome this limitation, Collin Wetzel worked with Siwei to investigate culturing conditions that would minimize stable isotope interferences. We used ^{12}C -enriched medium during cell culturing to essentially eliminate ^{13}C (and ^{15}N) isotope interferences during CARD (Fig. 6) [35]. This culturing strategy leads to identification of singlets and doublets that can be automated due to improved differentiation of doublets. Moreover, this strategy can again be combined with the SDP approach to provide more targeted tRNA analysis.

Another limitation of the CARD strategy was identified by limiting singlet and doublet measurements to only mass measurements. RNase digestion products having the same mass but different sequences in the two samples could co-elute and

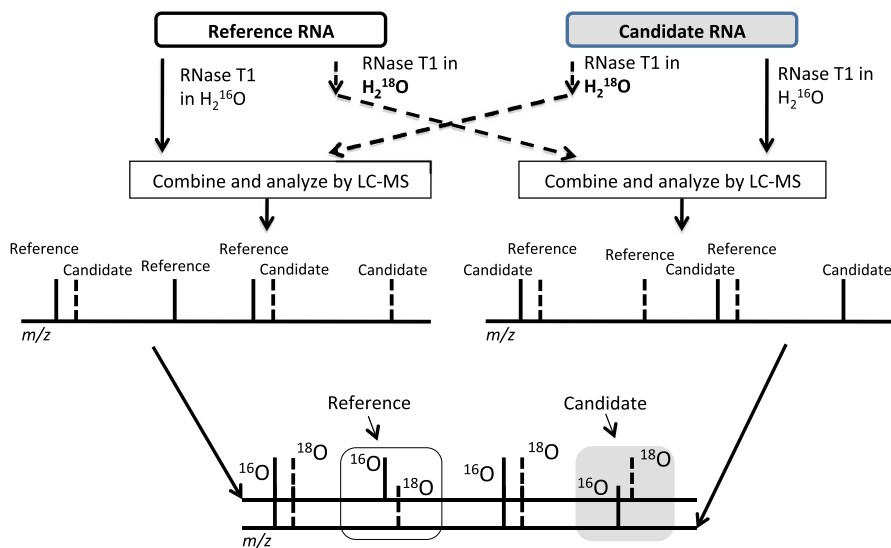


Fig. 4 Schematic outline of comparative sequencing by isotope labeling and LC-MS where *Escherichia coli* serves as the reference organism and *Citrobacter koseri* serves as the candidate (unknown) to be sequenced. tRNA endonuclease digestion products that are equivalent between organisms will appear as doublets (separated by 2 Da) in the mass spectral data; digestion products that are different between the two organisms will appear as a singlet. Reproduced with permission from Li et al. [33] Copyright 2012

appear as doublets, leading to a false positive in the analysis. To address this challenge, Siwei showed that ¹⁶O/¹⁸O-labeled digestion products can be differentiated based on MS/MS data [36], echoing our very first studies of ¹⁸O labeling during MALDI PSD experiments [28].

When oligonucleotides are fragmented during CID MS/MS, the most abundant fragment ions are the c-type and y-type ions, representing the oligonucleotide sequence from the 5'- and 3'-termini, respectively. The y-type ions contain the ¹⁶O or ¹⁸O label on the 3'-phosphate. Knowing this, Siwei demonstrated that the MS/MS data can be used to confirm that doublets detected in the mass spectrum are truly the same sequence rather than sequence isomers. Doublets detected in the y-type ions during MS/MS can only arise if the two sequences are identical. Sequence isomers are identified by singlets in the y-type ions, which occur wherever sequence differences are present in the original digestion products.

6 Future Outlook

As discussed above, phosphate labeling by using enzyme-mediated incorporation of ¹⁸O into RNase digestion products has been used in numerous ways to improve the MS-based characterization of modified RNAs. However, a fundamental limitation remains that the minimal mass difference between ¹⁶O and ¹⁸O limits the overall utility of this approach to those examples discussed previously. It will be of interest

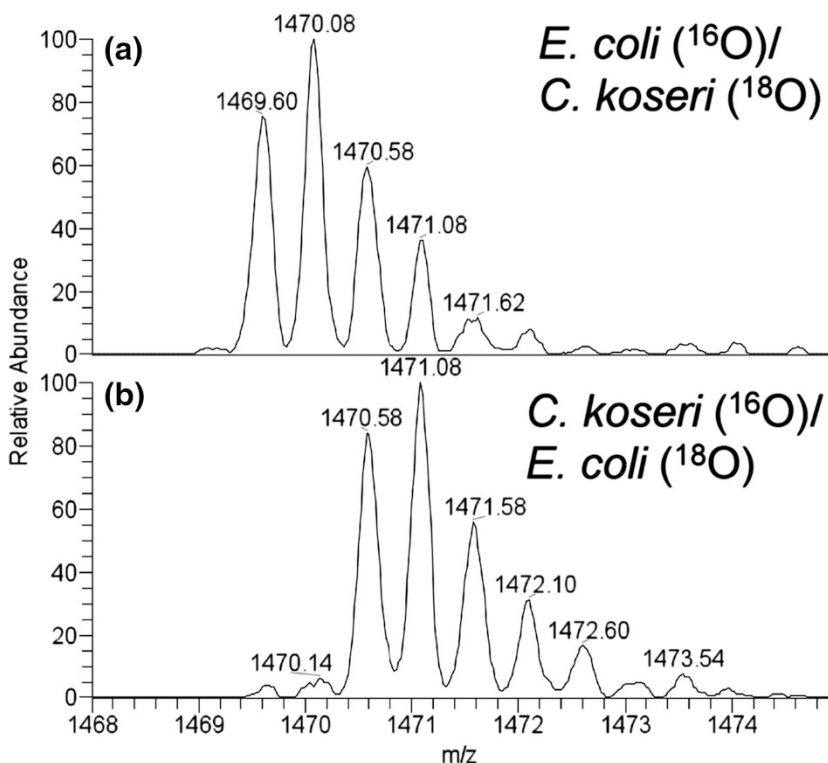


Fig. 5 Mass spectra corresponding to a detected singlet when **a** *C. koseri* is labeled with ^{18}O and **b** *E. coli* is labeled with ^{18}O . The singlet U[s ^{4}U]AACAAAGp (m/z 1469.6, 2-charge) arises from the *E. coli* tRNA-Cys(GCA) as confirmed by the +1 increase in the m/z isotopic envelope after ^{18}O -labeling of *E. coli*. Reproduced with permission from Li et al. [34] Copyright 2013

to examine alternative strategies for phosphate labeling, which are known in the field of RNA biology, but which have not yet entered the world of MS.

T1 ligase has been used in radiolabeling for visualizing RNA since the late 1970s [37, 38]. This ligase has not been explored in MS applications because it is known to generate a variety of different side-products. However, more recently, a T4 RNA ligase was created to reduce these unwanted ligation products [39]. With this advancement, T4 RNA ligase may become a more promising tool for stable isotope labeling of oligonucleotides, including RNase digestion products. Another potential enzyme for RNA labeling is Thg1 [40]. The role of Thg1 in the cell is to add a single guanosine nucleotide to the 5' terminus of tRNA-His. However, it has been shown that this enzyme has 3'-5' polymerase activity [41]. With additional study, it may be possible to use this unique function and activity to incorporate stable isotopes into specific RNA samples, which could be part of a broader MS strategy for sample characterization.

Another area where RNA MS in general, and RNA modification mapping in particular, can look to for inspiration and ideas for phosphate labeling and multiplexing strategies is the field of proteomics. A significant diversity of

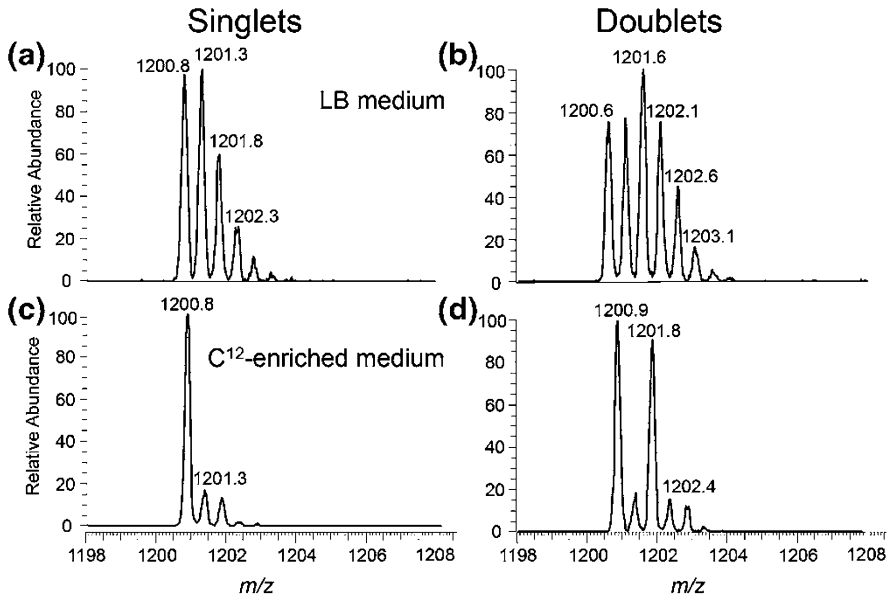


Fig. 6 Improvements in singlet and doublet identification using ^{12}C -enriched medium as illustrated with the doubly charged *E. coli* total tRNA RNase T1 digestion product $\text{A}[\text{ms}^{2,6}\text{A}]\text{AACCGp}$ (MW 2403.4 Da). **a** Mass spectrum from sample grown in rich medium and labeled with ^{16}O during RNase T1 digestion. **b** Same sample as in **a** except labeled with both ^{16}O and ^{18}O during RNase T1 digestion. **c** Mass spectrum obtained when sample grown in ^{12}C -enriched medium and labeled with ^{16}O during RNase T1 digestion. **d** Same sample as in **c** except labeled with both ^{16}O and ^{18}O during RNase T1 digestion. Singlet and doublet identifications are simplified in **c** and **d**, respectively, by use of ^{12}C -enriched medium. Reproduced with permission from Wetzel et al. [35] Copyright 2014

multiplexing strategies have been developed in proteomics [42]. Given the demonstrated advantages of relatively rapid characterization of multiple samples by these proteomics approaches, one can envision the development of tools and technologies for RNA MS that provide similar advantages, even if the particular chemistry and biochemistry may differ due to the unique characteristics of RNA. Regardless, MS as a platform for RNA analysis in general, and RNA modification mapping in particular, are now well-appreciated. It remains an ongoing challenge for the community to identify and develop the needed sample labeling tools to take full advantage of this platform.

Acknowledgements Financial support for our lab's decade-long work in the area of stable isotope labeling of RNA has been generously supported by the National Science Foundation, including our current NSF support (CHE1507357). The generous support of the University of Cincinnati and the Rieveschl Endowment for these studies are also appreciated.

References

1. Nachtergaele S, He C (2016) The emerging biology of RNA post-transcriptional modifications. *RNA Biol* 14(2):156–163

2. Song J, Yi C (2017) Chemical modifications to RNA: a new layer of gene expression regulation. *ACS Chem Biol* 12(2):316–325
3. Zhang X et al (2016) Small RNA modifications: integral to function and disease. *Trends Mol Med* 22(12):1025–1034
4. Kowalak JA et al (1993) A novel method for the determination of post-transcriptional modification in RNA by mass spectrometry. *Nucleic Acids Res* 21(19):4577–4585
5. Kowalak JA, Bruenger E, McCloskey JA (1995) Posttranscriptional modification of the central loop of domain V in Escherichia coli 23 S ribosomal RNA. *J Biol Chem* 270(30):17758–17764
6. Ni J et al (1996) Interpretation of oligonucleotide mass spectra for determination of sequence using electrospray ionization and tandem mass spectrometry. *Anal Chem* 68(13):1989–1999
7. McLuckey SA, Van Berkel GJ, Glish GL (1992) Tandem mass spectrometry of small, multiply charged oligonucleotides. *J Am Soc Mass Spectrom* 3:60–70
8. Kirpekar F, Douthwaite S, Roepstorff P (2000) Mapping posttranscriptional modifications in 5S ribosomal RNA by MALDI mass spectrometry. *RNA* 6:296–306
9. Meng Z, Limbach PA (2006) Mass spectrometry of RNA: linking the genome to the proteome†. *Brief Funct Genom* 5(1):87–95
10. Ross R et al (2016) Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. *Methods* 107:73–78
11. Limbach PA, Paulines MJ (2017) Going global: the new era of mapping modifications in RNA. *Wiley Interdiscip Rev: RNA* 8(1):e1367
12. Li X, Xiong X, Yi C (2016) Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat Methods* 14(1):23–31
13. Ong SE et al (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteom* 1(5):376–386
14. Ong SE, Mann M (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* 1(6):2650–2660
15. Brückl T et al (2009) Parallel isotope-based quantification of modified tRNA nucleosides. *Angew Chem Int Ed Engl* 48(42):7932–7934
16. Kellner S et al (2014) Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res* 42(18):e142
17. Waghmare SP, Dickman MJ (2011) Characterization and quantification of RNA post-transcriptional modifications using stable isotope labeling of RNA in conjunction with mass spectrometry analysis. *Anal Chem* 83(12):4894–4901
18. Popova AM, Williamson JR (2014) Quantitative analysis of rRNA modifications using stable isotope labeling and mass spectrometry. *J Am Chem Soc* 136(5):2058–2069
19. Taoka M et al (2015) A mass spectrometry-based method for comprehensive quantitative determination of post-transcriptional RNA modifications: the complete chemical structure of Schizosaccharomyces pombe ribosomal RNAs. *Nucleic Acids Res* 43(18):e115
20. Paulines MJ, Limbach PA (2017) Stable isotope labeling for improved comparative analysis of RNA digests by mass spectrometry. *J Am Soc Mass Spectrom*. doi:10.1007/s13361-017-1593-3
21. Sprinson DB, Rittenberg D (1951) Nature of the activation process in enzymatic reactions. *Nature* 167(4247):484
22. Desiderio DM, Kai M (1983) Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed Mass Spectrom* 10(8):471–479
23. Mirgorodskaya OA et al (2000) Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards. *Rapid Commun Mass Spectrom* 14(14):1226–1232
24. Yao X et al (2001) Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* 73(13):2836–2842
25. Qian WJ et al (2009) Large-scale multiplexed quantitative discovery proteomics enabled by the use of an (18)O-labeled “universal” reference sample. *J Proteome Res* 8(1):290–299
26. Lange S et al (2010) Identification of phosphorylation-dependent interaction partners of the adapter protein ADAP using quantitative mass spectrometry: SILAC vs (18)O-labeling. *J Proteome Res* 9(8):4113–4122
27. Hamasaki T et al (2013) Synthesis of (1)(8)O-labeled RNA for application to kinetic studies and imaging. *Nucleic Acids Res* 41(12):e126

28. Berhane BT, Limbach PA (2003) Stable isotope labeling for matrix-assisted laser desorption/ionization mass spectrometry and post-source decay analysis of ribonucleic acids. *J Mass Spectrom* 38:872–878
29. Meng Z, Limbach PA (2005) Quantitation of ribonucleic acids using 18-O labeling and mass spectrometry. *Anal Chem* 77:1891–1895
30. Hartmer R et al (2003) RNase T1 mediated base-specific cleavage and MALDI-TOF MS for high-throughput comparative sequence analysis. *Nucleic Acids Res* 31(9):e47
31. Castleberry CM, Limbach PA (2010) Relative quantitation of transfer RNAs using liquid chromatography-mass spectrometry (LC-MS) and signature digestion products. *Nucleic Acids Res* 38:e162
32. Hossain M, Limbach PA (2007) Mass spectrometry-based detection of transfer RNAs by their signature endonuclease digestion products. *RNA* 13(2):295–303
33. Li S, Limbach PA (2012) Method for comparative analysis of ribonucleic acids using isotope labeling and mass spectrometry. *Anal Chem* 84(20):8607–8613
34. Li S, Limbach PA (2013) Mass spectrometry sequencing of transfer ribonucleic acids by the comparative analysis of RNA digests (CARD) approach. *Analyst* 138(5):1386–1394
35. Wetzel C, Li S, Limbach PA (2014) Metabolic de-isotoping for improved LC-MS characterization of modified RNAs. *J Am Soc Mass Spectrom* 25(7):1114–1123
36. Li S, Limbach PA (2014) Identification of RNA sequence isomers by isotope labeling and LC-MS/MS. *J Mass Spectrom* 49(11):1191–1198
37. Bruce AG, Uhlenbeck OC (1978) Reactions at the termini of tRNA with T4 RNA ligase. *Nucleic Acids Res* 5(10):3665–3677
38. England TE, Bruce AG, Uhlenbeck OC (1980) Specific labeling of 3' termini of RNA with T4 RNA ligase. *Meth Enzymol* 65(1):65–74
39. Viollet S et al (2011) T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. *BMC Biotechnol* 11:72
40. Abad MG, Rao BS, Jackman JE (2010) Template-dependent 3'-5' nucleotide addition is a shared feature of tRNAHis guanylyltransferase enzymes from multiple domains of life. *Proc Natl Acad Sci USA* 107(2):674–679
41. Jackman JE, Gott JM, Gray MW (2012) Doing it in reverse: 3'-to-5' polymerization by the Thg1 superfamily. *RNA* 18(5):886–899
42. Timms JF, Cutillas PR (2010) Overview of quantitative LC-MS techniques for proteomics and activitomics. *Methods Mol Biol* 658:19–45

New Synthetic Methods for Phosphate Labeling

Amit K. Dutta¹ · Ilya Captain¹ · Henning Jacob Jessen¹

Received: 21 November 2016 / Accepted: 27 March 2017
© Springer International Publishing Switzerland 2017

Abstract The complexity of phosphorylation pathways and their downstream effects is vast. Synthetic chemistry has been working side by side with biology to develop phosphate labels for biological processes involving phosphorylated compounds. This chapter discusses recently employed methods for the preparation of several phosphate labels. Synthesis of biomolecules and their analogs and other useful or potentially useful phosphate derivatives is discussed.

Keywords Phosphorylation · Phosphitylation · Phosphate Analogs · Organophosphate

Abbreviations

A	Adenosine
Ade	Adenine
ATP	Adenosine triphosphate
B	Nucleobase
BMF ⁴ TPA	Bis(difluoromethylene)triphosphoric acid
BMT	Bismethylene triphosphate
Boc	<i>Tert</i> -Butyloxycarbonyl
Bop	Bis(2-oxo-3-oxazolidinyl)phosphinic
B ^P	Protected nucleobase
BTT	5-Benzylthio-1- <i>H</i> -tetrazole
C	Cytosine

This article is part of the Topical Collection “Phosphate Labeling and Sensing in Chemical Biology”; edited by Henning Jessen.

✉ Henning Jacob Jessen
henning.jessen@oc.uni-freiburg.de

¹ Institute of Organic Chemistry, Albertstrasse 21, 79104 Freiburg im Breisgau, Germany

Published online: 25 April 2017

Reprinted from the journal

105

 Springer

CDI	Carbodiimidazole
CE	β -Cyanoethyl
CEM	Cyanooxymethyl
CMPT	<i>N</i> -(cyanomethyl)pyrrolidinium triflate
CPG	Controlled pore glass
CTP	Cytidine triphosphate
Cyt	Cytidine
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCA	Dichloroacetic acid
DCI	4,5-Dicyanoimidazole
DEAE	Diethylaminoethyl
DIAD	Diisopropyl azodicarboxylate
DIPEA	Diisopropylethylamine
DMAN	1,8-Bis-(dimethylamino)naphthalene
DMF	<i>N,N</i> -dimethylformamide
DMS	Dimethylsulfide
DMTr	4,4'-Dimethoxytrityl
DTD	<i>N,N</i> -dimethylthiuram disulfide
EC ₅₀	Half maximal effective concentration
EDC/EDCI	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
ETT	5-(Ethylthio)-1H-tetrazole
Fm	9-Fluorenylmethyl
Fmoc	Fluorenylmethoxycarbonyl
G	Guanosine
Gua	Guanine
IC ₅₀	Half maximal inhibitor concentration
IEX-HPLC	Ion-exchange high performance liquid chromatography
KHMDS	Hexamethyldisilazide
LTMPA	Lithium 2,2,6,6-tetramethylpiperidine amide
NHS	<i>N</i> -hydroxysuccinimide
NMP	Nucleoside monophosphate
Np _n	Nucleoside polyphosphate
Np _n N	Dinucleotide polyphosphate
NPP	Nucleotide pyrophosphatase/phosphodiesterase
Ns	Nosyl
NTP	Nucleoside triphosphate
NTP	Nucleoside triphosphate
Nuc	Nucleotide or nucleoside
ODN	Oligodeoxynucleotides
ORN	Oligoribonucleotide
OTP	Oxathiaphospholane
PEP	Phosphoenolpyruvate
Pip	Piperidine
PK	Pyruvate dinase
ppGpp	Guanosin-3',5'-bispyrophosphate

ppp	RNA 5'-triphosphate RNAs
PRR	Pattern recognition receptors
Py	Pyridine
RP18	Reverse phase C ₁₈
RSH	RelA-SpoT homolog
SAX	Strong anion exchange
T	Thymine
TBAF	Tetrabutylammonium fluoride
TBHP	<i>tert</i> -Butylhydroperoxide
TBS	<i>tert</i> -Butyldimethylsilyl
TEA	Triethylamine
TEAB	Triethylammonium bicarbonate
Tf	Trifluoromethylsulfonyl
THF	Tetrahydrofuran
Thy	Thymidine
TMS	Trimethylsilyl
Tr	2,4,6-Triisopropylbenzenesulfonyl
Ts	<i>p</i> -Toluenesulfonyl
U	Uridine
Ura	Uracil
UTP	Uridine triphosphate

1 Introduction

The focus of this chapter is on new and recent phosphate labeling methods. The term “phosphate label” can be interpreted in many ways; in the purest sense it is a chemically indistinguishable phosphate unit that can be tracked with some sort of readout. This is possible using isotopic labeling or radioactive tracers, but the scope of these highly useful tools is narrow. We apply a more liberal definition of “phosphate labeling” and describe a range of phosphate derivatives, analogs, and bioisosteres [1]. Specifically, we focus on chemical substitution around the

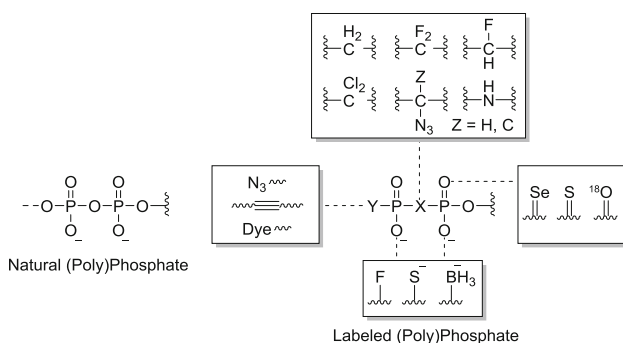


Fig. 1 Overview of labeled phosphates. X, Y = O, N, S

phosphorus center and its neighbors. These modifications can be invaluable when properties such as hydrolytic stability, external monitoring, and enzyme inhibition are desired.

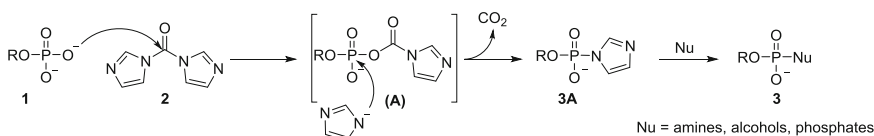
Figure 1 outlines some modifications that we discuss; separate reviews on each specific modification or label type may exist and are referenced when available. The aim of this chapter is to summarize the most recent developments in the field (ca. ~5–10 years), but earlier methods are discussed if they have not been reviewed before this soft cutoff. Finally, only structures that bear resemblance to the natural phosphate and could potentially be used in a biological setting are reviewed, though this classification is by no means ironclad.

2 General Modes of Activation and Coupling

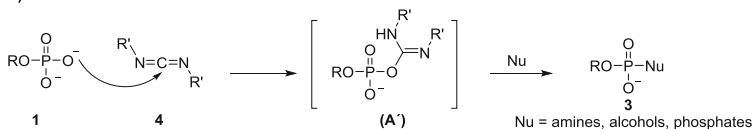
This chapter primarily focuses on two major modes of phosphate **1** activation and/or coupling (Fig. 2). In P(V) state, tri- or tetra-alkyl ammonium salts of phosphoric acids are activated by carbonyldiimidazole **2** (CDI) or various carbodiimides **4** [2] to form P(V)-imidazolide **3A** or O-phosphorylated urea intermediate (**A**), respectively. Relative to unmodified phosphates, these species are much more susceptible to nucleophilic attack affording **3**. In the P(III) state (**5**, Fig. 2), differently substituted tetrazoles or imidazoles (Activator) protonate and replace the amino moiety to form highly reactive P(III) intermediates of general formula (**B**), containing tetrazole or imidazole moiety [3]. Subsequent treatment with alcohols or phosphates leads to

Activation at the P(V) State

1) Carbonyldiimidazole Activation



2) Carbodiimide Activation



Activation at the P(III) State

1) P-amidite based

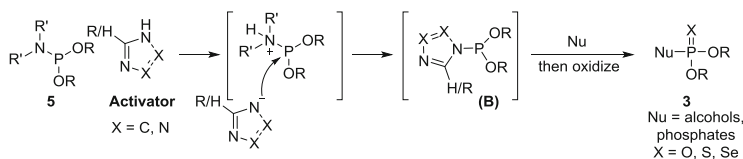


Fig. 2 Common P(V)/P(III) activation modes

P(III) or mixed P(III)–P(V) intermediates, respectively. The P(III) center is then oxidized to form substituted P(V) compounds **3**. P(III) reagents are generally more active than their P(V) counterparts, requiring shorter reaction times. These two activation modes are often mentioned as P(V)- and P(III)-standard protocols, respectively, in the subsequent text.

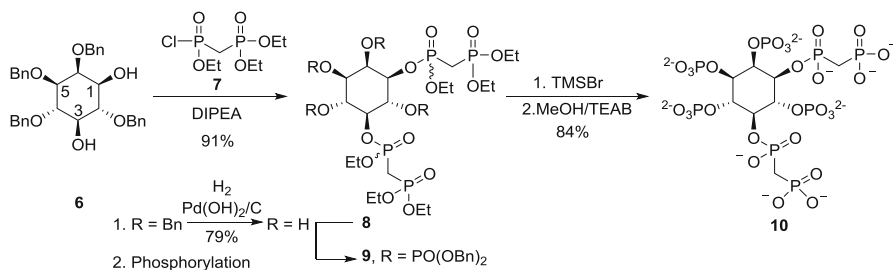
3 Non-hydrolyzable Phosphate Analog Synthesis

Condensed phosphates play a significant role in biological processes such as cellular signaling and metabolism through phosphorylation [4]. Phosphoanhydrides are thermodynamically inclined toward hydrolysis and are referred to as “high energy bonds” in biological settings. Non-hydrolyzable analogs provide useful chemical tools to study phosphoryl transfer mechanisms as they are unable to do such transfers. Hydrolytically stable bioisosteres containing a P–CH₂–P unit are the simplest analogs of P–O–P motifs.

3.1 P–CH₂–P Motifs

The *myo*-inositol phosphates (InsPs) or inositol phosphates are signaling molecules involved in cellular processes such as vesicle trafficking, cell motility and proliferation, apoptosis, transcriptional regulation, cancer development, diabetes, and obesity [5–9]. The non-hydrolyzable diphosphoinositol polyphosphates allow for differentiation between two distinct signaling pathways: protein binding and β -phosphoryl transfer. The bridging –CH₂– unit introduces only minimal structural perturbation; hence, signaling via protein binding remains unaffected, but the analogs are unable to participate in β -phosphoryl group transfer.

One synthetic example of non-hydrolyzable inositol phosphates [10] 1,5-(PCP)₂-InsP₄ starts with enantiomerically pure **6** (Scheme 1) [11]. Compound **6** was reacted with phosphonochloridate **7** at its two free hydroxyl groups providing **8** as a mixture of four diastereomers. Removal of benzyl-protecting groups liberated the 4 hydroxyls, which underwent the standard P(III)-tetrazole protocol affording **9**. Subsequent TMSBr deprotection followed by a methanol treatment produced the desired non-hydrolyzable InsP₈ analog **10**.



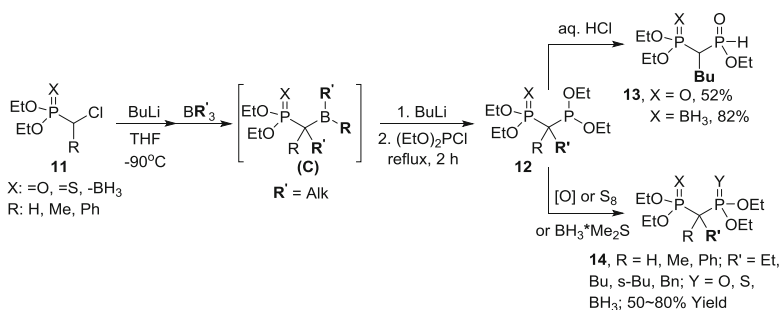
Scheme 1 Synthesis of non-hydrolyzable analogs of inositol polyphosphate **10**

The P–C–P subunit can principally be diversified into a broader range of 1,1-bisphosphorus compounds with regard to the oxidation state at phosphorous; one such method proceeds via organoborane intermediates while the other takes advantage of a modified Claisen condensation. The deprotonation of precursors **11** gives carbenoid intermediates that can add trialkylboranes to produce intermediates of type (C) [12], which upon addition of an extra equivalent of BuLi and (EtO)₂PCl yields compounds of type **12** [13]. **12** is then hydrolyzed to mixed phosphonate–phosphinate **13** using HCl. Alternatively, its oxidation with S₈ (**14**, X = S) or cumene hydroperoxide (**14**, X = O) or the functionalization with a borane source (**14**, X = BH₃) affords pyrophosphate analogs of type **14** (Scheme 2).

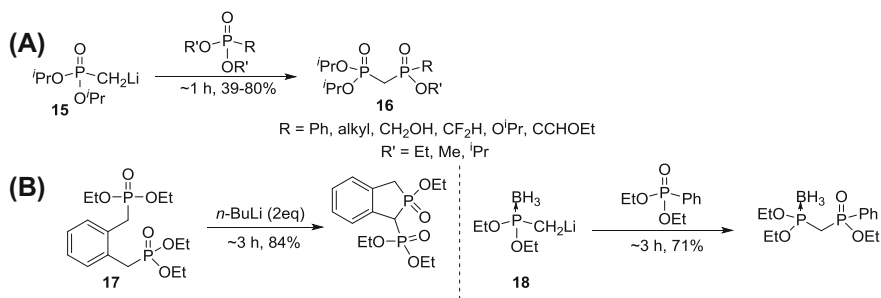
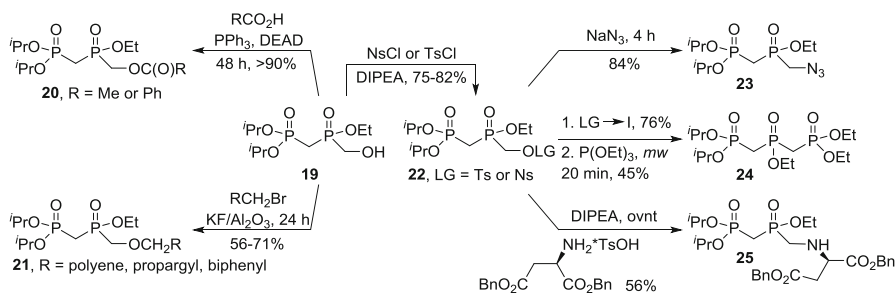
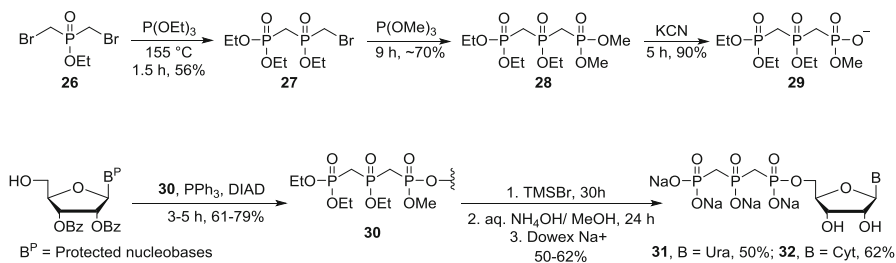
Bisphosphonates and their analogs may also be prepared using a combination of a phosphonate C-anion and a phosphonate electrophile. Phosphonate (iPrO)₂P(O)CH₃ forms a phosphonate nucleophile **15** upon treatment with *n*-BuLi at low temperature, which reacts with a number of phosphonate electrophiles to produce bisphosphonates of type **16** (Scheme 3a). This synthetic protocol can be also used for the preparation of phosphate–phosphonate derivatives. The scope of these transformations has been further explored for substrates like **17** for intramolecular phospho-Dieckmann reactions. Implementation of other nucleophiles such as boronated phosphinate analogs **18** were also successful (Scheme 3b) [14].

The phospho-Claisen condensation product **19** (see Scheme 3, **16**, R = CH₂OH) was used as a starting material for a family of non-hydrolyzable analogs (Scheme 4). The hydroxyl group of **19** can be activated using Mitsunobu conditions to yield esters **20** or potassium fluoride/alumina to yield ethers **21**. Standard conditions can be applied to convert the free OH into leaving group (LG) bearing compounds **22**, which react with NaN₃, P(III) nucleophiles, or primary amines affording compounds **23**, **24**, and **25**, respectively (Scheme 4) [15].

Bearne and coworkers [16] reported synthetic bismethylene triphosphate (BMT) analogs of UTP and CTP that contain a nonhydrolyzable P–C–P–C–P phosphonate unit (Scheme 5). The unsymmetrical bismethylene triphosphate analog **29** was synthesized starting from ethyl bis(bromomethyl)phosphinate **26** via sequential Michaelis-Arbuzov reactions [17], using first triethylphosphite to yield **27** and then trimethylphosphite to yield **28**. The methoxy group was then selectively mono-deprotected at the terminal phosphonate moiety using KCN to produce **29**.

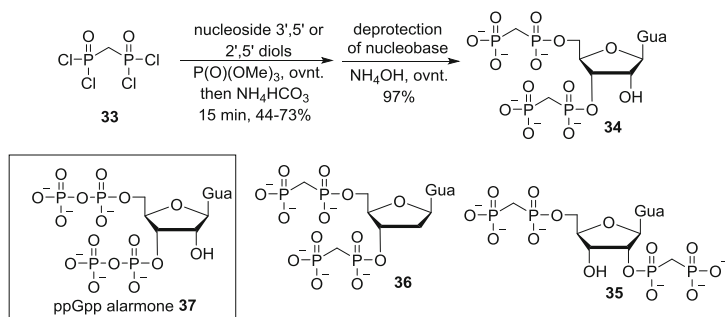


Scheme 2 Trialkylborane-mediated synthesis of C-substituted P–C–P compounds


Scheme 3 Bisphosphonates and analogs via phospho-Claisen condensation

Scheme 4 Further reactions of bisphosphonate analogs

Scheme 5 Synthesis of P-C-P-C-P non-hydrolyzable nucleotides

Nucleosides and **30** were then condensed at the 5'-position using a Mitsunobu reaction to afford protected NTP analogs **30**. Global deprotection was achieved using TMSBr followed by treatment with aqueous NH₄OH in methanol. The crude products were further purified by RP-HPLC and converted into corresponding Na⁺ salts to afford pure non-hydrolyzable NTP analogs **31** and **32**.

Alarmones, such as ppGpp **37** [18], are synthesized by a family of RelA-SpoT enzymes and accumulate during the bacterial stress response. Analogs **34**, **35**, and **36** are competitive inhibitors of Rel-SpoT in vitro with **36** showing the highest IC₅₀ (~1 mM). Their syntheses commenced with double bisphosphonylation of protected 5'- or 3'-guanosine diol and protected 5'-2'-guanosine diol. The phosphonylation was carried out in the presence of methylene-bis-(phosphonic



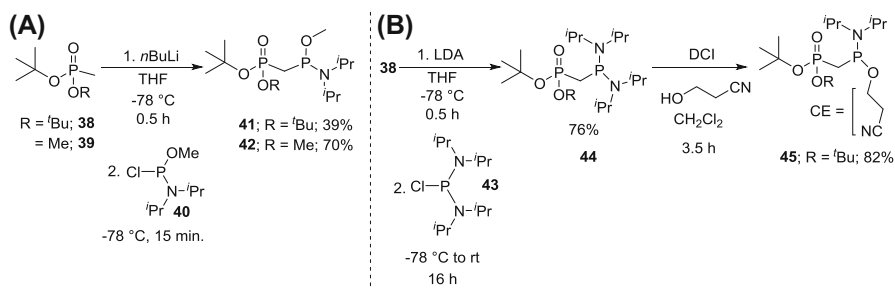
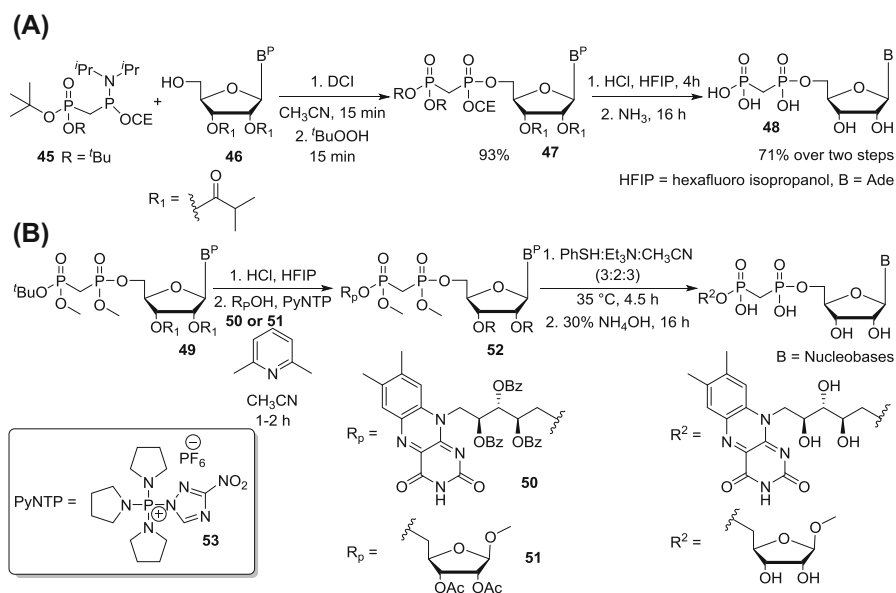
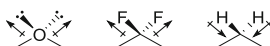
Scheme 6 Synthesis of ppGpp alarmone analogs

dichloride) **33** and trimethyl phosphate as solvent yielding the corresponding regioisomeric bisphosphonate analogs. Deprotection of the nucleobase by treatment with ammonium hydroxide gave **34** and **35**. The same bisphosphonylation protocol was applied to synthesize 2'-deoxyguanosine 5'-3'-di(methylene bisphosphonate) analog **36** starting from unprotected deoxyguanosine [19] (Scheme 6).

Mixed phosphoramidite-phosphodiester reagents **41**, **42**, and **45** were recently synthesized in two steps from methylphosphonates **38** and **39** [20]. (1) Lithiation of **38** and **39** followed by (2) an electrophilic quenching with chlorophosphine **40** produced **41** and **42** (Scheme 7a). A similar reaction between **38** and bis(diisopropylamino)chlorophosphine **43** produced reagent **44**. Further activation of one of the diisopropylamine groups was possible, converting it into same type of reagent as **41** and **42** (**45**, Scheme 7b), but with a different pattern of protecting groups.

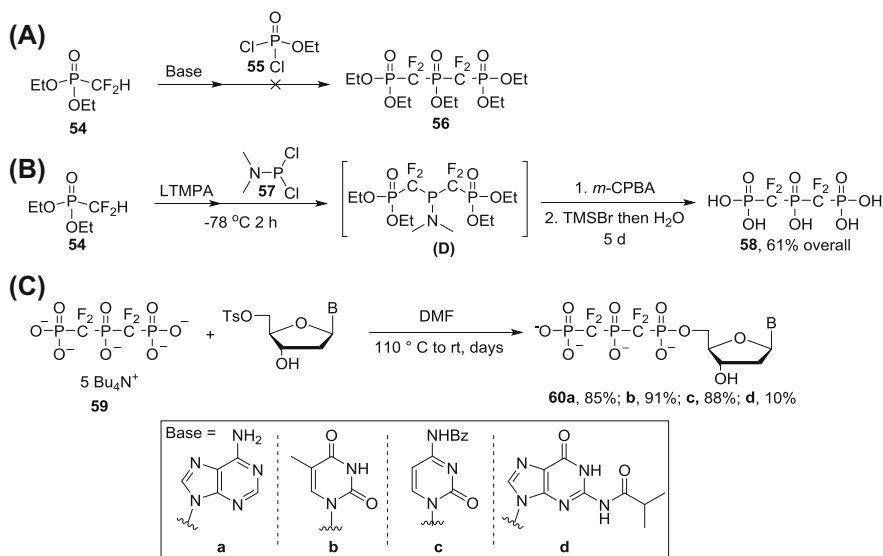
The synthetic scope of **41**, **42**, and **45** was further explored in the presence of an activator (DCI) and nucleophiles, such as adenosine derivative **46**. Subsequent oxidation with *tert*-butyl hydroperoxide gave product **47** (reaction time: 30 min, 93% over two steps). Removal of the protecting groups released **48**, a non-hydrolyzable methylene bisphosphonate monoester analog of ADP, with 66% overall yield (4 steps, Scheme 8a). Furthermore, when the strategy was transferred into an automated solid-phase oligonucleotide synthesizer, methylene bisphosphonate analogs of all four nucleosides were isolated efficiently with 39–52% of overall yields (four cycles).

Moreover, the terminal methylene bisphosphonates were used for additional phosphorylations (via the P(III)-amidite protocol) as well as condensations, forming either triphosphate analogs or non-symmetrical diesters, respectively. Selective deprotection of *tert*-butyl protecting group of **49** and successive condensation of the intermediate with alcohols, such as flavin adenine dinucleotide (FAD) derivative **50** and adenosine bisphosphonate ribose (ADPR) derivative **51**, were also successful, yielding products of type **52**. The condensation reactions were particularly successful in the presence of condensation reagent 3-nitro-1,2,4-triazol-1-yl-tris(pyrrolidin-1-yl)phosphoniumhexafluorophosphate [21] (PyNTP) **53** (Scheme 8b).


Scheme 7 Synthesis of phosphoramidite-phosphodiesters

Scheme 8 Synthesis of nucleoside methylene bisphosphonate monoester and diester analogs

Fig. 3 Dipoles in P–O–P and non-hydrolyzable P–C–P motifs

3.2 P–CF₂–P Analogs

Like CH_2 , CF_2 analogs are hydrolytically stable and possess a very similar steric perturbation profile. Since fluorine is more electronegative than carbon, the direction of dipoles in the CF_2 bridging of P– CF_2 –P motif is the same as in condensed phosphates. Hence, it is often referred to as the electronically “correct” non-hydrolyzable motif, unlike CH_2 in which the resulting direction of the dipole is reversed (Fig. 3) [22].



Scheme 9 Synthesis of P-CF₂-P-CF₂-P non-hydrolyzable NTP analogs

Several methods for incorporation of bis-difluoromethylene units into biomolecules exist, and a number of them have been reviewed [23, 24]. Like nucleoside bismethylene triphosphates **31** and **32** (Scheme 5), bis(difluoromethylene)triphosphate containing NTP analogs have also been described. The synthesis of pentaethyl protected bis(difluoromethylene)triphosphoric acid **56** (BMF⁴TPA) [22] via a double substitution reaction between two equivalents of the (diethylphosphinyl)difluoromethyl anion (after deprotonation of **54**) and one equivalent of dichlorophosphate **55** was unsuccessful (Scheme 9a). However, use of a sterically hindered base, lithium 2,2,6,6-tetramethylpiperidine amide (LTMPA), and a P (III) electrophile **57** giving intermediate (**D**), followed by in situ oxidation and deprotection yielded the P-CF₂-P-CF₂-P motif **58** (Scheme 9b).

58 can be incorporated into nucleosides to produce nonhydrolyzable NTP mimics **60a-d**. This is accomplished using a 5'-tosylate substitution approach with the tetrabutylammonium salt **59** as nucleophile (Scheme 9c). The corresponding tributylammonium salt failed in the same reaction presumably because of the presence of intramolecular hydrogen bonding that weakens the nucleophile [22].

3.3 P-C(XY)-P (X ≠ Y) Analogs Containing an Asymmetrical Carbon Atom

Analogs containing the PC(XY)P motif have been overlooked many years because of the synthetic challenges associated with stereochemical control. Recent X-ray crystallographic studies display stereospecific interactions between the F atoms of β,γ-CXY dGTP analogs (X = F; Y = H, Cl, Me) and Arg183 in the active site of DNA-pol β and therefore provide a model system that allows a better understanding

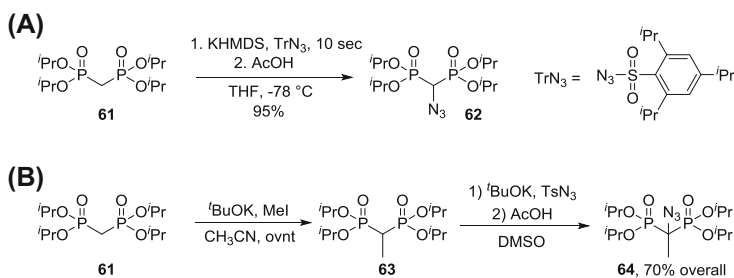
of polymerase fidelity [25]. First, we describe examples of synthetic procedures giving racemates and then discuss stereoselective approaches targeting the PC(XY)P motifs.

In general, sulfonyl azides react with carbanions to form triazene intermediates, which rearrange into either azido or diazo compounds, a rearrangement that is determined by the nature of sulfonyl azides. Bulky electron-donating azides, such as trisyl azide (TrN_3), favor electrophilic azido transfer over diazo transfer [26, 27]. On the contrary, electron-withdrawing azides such as trifluoromethylsulfonyl azide (TfN_3) preferably transfer the diazo group [28]. This reactivity difference has been successfully applied in the syntheses of two novel azidobisphosphonates **62** and **64**.

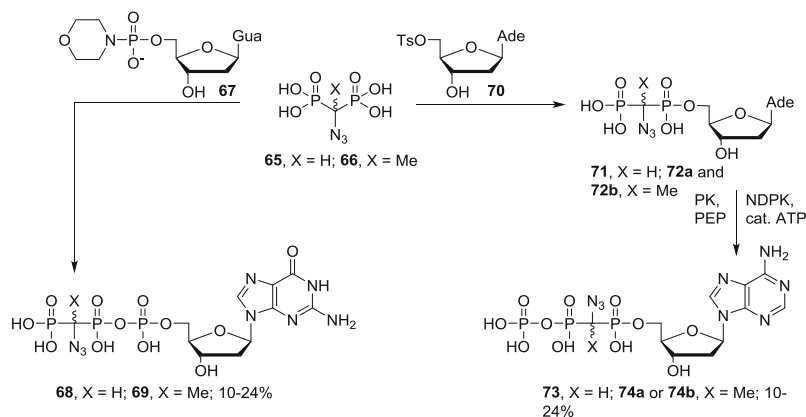
Trisyl azide selectively forms α -azido bisphosphonate ester **62** (95%) when treated with deprotonated bisphosphonate **61** (Scheme 10a) [29]. Replacement of an α -H of tetraisopropyl ethylenedibis(phosphonate) **61** by a methyl group as in intermediate **63** also generates predominantly the α -azido derivative **64** when exposed to similar reaction conditions but only with *p*-toluenesulfonyl azide as reagent (tosyl azide; TsN_3) (Scheme 10b).

Compounds **62** and **64** are readily converted into the corresponding α -azido bisphosphonic acids **65** and **66** using TMSBr followed by aqueous ethanol workup (>98% overall yield). Bu_3NH^+ salts of **65** and **66** are coupled to a dGMP-morpholidate **67** to generate β,γ -nonhydrolyzable azidomethylene dGTP analogs **68** and **69**, obtained as diastereomeric mixtures. A nucleophilic substitution reaction between the Bu_4N^+ salt of **65** and dA-5'-tosylate **70** provided inseparable diastereomers of α,β -azidomethylene dADP **71**. A similar reaction between **66** and **70** produced diastereomers **72a–b**, which were then separated by RP-HPLC. Subsequent enzymatic phosphorylation of diastereomers **72a** and **72b** with ATP and nucleoside diphosphate kinase (NDPK) [30] gave diastereopure α,β -nonhydrolyzable azidomethylene dATP analogs **74a** and **74b**. Phosphoenolpyruvate (PEP) and pyruvate kinase (PK) regenerated ATP from ADP into the cycle requiring ATP only in catalytic amounts to carry out this transformation. A mixture of the α,β -nonhydrolyzable azidomethylene dATP diastereomers **73** was obtained from **71** (Scheme 11) under the same conditions. All final dNTP analogs were purified by SAX chromatography and RP-HPLC.

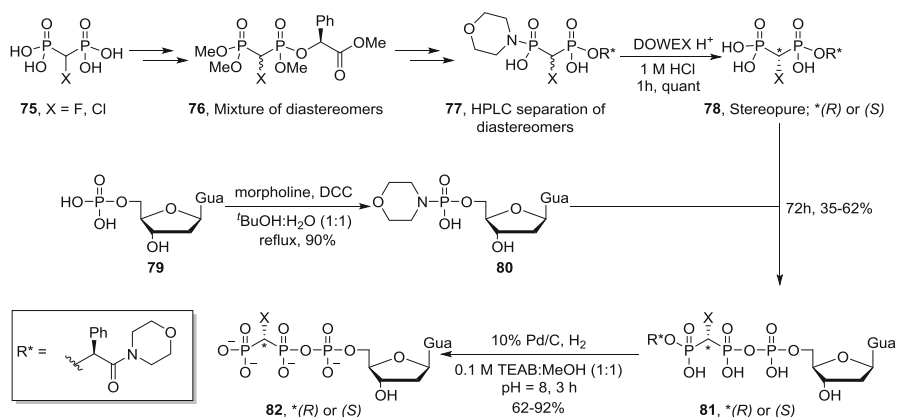
More recently, four diastereopure conjugates β,γ -CHX-dGTP (X = F, Cl) were prepared using a new chiral auxiliary strategy (Scheme 12) [31]. α -Halo



Scheme 10 Synthesis of azido-bisphosphonates



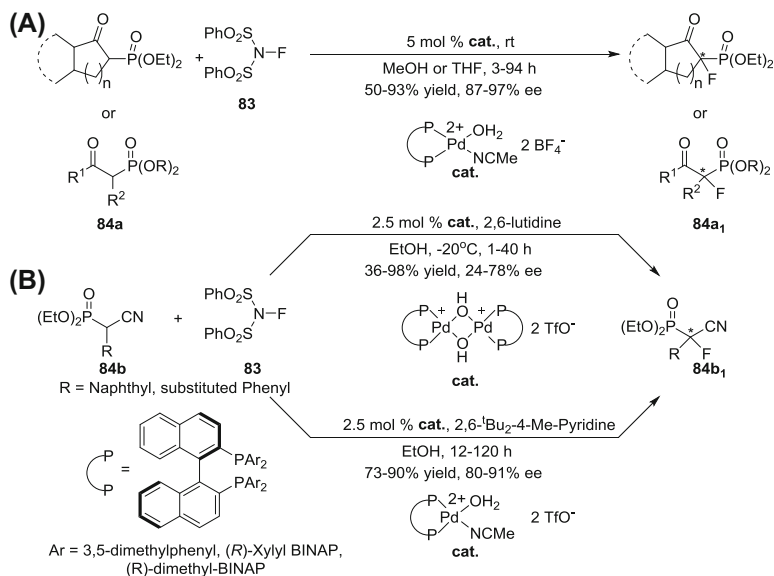
Scheme 11 Synthesis of non-hydrolyzable azido containing NTP analogs



Scheme 12 Synthesis of enantiopure β,γ -CHX(X = Cl, F)-dGTP analogs

bisphosphonates **75** were subjected to a series of protection/deprotection reactions followed by a Mitsunobu esterification with chiral auxiliary (R)-(-) methyl mandelate giving rise to a mixture of inseparable diastereomers **76**. Diastereomers **76** were further elaborated to **77**, and at this stage the diastereomers were separated by preparative RP-HPLC. The morpholidate group was then removed using DOWEX H⁺ affording the coupling partner R- or S-**78**. The complementary coupling partner **80** was prepared by condensation of deoxyguanosine monophosphate **79** with morpholine using DCC as activator. The coupling partners **78** and **80** were conjugated to yield the dNTP analogs **81** modified with a chiral auxiliary. The auxiliary was removed using hydrogenolysis to afford the enantiomerically pure dNTP analogs **82**.

The groups of Kim and Sodeoka have pioneered catalytic enantioselective electrophilic α -fluorination of several β -ketophosphonates of type **84** using *N*-fluoro-benzenesulfonamide **83** as the fluorinating agent [32, 33]. Using slightly



Scheme 13 Catalytic enantioselective fluorination of substituted phosphonates

varied conditions and chiral Pd-catalysts, they independently described fluorinations of substituted phosphonates in a highly enantioselective fashion (Scheme 13a).

The method was extended to α -cyanophosphonates **84a** (Scheme 13b) [34, 35]. These reactions require the addition of base; Sodeoka reported the application of 2,6-lutidine, whereas Kim used 2,6-*tert*-butyl-4-methylpyridine, with minor differences in ligand selection. The highest enantioselectivity (78% ee) was observed by Sodeoka's group at -20°C , whereas Kim's group isolated the products with >80% enantioselectivity at room temperature. Both groups report that aliphatic substitution (instead of aromatics) at the α -position precludes the desired transformation. Overall, the Pd catalyst used by Kim results in higher enantioselectivity. These strategies open up possibilities for future stereoselective fluorinations in the context of phosphate labeling.

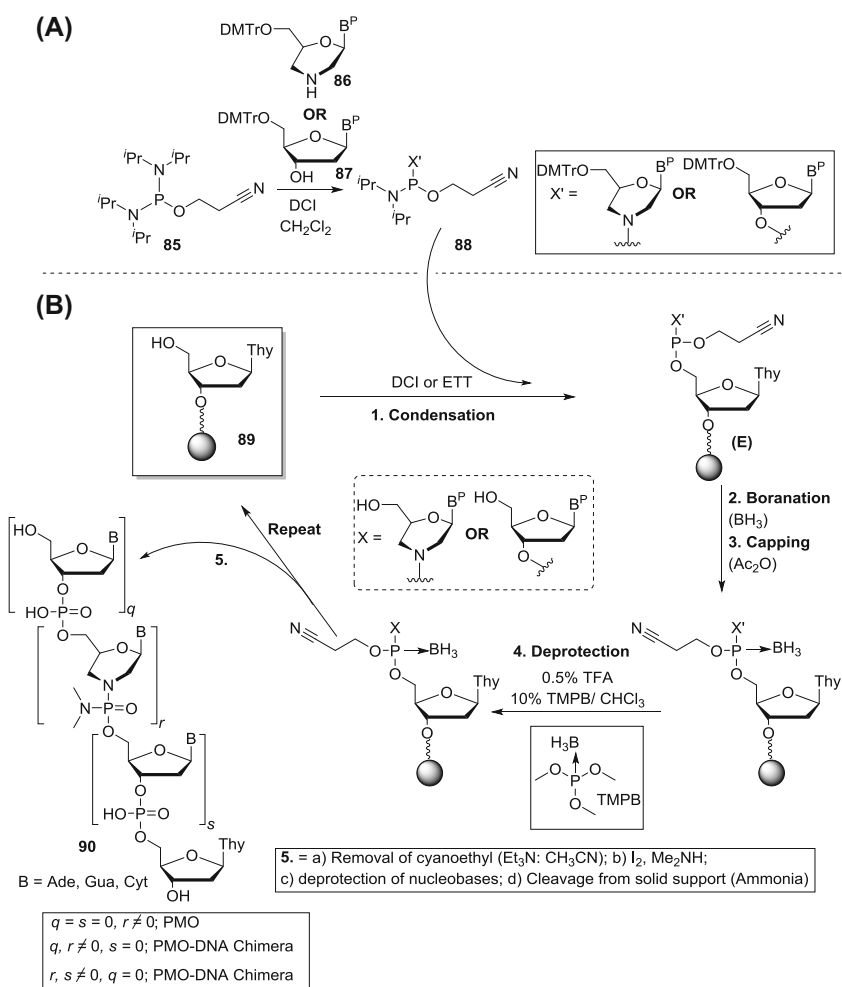
4 Phosphoramidates

The phosphoramidates (see also thiophosphoramidates; Sect. 5.6 of this chapter) are structural motifs where one amine functional group is directly attached to the phosphorus atom of a phosphate moiety (instead of one of its oxygen atoms). Numerous phosphoramidates have been synthesized, isolated, and documented for their diverse roles of biological interests, such as antiviral or antitumor activities. The pharmacology and medicinal chemistry of phosphoramidates have been comprehensively reviewed [36].

In a recent report, Caruthers and coworkers demonstrated an automated solid phase synthesis that facilitates unprecedented synthetic access to several

phosphorodiamidate morpholino oligonucleotides (PMO oligonucleotides) and PMO-DNA chimeras [37]. Previously reported solid phase syntheses of PMO primers were only possible to extended from the 5' to 3' direction and hence could not be automatized on a DNA synthesizer.

To obtain PMO oligonucleotides and PMO-DNA chimeras, first, morpholine and nucleoside phosphoramidites of type **88** were prepared via phosphitylation of morpholine and nucleoside derivatives **86** and **87**, respectively. The phosphitylation reaction was carried out by activating the P(III) center of diamidite **85** in the presence of DCI and a nucleophile (Scheme 14a). To couple phosphoramidites of type **88** with a 5'-OH-2'-deoxyribothymidine (attached to a polystyrene support at the 3'-position) **89**, the P(III) activation strategy was employed. The choice of the activator played a crucial role as more acidic 5-(ethylthio)-1H-tetrazole (ETT);



Scheme 14 Automated synthesis of PMOs and PMO-DNA chimeras

$pK_a = 4.3$) not only activated the diisopropylamine leaving group but also the morpholine moieties, resulting in poor yields. Less acidic DCI ($pK_a = 5.2$) was found more suitable for chemoselective activation of the diisopropylamine moiety providing significant improvements in the reaction of morpholino P-amidites of type **88** and immobilized nucleoside **89**. Next, the intermediates of type (**E**), phosphoramidite diesters (for morpholines) or phosphite esters (for nucleosides), were boranated (boranophosphates are discussed separately in detail in Sect. 7 of this chapter). After boronation, subsequent detritylation and repetition of the cycle generated chimeras up to 21 mers. Finally, oxidative aminations at the respective boranophosphonate linkages via iodine and different amines produced ranges of PMOs and PMO-DNA chimeras of general structure **90** (Scheme 14b).

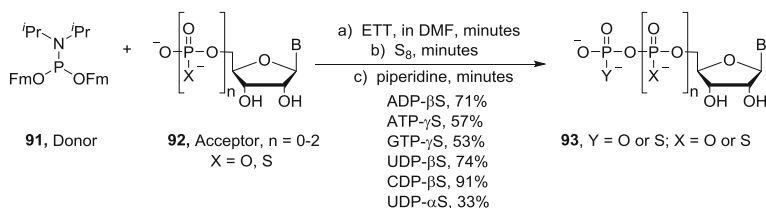
5 Thiophosphates

Sulfur is isoelectronic with oxygen. Therefore, phosphate analogs in which one or more oxygen atoms are replaced with sulfur atoms provide potential phosphate mimics. In thiophosphates, non-bridging oxygen atoms are replaced by sulfur; the analogs display greater hydrolytic stability and are widely used as probes for investigating cellular mechanisms. The syntheses and applications of thiophosphate analogs have been reviewed in 2011 [38].

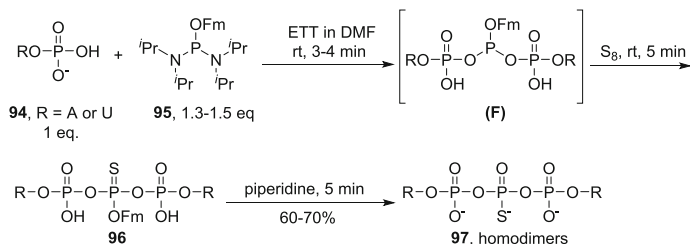
5.1 Thiophosphorylation Giving Racemic Mixtures

A recently reported P-(mono)amidite based iterative coupling strategy allows for syntheses of nucleotide polyphosphate analogs **93** [39, 40]. First, the coupling of donor phosphoramidite **91** with acceptor nucleosides **92** ($n = 0$) or their mono- ($n = 1$) and di- ($n = 2$) phosphates proceeds in the presence of a tetrazole-derived activator (ETT), followed by oxidation with elemental sulfur. After deprotection of the Fm groups, the obtained products can be further phosphorylated or thio-phosphorylated by iteration of the same protocol. Reagent **91** exclusively modifies the oxygen of thio-phosphates and is position specific (terminal), giving positional flexibility for sulfur incorporation (Scheme 15).

A similar strategy used P-diamidite **95** for the synthesis of C_2 -symmetrical dinucleoside triphosphate analogs Ap-p(S)-pA and Up-p(S)-pU **97** (Scheme 16) [41]. The approach demonstrates the general ability of P-diamidites to selectively



Scheme 15 Iterative synthesis of thiophosphate-containing nucleosides

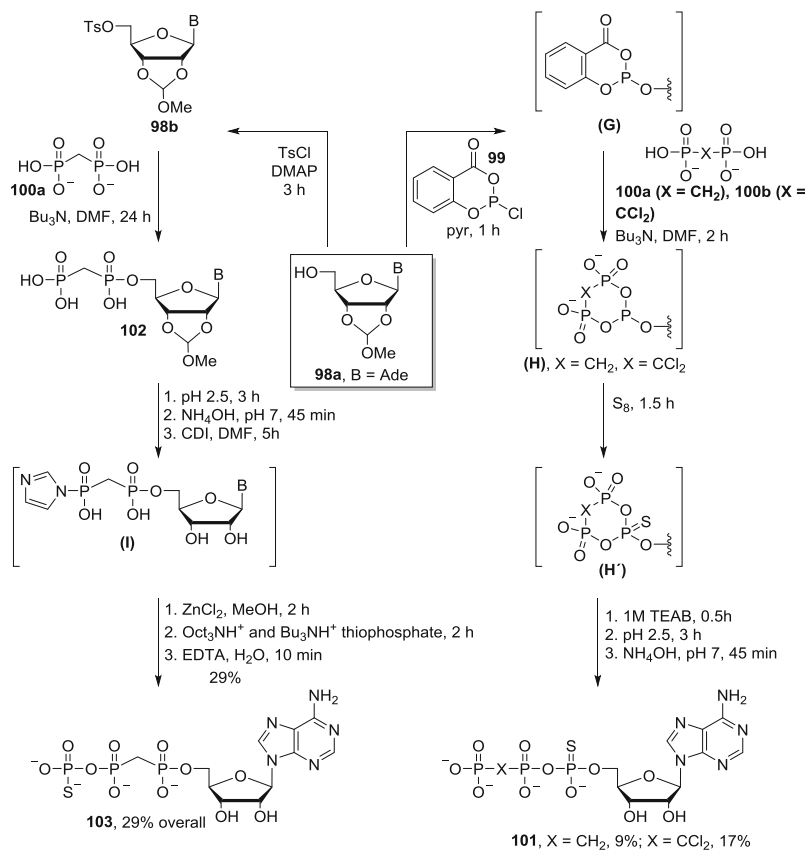


Scheme 16 Synthesis of Npp(S)pN homodimers

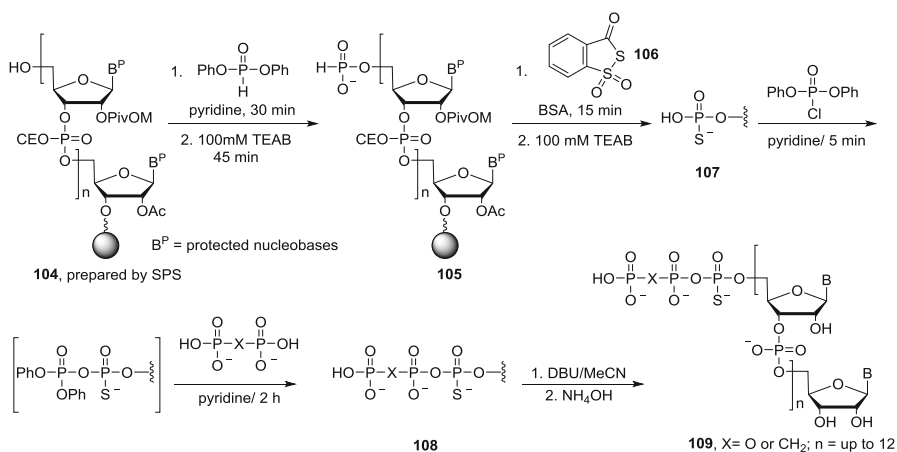
react with phosphates **94** under ambient conditions without protecting groups on either the sugar or nucleobase (B). A protecting group on the P-diamidite is required to stabilize the reagent; Fm is used because of its rapid and quantitative removal by piperidine. The coupling conditions rapidly lead to the formation of mixed P(V)–P(III)–P(V) intermediates (**F**). Subsequent oxidation using elemental sulfur followed by Fm-deprotection of **96** gave analogs **97**, which are isolated by simple precipitation.

Abnormal activity of nucleotide pyrophosphatase/phosphodiesterase-1 (NPP1) correlates with diseases such as chondrocalcinosis, osteoarthritis, and type 2 diabetes. Recently, a combination of thiophosphorylated and nonhydrolyzable ATP analog NPP1 inhibitors **101** and **103** were synthesized ($\text{IC}_{50} = 0.39\text{--}0.57 \mu\text{M}$) [42]. The synthesis of **101** and **103** started from the protected adenosine **98a**, which was activated with **99** to afford intermediate (**G**). (**G**) was then coupled with pyrophosphate analogs **100a–b** to yield the cyclic trimetaphosphate intermediates of type (**H**). Sulfurization followed by addition of TEAB buffer opened the ring to produce the corresponding 2',3'-protected analog in a one-flask reaction. Subsequent treatments with aqueous HCl and NH_4OH yielded a mixture of diastereomers **101**, which was further purified by DEAE-Sephadex or RP18 chromatography. To obtain the γ -thiophosphate analog **103**, compound **98a** was first tosylated affording **98b** and then reacted with methylenebisphosphonate **100a** to give the ADP analog **102**. After removal of the methoxymethylidene-protecting group, the terminal phosphate was activated with carbonyl diimidazole (CDI). The activated intermediate (**I**) reacted with either tributylammonium or tri-octylammonium thiophosphate salts in the presence of Lewis acid ZnCl_2 to afford **103** in 29% yield (Scheme 17).

Thillier et al. [43] recently described solid-supported syntheses of oligoribonucleotide 5'-(α -P-thio)triphosphates and 5'-(α -P-thio)(β,γ -methylene)triphosphates **109** (Scheme 18). First, synthetic solid-supported oligoribonucleotides **104** were transformed into H-phosphonates **105**. Afterwards, silylation with *N,O*-bis(trimethylsilyl)acetamide (BSA) was used to increase the reactivity of H-phosphonate (also see Scheme 36c and 38a) **105**, followed by sulfurization with Beaucage's reagent **106** [44] giving oligoribonucleotide thiophosphate monoesters **107**. These were activated with diphenyl phosphoryl chloride and reacted with an alkylammonium salt of pyrophosphate or its bisphosphonate analog to give solid-supported protected oligonucleotide triphosphate and its analogs **108**, respectively. Finally, all protecting groups were removed under basic conditions, which also released the products from the solid support. The crude ppp(S)-



Scheme 17 Synthesis of pyrophosphatase/phosphodiesterase-1 (NPP1) inhibitors

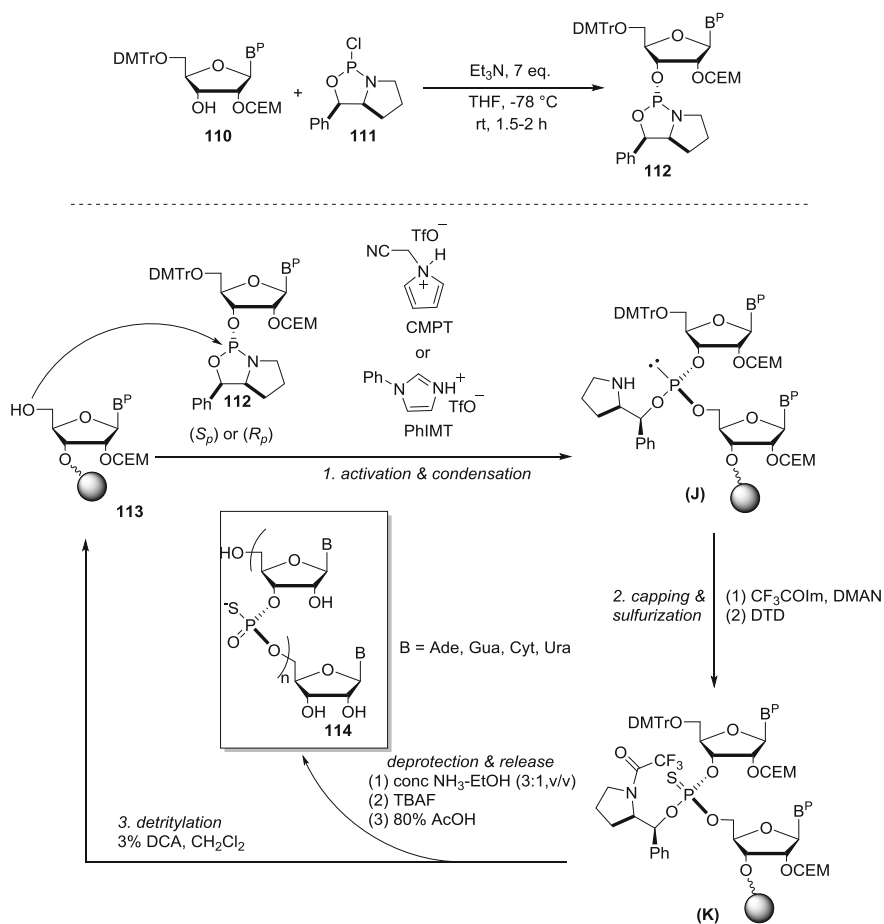


Scheme 18 Solid phase synthesis of 5'-p(S)(O/CH₂)p capped ribonucleotides

ORN analogs were further purified by IEX-HPLC giving pure compounds of type **109**. This serves as a general strategy for syntheses of 5'-(α -P-thio)triphosphate RNA sequences with high purity (>95%) and overall yields around 20%.

5.2 Stereoselective Thiophosphorylation

Generation of a stereogenic P-atom at internucleotide linkages is a long-standing problem in the synthesis of thiophosphate oligonucleotides (PS-ODNs and PS-ORNs). Biomolecules interact in a stereospecific manner, and therefore stereodefined P(S)-oligonucleotides would greatly aid their exploration as sophisticated functional tools. Stec et al. [45] and Wada and colleagues [46] are pioneers for syntheses of stereoregular oligonucleotides; Stec's methods are based on oxathiaphospholanes (OTPs), whereas Wada uses oxazaphospholidines to introduce stereochemical bias. Wada and coworkers successfully demonstrated a solid-



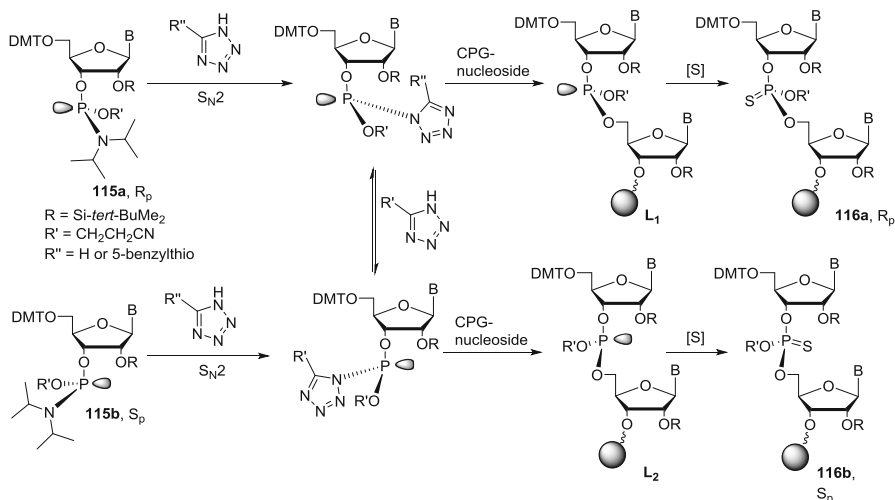
Scheme 19 Solid-supported stereocontrolled synthesis of PS-ORNs via oxazaphospholidine

supported and stereodefined ($\geq 98:2$) synthesis of a mixed PS-ORN sequence (Scheme 19) [47].

Stereoselective trans-oxazaphospholidines of type **112** are synthesized by reactions between free 3'-OHs of otherwise protected nucleosides, e.g., **110** and enantiomerically pure 2-chlorooxazaphospholidine derivatives, D- or L-**111**, which are generated from D- or L-proline precursors. The reactions between **110** and **111** generated the anti-isomers at phosphorus almost exclusively ($>99:1$). 2'-Cyanooxymethyl (CEM) oxazaphospholidines **112** are not stable on silica gel but were chosen because of their higher reactivity; 3-aminopropyl functionalized silica gel was employed to minimize their decomposition during purification.

Stereoregular thiophosphate oligoribonucleotides (PS-ORNs) were obtained in four steps: (1) condensation, (2) capping and sulfurization, (3) detritylation, and (4) deprotection and release (Scheme 19). First, solid-supported 5'-OH nucleosides or PS-ORNs **113** were condensed with one of the 2'-O-CEM-protected 3'-O-oxazaphospholidine monomers **112** in the presence of either CMPT or PhIMT activators. Capping of unreacted 5'-OHs and the free secondary amino group of the intermediate (**J**) were then carried out using trifluoroacetylimidazole (CF_3COIm) and 1,8-bis-(dimethylamino)naphthalene (DMAN). Subsequent oxidative sulfurization in the presence of DTD produced (**K**) with retention of configuration. Detritylation with 3% dichloroacetic acid (DCA) in dichloromethane liberates the 5'-OH group, and the cycle can be repeated (up to 11 times). Oligomers of different chain lengths **114** were deprotected and cleaved from solid support using standard procedures at the end of the cycle.

Another notable SPS method introduces tetrazole-induced stereochemical bias into thiophosphate-containing small interfering RNAs (PS-siRNAs) [48]. Tetrazoles

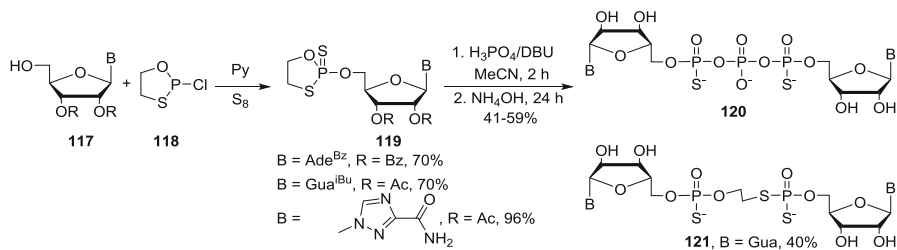


Scheme 20 Solid-phase diastereocontrolled synthesis of P(S)-ORN via phosphoramidite coupling on controlled-pore glass (CPG)

are typically used as activators in solid phase ORN synthesis to enhance the nucleophilic attack rate at the P(III) stereogenic center of **115** by the 5'-hydroxyl of the ribonucleoside (Scheme 20). The stereochemical outcome of this reaction is influenced by activator choice and not by initial R_p/S_p ratios of the phosphoramidites **115a** and **115b**. 5-Enzylthio-1-*H*-tetrazole (BTT) and tetrazole (tet) as activators influence the product diastereomeric ratios L_1 and L_2 (R_p vs. S_p) in a complementary way; BTT preferably produces R_p isomers, whereas tetrazole favors formation of the S_p isomer. The product R_p/S_p ratios also appear to be determined by the substituents at the 2'-position of the sugar moiety. The final oxidative sulfurization step proceeds with retention of configuration to yield either diastereomerically enriched **116a** or **116b** with lengths of more than 20 units.

Activator-based diastereoselectivities attained from a single PS coupling are small; however, small biases are accumulated in ORNs comprised of several PS linkages. Such stereo-augmented ORNs are then capable of influencing bulk biophysical properties. Although P(O) siRNAs exhibit the highest inhibitor potency against target proteins Lin28 and p53 in Huh7 and HeLa cells, P(S)_{BTT} siRNAs (higher R_p content) display superior stability. Higher R_p content siRNAs also possess higher potency over their P(S)_{tet} siRNA (higher S_p content) counterparts. The hydrolytic stability of R_p enriched siRNAs together with their improved pharmacokinetic properties demonstrates their potential future applications in siRNA therapeutics.

Fragile histidine triad (Fhit) protein-dinucleoside polyphosphate (NP_nN) complex (Fhit-NP_nN complex) is postulated to function as a signaling complex in tumor suppression mechanisms, which upon NP_nN hydrolysis inhibits signal transduction [49]. In the search for hydrolytically stable dinucleoside polyphosphates as Fhit substrates, three Np(S)pp(S)N analogs of type **120** were synthesized (Scheme 21) [50]. Oxathiaphospholane (OTP) intermediates **119** were prepared by coupling protected nucleosides **117** and 2-chloro-1,3,2-oxathiaphospholane **118** followed by oxidative sulfurization in the presence of elemental sulfur. The diastereomers were not separated at this stage, and further reaction was carried out between the diastereomeric mixtures and phosphoric acid in the presence of DBU. This procedure yielded cap analogs **120** as mixtures of corresponding diastereomers. These were separated by RP-HPLC, and seven out of nine possible diastereomers were isolated. The double condensation of **118** and the nucleoside **117** (B = Gua) also produced a by-product **121** (4 diastereomers). Type **120** analogs displayed a



Scheme 21 Synthesis of dinucleotide polyphosphate analogs Np(S)pp(S)N

two to three orders of magnitude lower hydrolysis rate catalyzed by Fhit as compared to their natural triphosphate counterparts.

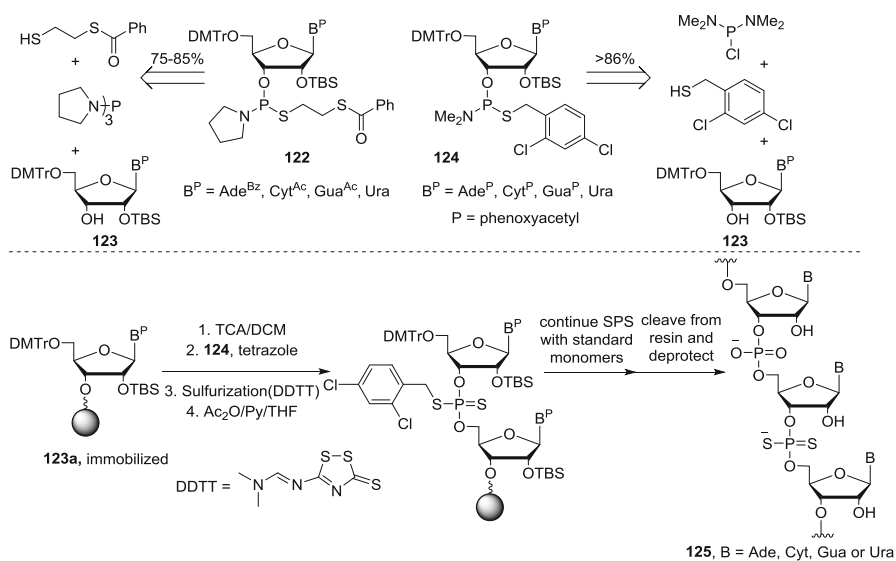
5.3 Dithiophosphates

Dithiophosphates ($P(S)_2$) are isosteres of phosphate where both non-bridging oxygens are replaced by sulfur. Unlike $P(S)$ analogs, $P(S)_2$ analogs are symmetrical like natural phosphates. Syntheses and applications of nucleotide and oligonucleotide dithiophosphates were reviewed in 2010 [51].

Despite their usefulness, only few reports on $P(S)_2$ -ORNs are documented. Recently, Yang et al. and Piccirilli and colleagues independently reported two different ribonucleoside 3'-phosphorothioamidite monomers **122** and **124**, respectively (Scheme 22). These monomers facilitate the incorporation of at least one $P(S)_2$ linkage and have been used to introduce several modifications using a DNA/RNA synthesizer [52, 53].

Phosphorothioamidites **124** were prepared by treating protected nucleosides **123** with bis(dimethylamino)phosphorochloridite followed by addition of 2,4-dichlorobenzylmercaptan in a one-flask reaction with yields of 86–99%. Phosphorothioamidites **122** were prepared in a similar fashion starting from the same protected nucleosides **123**. First, the phosphorylation of **123** was carried out in the presence of tris-(pyrrolidino)phosphine and 1*H*-tetrazole. Subsequent in situ addition of monobenzoylthanedithiol, trimethylsilylimidazole and additional 1*H*-tetrazole yielded **122** (75–85%).

Modified $P(S)_2$ linkages were incorporated in an automated synthesizer by tetrazole-activated coupling between solid-supported oligonucleotides **123a** and one



Scheme 22 Solid-supported automated synthesis of dithiophosphate-containing ribonucleotides

of the phosphorothioamidites **124** followed by sulfurization using 3-(*N,N*-dimethylaminomethylidene)amino-3-*H*-1,2,4-dithiazole-5-thione (DDTT) prior to the capping procedure. After capping and detritylation, the protected support-bound modified oligonucleotides were deprotected and isolated in three steps: (1) treatment with thiophenol/trimethylamine, (2) heating in ammonium hydroxide/ethanol mixture, and (3) heating with triethylamine (TEA) and TEA-3HF. This sequence gave the desired modified P(S)₂-linked oligoribonucleotides **125** (Scheme 22). However, attempts to obtain PS₂-oligoribonucleotides containing exclusively this modification have remained unsuccessful so far.

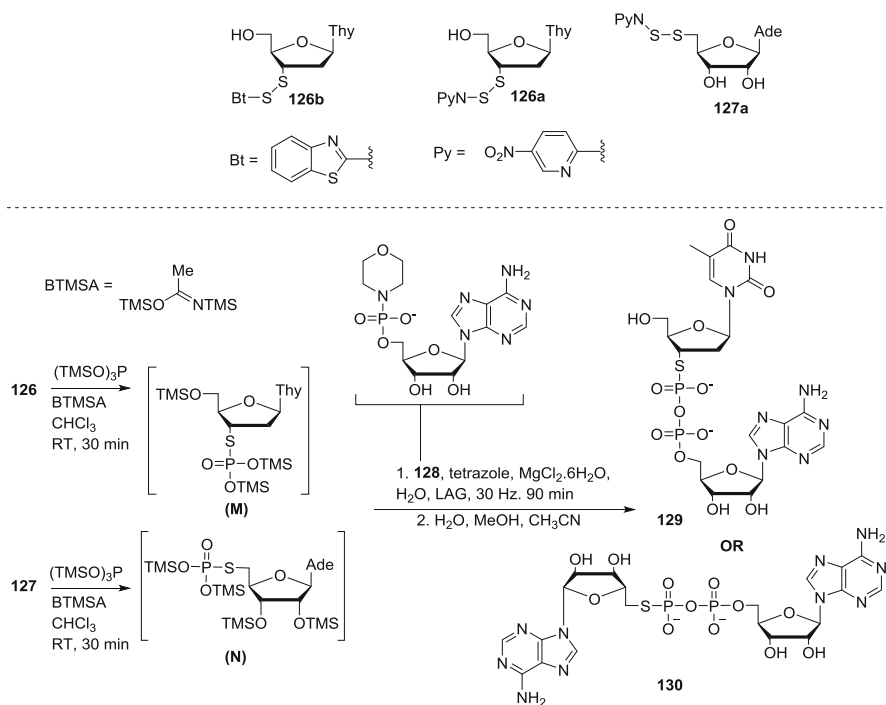
Gene silencing ability of the P(S)₂-bearing duplexes are sometimes higher compared to P(S) modified or unmodified duplexes. Silencing is also position dependent with regard to the modification. Consistently higher serum stability of P(S)₂-siRNA duplexes is observed as compared to P(S)-siRNA in 10% fetal bovine serum.

5.4 Phosphorothiolates

Phosphorothiolates contain an O-P-S substitution at the bridging oxygen. Nucleotides bearing 3'-S phosphorothiolates are useful probes for detection of metal ion-dependent phosphotransesterifications or determination of the rate-limiting steps of enzymatic reactions. The rate of hydrolysis is increased in phosphorothiolates, which often renders them difficult to synthesize and isolate [54]. The internucleosidic pyrophosphorothiolates were unknown until recently. They were prepared in a ball mill where a liquid-assisted grinder (LAG) provides mechanical energy for coupling between 5'-NMP-morpholidates **128** and 3'-NMP-phosphorothiolates **126** or 5'-NMP-phosphorothiolates **127** to form pyro-phosphorothiolate-linked dinucleoside diphosphate analogs **129** and **130**. To promote the coupling, tetrazole and MgCl₂ activators were also used. The starting materials **126** and **127** were synthesized from their precursor 3' or 5' thio-thymidines **126a-b** or **127a**, respectively (Scheme 23) [55]. This novel strategy used stoichiometric amounts of water; hence, rigorous cation exchange and pre-drying were not necessary, which also preserved the labile thiolate linkage.

5.5 Methylenebisphosphonodithiolates

Methylenebisphosphonodithiolates are multi-thiolated analogs of bisphosphonates. Fischer and coworkers recently reported novel methylenebisphosphonodithiolate nucleoside analogs **136** and **137**. The non-hydrolyzable analogs **136** and **137** show metabolic stability toward nucleotide pyrophosphatase/phosphodiesterase (NPP1,3) and nucleoside triphosphate diphosphohydrolase (NTPDase1,2,3,8). The compounds show high chemical stability against acidic (for **136**, pD 1.5, $t_{1/2}$ = 44 h), basic (no decomposition 2 weeks at pD 11), and oxidative degradation (for **137**, $t_{1/2}$ = 3 days) [56]. While **136** and **137** are not efficient inhibitors of NPP1,3, they displayed strong Zn²⁺ chelation and could potentially be useful in metal complexation applications.



Scheme 23 Synthesis of pyrophosphorothiolate-linked dinucleotides

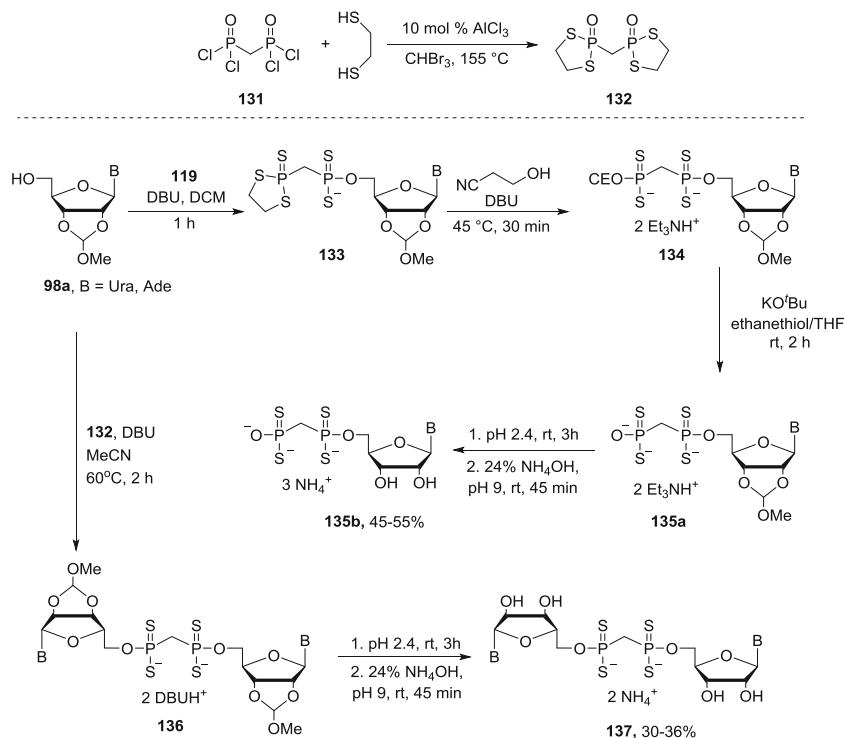
Dithiaphospholane **132** is the key reagent for introduction of the $\text{P}(\text{S})_2\text{-CH}_2\text{-P}(\text{S})_2$ motif into nucleosides. **132** can be prepared from methylene bis(phosphonic dichloride) **131** and 1,2-ethanedithiol in the presence of catalytic amounts of AlCl_3 (Scheme 24) [57].

The free 5'-OH of the 2',3'-methoxymethylidene protected nucleosides **98a** reacts with **132** in the presence of DBU to form intermediates **133**. Addition of excess 3-hydroxypropionitrile and one equivalent of DBU into the reaction mixture converts **133** into **134**. Base promoted β -elimination from **134** followed by the removal of methoxymethylidene protecting group gives $\text{NP}(\text{S})_2\text{-CH}_2\text{-P}(\text{S})_2$ analogs **135a–b** (Scheme 24). The same strategy was used for syntheses of dinucleotide $\text{P}(\text{S})_2\text{-CH}_2\text{-P}(\text{S})_2$ analogs **137**, via intermediate **136**, in 30 and 36% yield, respectively (Scheme 24).

5.6 Thiophosphoramidates

Thiophosphoramidates are a class of phosphate mimics where a nitrogen and a sulfur replace two oxygens. Thiophosphoramidates could potentially serve as therapeutic agents [58] and telomerase inhibitors [59, 60].

Nucleoside-based uncharged N- and S-bridging thiophosphoramidate analogs are phosphate triester mimics, can be used as prodrugs, and can display antiviral activity. They may also be used as a general tool for mechanistic studies of natural



Scheme 24 Synthesis of methylenebisphosphonodithioate analogs

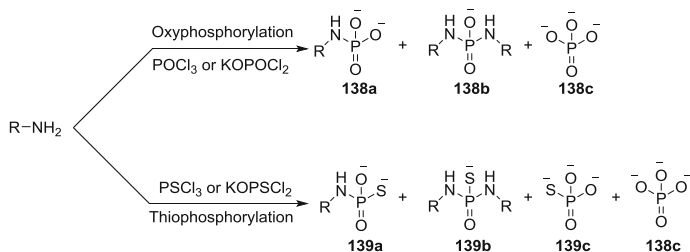
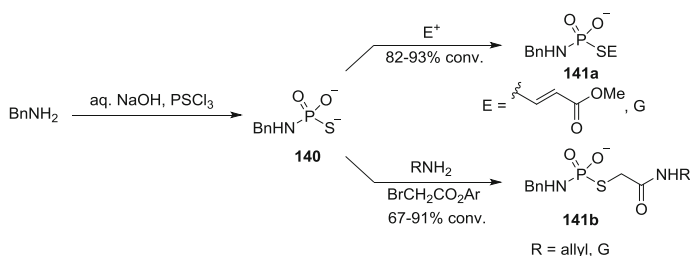


Fig. 4 Distribution of products during phosphorylation and thiophosphorylation of amines

phosphodiesterases, such as sugar nucleotide diphosphatases [61]. Hodgson and coworkers developed a new ‘click’ methodology [62] for the synthesis of N- and S-dialkylated thiophosphoramidates, which has been further extended for the synthesis of novel N- and S-bridging nucleoside analogs. The preparation of these analogs now facilitates their exploration as potential mechanistic probes.

Many synthetic protocols in phosphorylation chemistry require strictly dry conditions and are restricted to specialized laboratories [63]. Hodgson and coworkers were able to transform generic alkyl amines into N,S-dialkyl thiophosphoramidates in one flask [62, 64] in an aqueous medium with THF as a co-solvent



Scheme 25 Heterobifunctional thiophosphoramidate cross-linkers from P(S)Cl₃

to dissolve the reagent P(S)Cl₃. Formation of undesired by-products (Fig. 4) during thiophosphorylation and phosphorylation reactions, such as inorganic phosphates **138c**, thiophosphates **139c**, and aminolysis products **139b**, was greatly suppressed if the reactions were carried out at pH 12 [65].

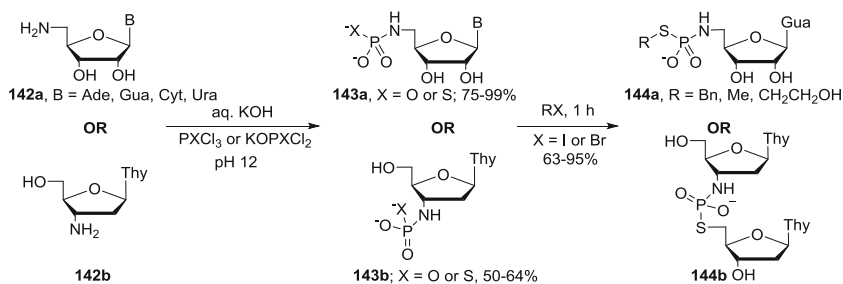
After completion of the reaction under optimized conditions, the organic solvents were first removed followed by lyophilization. The thiophosphoramidate reaction mixtures **140** were then alkylated at sulfur and the products were purified by anion exchange DEAE-Sephadex resin. The scope of the approach to rapidly access N,S-dialkyl thiophosphoramidates has been explored using several alkyl amines and alkylating agents to produce compounds of type **141a** (some representative examples are shown) [62]. The methodology was also used to accomplish amine–amine ligation via heterobifunctional cross-linking between thiophosphoramidate anionic intermediate **140** and aryl bromoacetate electrophiles [64]. The subsequent aminolysis of the active acylating agents in the presence of allylamine or 5′-amino-5′-deoxyguanosine produces thio-phosphoramidate-acetamide compounds of type **141b** (Scheme 25).

Unprotected aminonucleosides **142a–b** were also phosphorothioated using a similar pH-controlled protocol to yield products **143a–b**; **143a** was then further elaborated into **144a** using S-alkylation (Scheme 26) [65]. Notably, dinucleoside thiophosphoramidate **144b** can also be prepared [66]. All crude S-alkylated products were purified over anion exchange DEAE-Sephadex resin [65, 66].

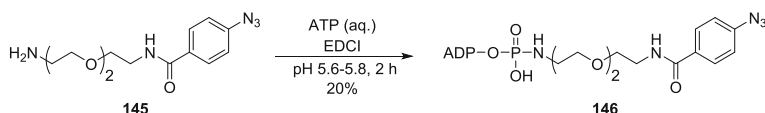
6 Tagged Phosphates

Terminal phosphate-tagged nucleotides provide useful tools for investigating essential cellular activities, such as DNA synthesis [67, 68] and protein phosphorylation [69]. Phosphorylation occurs in numerous occasions during signal-transduction cascades; aberrations in these pathways can lead to diseases such as Parkinson's [70] and cancer [71].

Adenosine 5′-triphosphate (ATP) is the central phosphate donor in biology; its γ-phosphate is most commonly transferred to serine, threonine, or tyrosine residues of proteins (other amino acids are also phosphorylated). ATP analogs modified at the γ position are often accepted as kinase co-substrates, a feature exploited in chemical



Scheme 26 Further scope of thiophosphoramidate synthesis



Scheme 27 Synthesis of γ -modified ATP-photo cross-linker **146**

biology. However, the high abundance of kinases in eukaryotes [72] makes it difficult to identify specific kinases responsible for phosphorylations [73, 74].

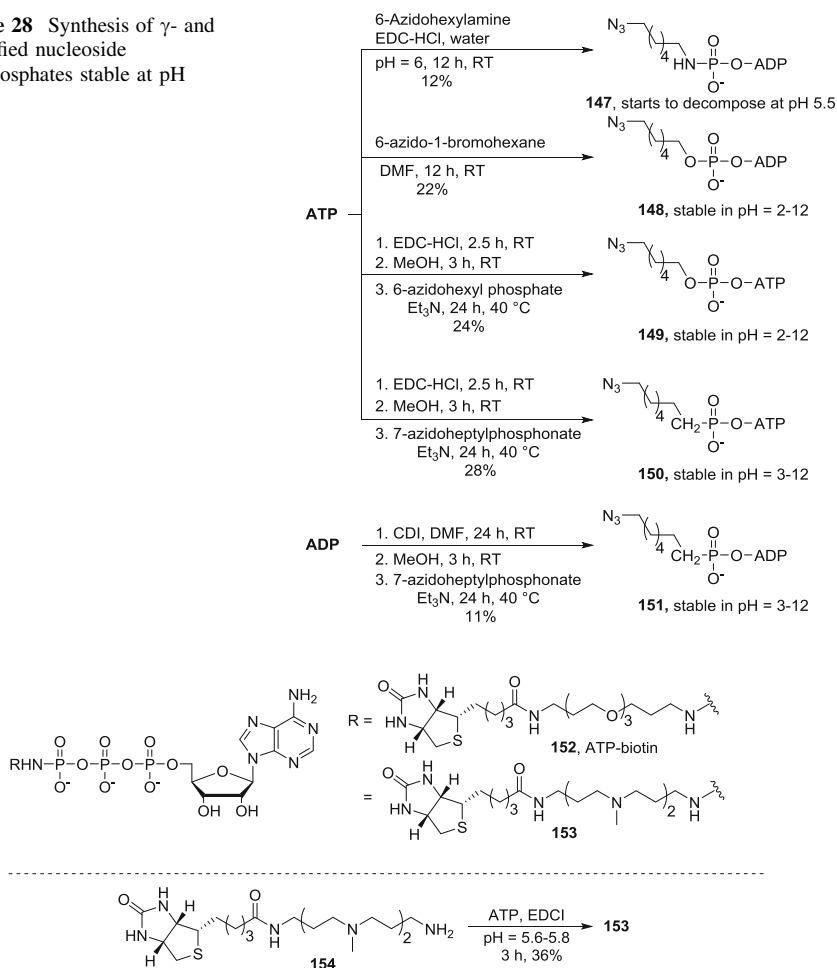
A photo cross-linking unit such as an aryl azide, when attached to a γ -phosphate modified ATP analog, can help to identify the operating kinases or targets of interest in vitro [75]. For example, in the presence of co-substrate, UV irradiation during CK2-mediated phosphoryl transfer to α -casein forms a phosphorylated protein bearing a highly reactive nitrene species, which covalently binds to the kinase, thus enabling its identification.

The ATP γ -phosphoramidate photo cross-linker **146** was synthesized by condensing γ -ATP and *N*-(4-azido benzoyl)-2,2'-ethylenedioxy bis-(ethylamine) **145** using an established reaction protocol in the presence of EDCI as an activator (Scheme 27).

Marx and coworkers also reported synthetic routes for γ - and δ -phosphate modified nucleotides **147–151** [76]. *N*- and *O*- γ -modified nucleotide analogs have been described previously [77–83]. However, the group also developed novel γ - and δ -CH₂ tagged nucleotides **150** and **151** (Scheme 28). Disodium ATP undergoes a *N*-6-azidoheptyl modification at the terminal phosphate at slightly acidic pH in the presence of 6-azido-hexylamine and EDC as an activator (12% yield) affording **147**. *O*-6-Azidoheptyl tags **148** and **149** were attached either by alkylation of tetrabutylammonium ATP using 6-bromo-1-azido-hexane in DMF under anhydrous conditions (γ -modified, **148**: 22% yield) or via reaction between tetrabutylammonium ATP and 6-azido-hexylphosphate mediated by EDC (δ -modified, **149**: 24% yield). Finally, novel γ - and δ -CH₂ tagged nucleotides **150** and **151** were synthesized from tetrabutylammonium ATP and tetrabutylammonium ADP, respectively, through coupling with 7-azido-heptyl-phosphonate. Different activators enabled the synthesis of tri- (CDI, yield 11%) and tetra- (EDC, yield 28%) phosphate analogs.

Azide-tagged nucleotide analogs can be further subjected to azide-alkyne click reactions or Staudinger ligations; reduction into corresponding amines opens up the

Scheme 28 Synthesis of γ - and δ -modified nucleoside oligophosphates stable at pH 5.5–12



Scheme 29 Synthesis of cell-permeable ATP analog, ATP–polyamine–biotin **153**

possibility to further explore amide chemistry on these substrates as well. When subjected to a range of different pH, O- and CH₂ analogs **148–151** show remarkable and almost identical hydrolytic stabilities, whereas the P-amidate analog **147** is less stable, especially at acidic pH (Scheme 28).

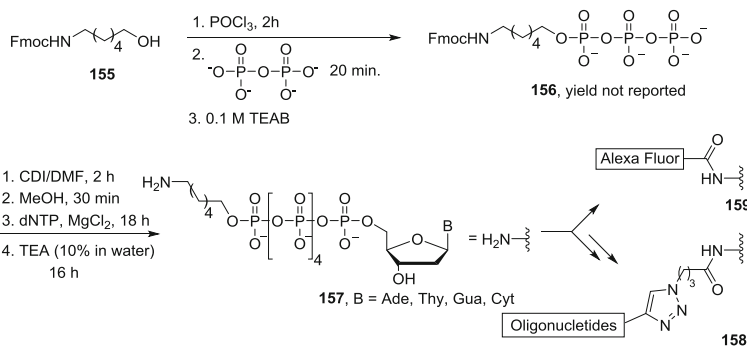
ATP-biotin **152** is another recently reported γ -phosphate-labeled ATP analog [84]. During a kinase-catalyzed phosphorylation reaction, **152** acts as a co-substrate, and the phosphorylbiotin unit is transferred to target proteins [85]. These biotinylated phosphoproteins can then be pulled down by several commercially available streptavidin-conjugated reagents [86]. This approach can only be applied in vitro because of the cell impermeability of **152** [85]. Recently, Fouda and Pflum synthesized a cell-permeable ATP–polyamine–biotin **153** that can be used in cellulose [87]. Synthesis of **153** was accomplished as described in Scheme 29; polyamine–biotin **154** was condensed with γ -ATP phosphate following the above-mentioned

standard protocols (Fig. 1) to give **153**. *In vitro*, **153** was less efficient than ATP and ATP-biotin **152**. However, in live cells, it successfully transfers its γ -phosphate and generates biotinylated phosphoproteins. **153** is a promising candidate to aid in the study of the cellular phosphoproteome, and therefore this strategy may provide additional insights into cell-signaling research.

Oligonucleotide-tagged or fluorescently labeled nucleoside hexaphosphate analogs **158** or **159** are used in real-time single-molecule DNA sequencing [88, 89]. All tagged and labeled analogs of family **158** and **159** were synthesized from aminohexyl-nucleoside hexaphosphate (dNP₆) intermediates **157**. These amines **157** are then modified with activated esters that can contain azides or also other labels. The azides further undergo a click reaction to form compounds **158** or nucleobase-specific *Alexa fluor*-labeled activated esters to form members of the family **159**. The intermediates **157** were synthesized starting from Fmoc-6-aminohexanol **155**, which was converted into Fmoc-aminohexyl-triphosphate **156** via sequential treatments with phosphorous oxychloride followed by tributylammonium pyrophosphate (Scheme 30). Compound **156** was isolated and purified by Sephadex A-25 ion-exchange chromatography prior to its use in subsequent steps. Pure **156** was then condensed with nucleoside triphosphates (as tributylammonium salts) in the presence of CDI as an activator and a Lewis acidic promoter (MgCl₂). Deprotection delivers crude aminohexyl-nucleoside hexaphosphates **157**.

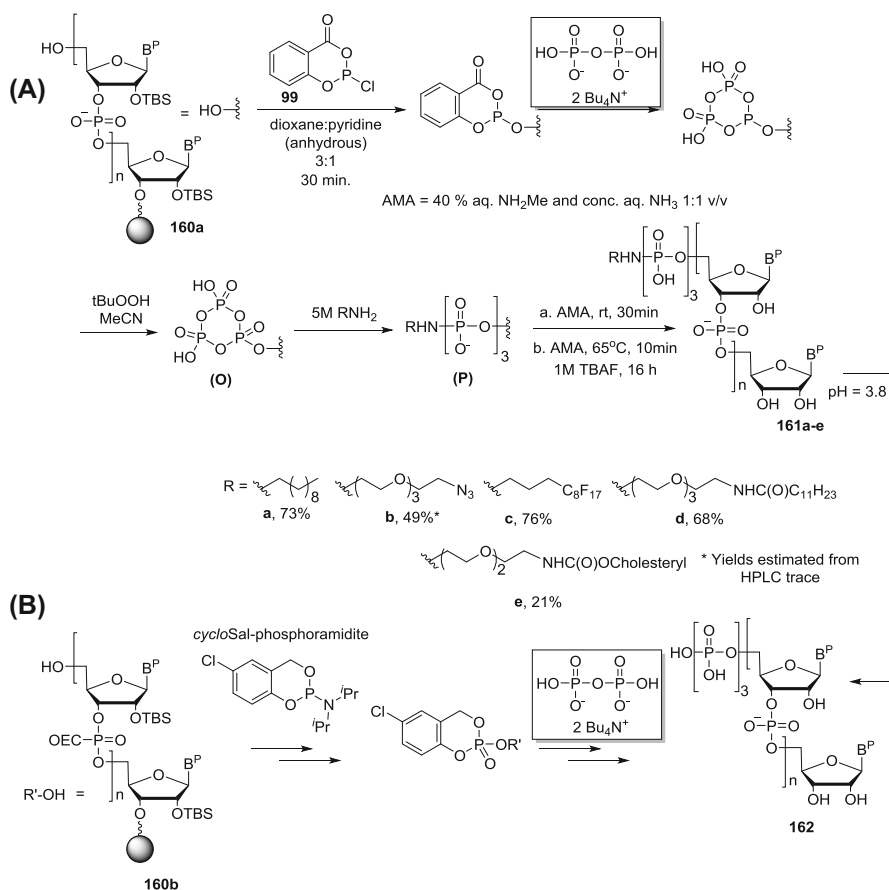
RIG-I is a class of pattern recognition receptors; RIG-Is are located in the cytoplasm and recognize 5'-triphosphate end-capped viral RNAs. Recognition triggers type 1 interferon production, which helps to maintain the antiviral state of the cell [90, 91]. A great deal of interest is focused on creating synthetic RIG-I agonists to aid understanding of signal responses [92–94]. Future studies will rely heavily on the availability of pure 5'-triphosphate RNAs (pppRNA). Although several synthetic methods have been reported so far [95, 96], almost all of them require rigorous purification protocols [97, 98].

Ludwig and coworkers prepared a solid-supported cyclic trimetaphosphate RNA intermediate (**O**) from its solid-supported precursor **160a** using salicyl phosphorochloridite following a standard procedure. The cyclic trimetaphosphate ring was opened with a series of amines to produce an immobilized γ -N-modified 5'-



Scheme 30 Synthesis of labeled nucleoside hexaphosphates

triphosphate analog of 24-mer RNA (**P**), which was then cleaved from solid support to give oligoribonucleotide P-amidate (for phosphoramidates see Sect. 4 of this chapter) **161a–e**. The terminal alkyl chain enhances lipophilicity and simplifies the purification of **161a–e** significantly (Scheme 31) [99]. The cyclic trimetaphosphate RNA intermediate (**O**) provides a common intermediate (see schemes 17, 35, and 43) for the synthesis of various P-amidate γ -tagged 5'-triphosphate RNAs. Modified *N*-azidoalkyl, perfluoroalkyl, monoacetylated amine, 2,2'-(ethylenedioxy) ethylamine tagged triphosphate RNAs were synthesized. The retention times for **161a–e** are independent of sequence and are solely determined by tag properties, which add lipophilicity to simplify purification. Perfluoroalkyl derivative **161c** provides the option to conduct fluorophilic affinity chromatographic separation. All analogs **161a–e** can be deprotected to give corresponding free 5'-triphosphate RNAs **162** through acid-catalyzed (pH 3.8) P–N bond scission at 60 °C. This is in accordance with the above study by Marx and coworkers on the stability of such linkages (Scheme 31a). An alternative solid-supported synthesis of triphosphate RNA **162** proceeded from



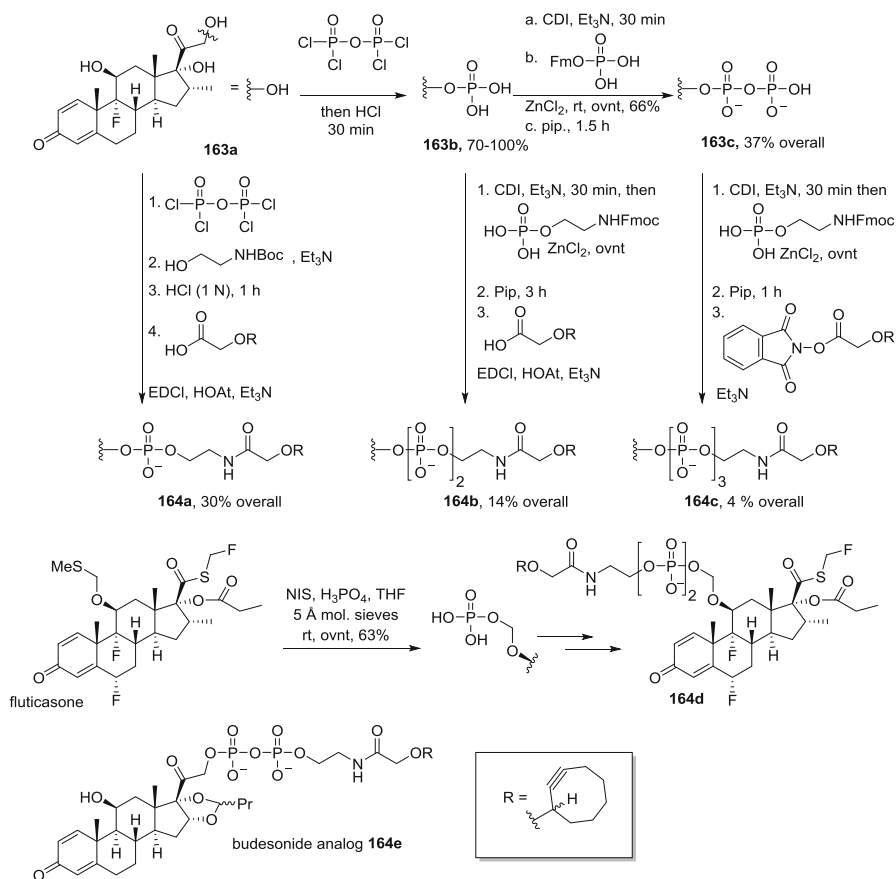
160b via *cycloSal*-phosphoramidite developed by Meier; the intermediate was further oxidized and condensed with pyrophosphate salt (Scheme 31b) [100]. In principle, using different pyrophosphate analogs, one can envision the facile generation of RNA triphosphate analogs. Notably, the *cycloSal* approach does not proceed via a cyclic trimetaphosphate intermediate as the Ludwig approach. This approach has recently been discussed in a detailed review demonstrating several other applications as well [101].

Glucocorticoids (GC) are an important class of steroid hormones, which bind cytosol-localized glucocorticoid receptors (GR). The GC-GR complex either upregulates the expression of anti-inflammatory proteins in the nucleus or downregulates the expression of pro-inflammatory proteins in the cytosol. The pleiotropic pharmacology of glucocorticoids is associated with many side effects [102] such as an increase in sugar level and mood swings.

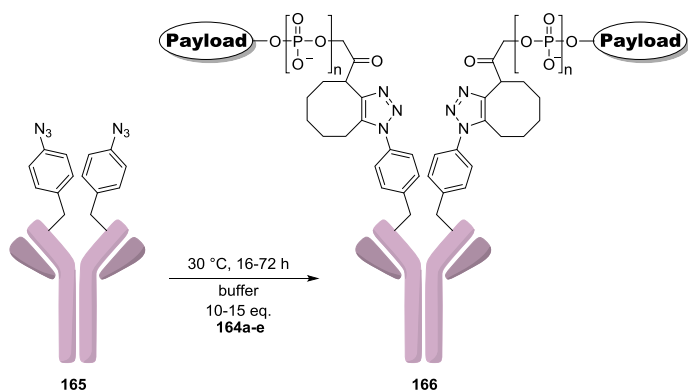
In a recent article, Kern et al. successfully reported several syntheses of glucocorticoid tagged with cyclooctyne-mono-, di- and tri-phosphates **164a–e** following standard procedures as shown in Scheme 32 [103]. The presence of the phosphate unit renders lipophilic steroids partially hydrophilic, suppressing their aggregation in water. The sterically least hindered –OH groups of the glucocorticoids (e.g., C-21 for **163a**) were mono- or di-phosphorylated. Both mono- and di-phosphates were further condensed with either N-Boc protected ethanolamine or N-Fmoc protected aminoethyl-phosphates in the presence of imidazole and/or Lewis acid activators to produce di- and tri-phosphate analogs (not shown). Subsequent Boc and Fmoc deprotection and coupling of the free amines with activated cyclooctynes using EDC gave the desired dexamethasone-derived cyclooctyne-tagged mono-, di-, and tri-phosphate analogs **164a–c**. Employing a similar strategy, Kern et al. also readily synthesized fluticasone- and budesonide-derived cyclooctyne-tagged pyrophosphate analogs **164d** and **164e**, respectively (Scheme 32) [103]. These small molecules were stable in mouse blood up to 6 h but released 50% of their glucocorticoid payloads in lysosome lysates within 30 min (di- and triphosphates); the payload release was slower for the monophosphate **164a** (about 10% in 6 h).

An unnatural amino acid (*para*-azido-phenylalanine) modified α -human CD70 (α -hCD70) antibody (represented as **165** in Scheme 33) allows for bioorthogonal reaction with cyclooctyne-tagged-glucocorticoid-carrier phosphate analogs **164a–e** to form antibody–drug conjugates of general type **166**. Likewise, the environment-specific release of glucocorticoids (blood vs. lysosomal lysate) from such antibody–drug conjugates could provide a solution for glucocorticoid delivery to target-antigen positive cells (Scheme 33) [103].

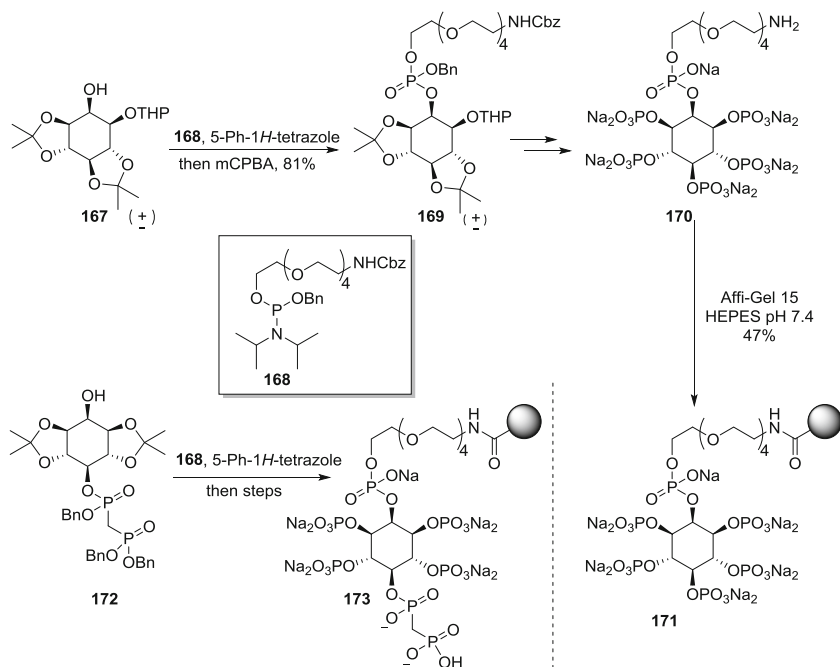
The group further established in vitro proof for efficient glucocorticoid delivery of type **166** in 786-O cells [103], a cell line with high CD70 (expressed on activated lymphocytes) on the cell surface. When 786-O cells were incubated with **166**, the glucocorticoid-induced leucine zipper (GILZ) activity was found to correlate well with previous lysosomal lysate assays. GILZ is a gene that expresses rapidly in the presence of glucocorticoids (GCs). Furthermore, fluticasone ($EC_{50} = 0.25$ nM) and budesonide ($EC_{50} = 0.58$ nM) antibody conjugates were found more potent compared to dexamethasone ($EC_{50} = 0.58$ nM).



Scheme 32 Synthesis of cyclooctyne-tagged mono-, di-, and tri-phosphate analogs of glucocorticoids



Scheme 33 Synthesis of antibody–drug conjugates



Scheme 34 Synthesis of modified inositol phosphates as affinity reagents for pull-down

As previously described in the non-hydrolyzable analog section, inositol pyrophosphates signal in two distinct ways, but many of their protein targets remain elusive. Fiedler and coworkers synthesized InsP₆ and 5PCP-InsP₅ affinity reagents **171** and **173** (Scheme 34) for pull-down assays. They identified more than 150 proteins that bind to these reagents and are responsible for nucleotide as well as glucose metabolism, ribosome biogenesis, phosphorylation-based signal transduction pathways and substrates for protein pyrophosphorylation in *Saccharomyces cerevisiae* cell lysate [104]. 5PCP-InsP₅ is a nonhydrolyzable structural mimic of 5PP-InsP₅ [105] to preclude decomposition during pull-down in lysates.

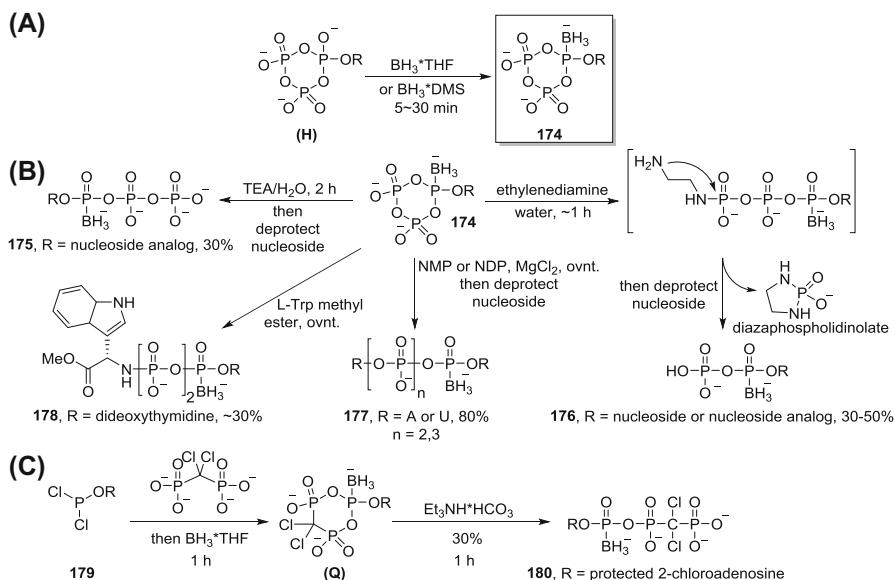
The synthesis of the InsP₆-affinity reagent **171** was accomplished starting from racemic alcohol **167**. The alcohol was phosphitylated in the presence of 5-Ph-1*H*-tetrazole and phosphoramidite **168** followed by oxidation to yield **169**. Subsequent deprotection followed by a global phosphorylation-oxidation sequence and global hydrogenolysis gave **170**. The polyethylene glycol (PEG)-linked primary amine attached at the 2-position of **170** enabled further coupling with N-hydroxysuccinimide (NHS)-activated Affi-Gel 15 beads to give InsP₆-affinity tag **171** in 47% yield (Scheme 31). Starting from **172**, PEG-linked 5PCP-InsP₅-affinity reagent **173** was obtained by applying the same synthetic sequence.

7 Boranophosphates

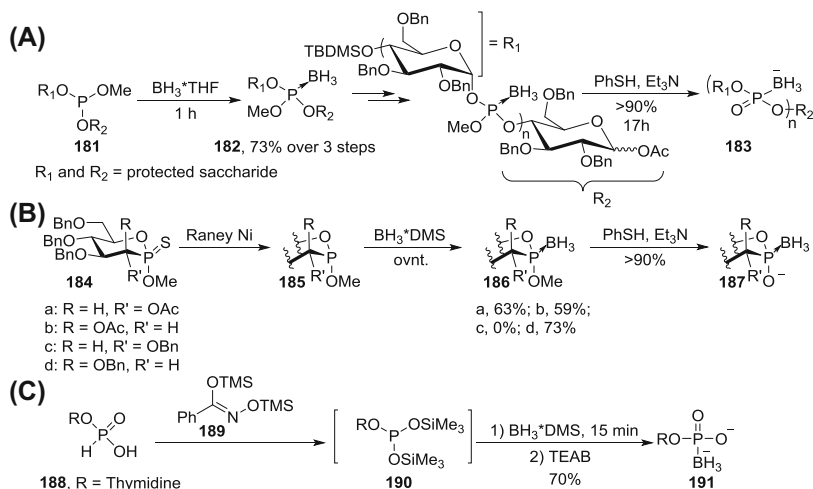
Boranophosphate analogs are isoelectronic to phosphates and bear similar charge distribution, but also possess higher lipophilicity and in vivo stability [106]. Boron is typically introduced in the borane form using three common synthetic strategies: (1) boronation of a P(III) center, (2) using P-BH₃ species as nucleophiles, and (3) using P-BH₃ species as electrophiles. O-to-BH₃ substitution in non-symmetrical phosphodiester creates a stereocenter, resulting in diastereomeric mixtures. For simplicity, all subsequent yields are reported as ~1:1 mixtures of diastereomers unless otherwise noted; the diastereomers can typically be separated by HPLC. Note that this section only covers borane-containing *phosphate* mimics. For a review of a broader class of P-BH₃ compounds, see [107].

7.1 Boronation of a P(III) Center

Cyclic mixed P(III)–P(V) species of type **(H)** [for synthesis of type **(H)**, see Scheme 17] often act as starting materials for borano-containing polyphosphates (Scheme 35a). Type **(H)** compounds are treated with a borane complex such as BH₃*THF to afford the cyclic boranophosphate intermediates **174**, which act as a precursor to a wide range of compounds (Scheme 35b). Treatment of **174** with a solution of triethylamine and water followed by deprotection of the nucleoside yields α-P-borano-NTP analogs **175** in ~30% yield (Scheme 35b) [108]. Use of ethylenediamine in place of trimethylamine promotes cleavage of the terminal phosphate to afford α-P-borano-NDP-**176** in 30–50% yield [109, 110]. The cleavage proceeds via intramolecular cyclization, and elimination of diazaphospholidinolate.



Scheme 35 Boronation of cyclic P(III) centers



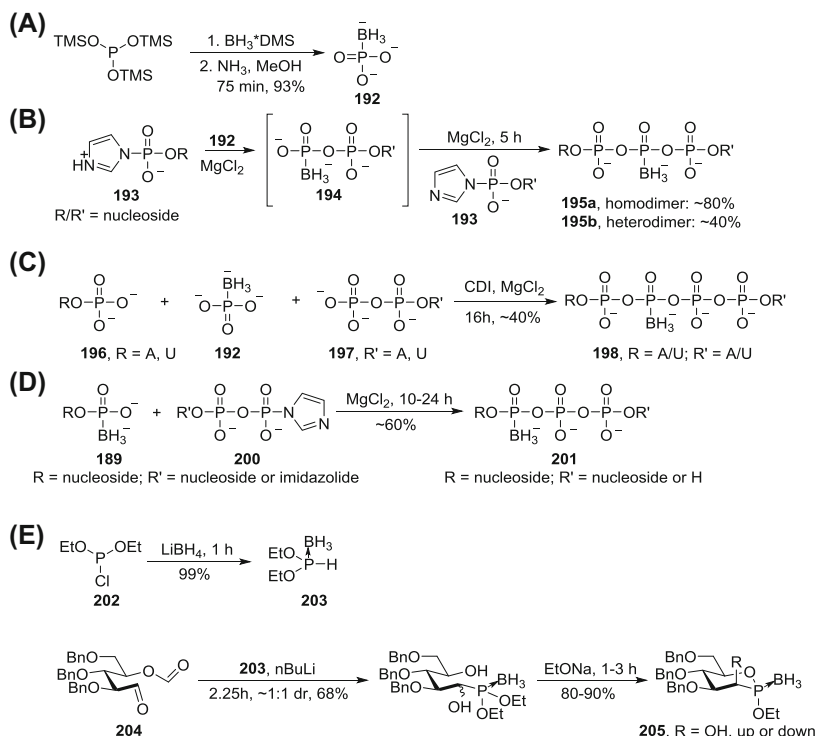
Scheme 36 Boronation of open-chain P(III) centers

174 may also be opened with NMPs or NDPs to yield P-borano homodimers **177** in yields of up to 80% [111]. Amines can be used to open the modified metaphosphating **174**, as recently shown using tryptophan methyl ester yielding **178** in 30% yield [112]. A notable modification of this strategy reacts the phosphordichloridite **179** with a non-hydrolyzable analog of pyrophosphate followed by boronation to yield the intermediate (**Q**), an analog of **174** (Scheme 35c). (**Q**) is then hydrolyzed with triethylammonium bicarbonate to afford the heavily modified ATP analog **180** in 30% yield starting from **179** [113].

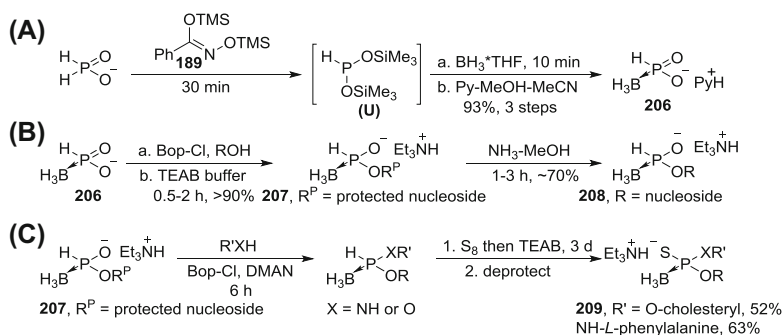
Much like cyclic P(III)–P(V) species, open-chain phosphite esters may also be boranated using the approach discussed in Scheme 36. Disaccharide-derived phosphite triester **181** undergoes smooth boronation to yield the disaccharide boranophosphate **182**, which after oligomerization is demethylated to afford the boranophosphodiester-linked polysaccharides of type **183** (Scheme 36a) [114]. Note that the yield for **182** is reported including two preceding steps, so the boronation is likely more efficient than shown. Similarly, phosphonites **185**, derived from their precursor thiophosphonates **184**, undergo boronation (**186a,b,d**) in good yields with retention of configuration (Scheme 36b) [115]. The preceding reduction (Raney Ni) was reported as a more troublesome step, which accounts for lower yields compared to the previous example; the relative stereochemistry of R and R' groups is the main consideration for the efficiency of this reduction. Subsequent deprotection leads to boranophosphates of type **187** in good yields (97%). H-phosphonates **188** do not readily undergo boronation without prior activation (Scheme 36c). Reagent **189** is used to circumvent the diminished reactivity of H-phosphonate **188** by silylation. Silylation shifts the equilibrium of **188** to the more reactive phosphite tautomer **190**, which can then be boranated under standard conditions. Subsequent deprotection of the nucleoside yields the mono-substituted boranophosphate **191** [116].

7.2 P-BH₃ Species as Nucleophiles

Inorganic boranophosphate **192**, or BPi, is an analog of inorganic phosphate; BPi is prepared by boranation of tris-trimethylsilylphosphite with a borane source (Scheme 37a) [117]. BPi displays similar acid–base properties [106] and reactivity of inorganic phosphate and acts as a nucleophile toward imidazole-activated phosphates. Reactions of NMP-imidazolides **193** with BPi and MgCl₂ activator yield the intermediate β-borano NDPs **194**. **194** can then be reacted with another equivalent of **193** to yield homodimers **195a** or a different NMP-imidazole to yield Npp(B)pN hetero-dimers **195b** (Scheme 37b) [118]. Heterodimers **195b** are prepared in lower yield compared to that of homodimers **195a**, presumably because of competing hydrolysis. BPi may also be used as a nucleophile in the synthesis of Npp(B)ppNs **198**. In this case, mixtures of NMPs **196** and NDPs **197** are activated by CDI in one flask (Scheme 37c), followed by the addition of BPi [111]. BPi monoesters, such as **199**, retain their nucleophilic character and may be elaborated further when reacted with imidazole phosphates of type **200** (Scheme 37d); Np(B)ppNs or Np(B)pps **201** are obtained using this method [118]. MgCl₂ addition is required in all cases to boost the nucleophilicity of BPi.



Scheme 37 Examples of P-BH₃ nucleophiles



Scheme 38 Reactions of P-BH₃ electrophiles

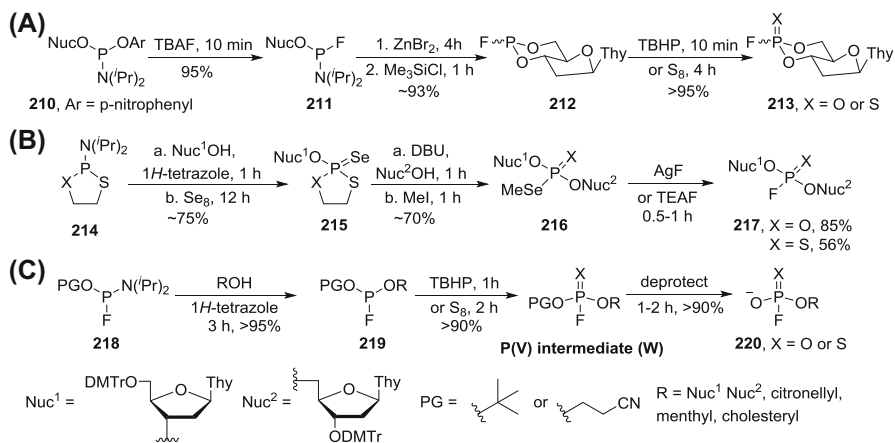
Boranophosphonates may be accessed using the borano-*H*-phosphonite **203b** reagent obtained from **202** and LiBH₄ (Scheme 37e) [119]. **203** is deprotonated with *n*-BuLi and reacted with the modified pentose electrophile **204**. Separation of diastereomers followed by a base-mediated boranophosphonodiester formation yields the pure stereoisomers of boranophosphonates **205** [120].

7.3 P-BH₃ Species as Electrophiles

Like unmodified phosphates, P-BH₃ species may be activated to act as electrophiles. Borano-*H*-phosphonate **206** is prepared from phosphinic acid, via intermediate (**U**), as shown in Scheme 38a. **206** is then activated using Bop-Cl and attacked by the 5' alcohol of a nucleoside (Scheme 38b). The resulting species **207** is deprotected to yield **208**, a borano-*H*-phosphonate nucleoside; acidic nucleoside deprotection conditions caused boranophosphonate decomposition in this case [121]. A similar example repeats the coupling on boranophosphonate **207** with an alcohol or an amine (Scheme 38c). Sulfurization followed by deprotection on the nucleoside provides the boranothiophosphate diesters **209** in 52% or 63% overall yield [122].

8 Fluorophosphates

¹⁹F is a desirable label for probing biological systems: it can be read by NMR with 83% sensitivity relative to ¹H, it has a wide magnetic window, and its resonance is environment-sensitive. Fluorine is nearly absent from the realm of natural biology, allowing for undisturbed NMR observation in complex biological matrices [123]. Fluorophosphate-containing biomimetics often display minimally perturbed functionality, owing to the F-P and O-P bonds' electronic similarities [123]. A previous review may be consulted for methods appearing before 1999 [124]. The radioactive ¹⁸F, though very useful for PET imaging, will not be discussed as its short half-life (*t*_{1/2} ~ 110 min) renders it out of reach for many chemical methods. Recent synthetic methods can be divided into three categories:



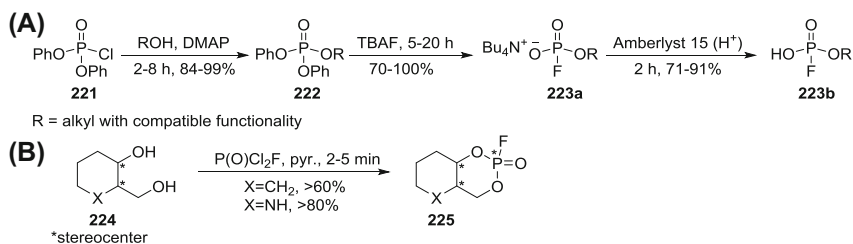
Scheme 39 Synthesis of fluorophosphates and their analogs using P(III) chemistry

P(III) chemistry, P(V) chemistry, and late-stage fluorination of biomolecules and their analogs.

8.1 P(III) Chemistry

Leaving group-bearing phosphoramidites of type **210** can be elaborated to P(III) fluorophosphite **211** using TBAF (Scheme 39a). Compound **211** contains a nucleoside attached at the 3' alcohol and an activatable diisopropylamine moiety. Deprotection of the nucleoside 5' alcohol with ZnBr₂ and activation of the amidite with TMSCl provide a cyclic intermediate **212**, which is then oxidized to the corresponding fluorophosphate or fluoro-thio-phosphate **213**. This method provides chromatographically stable materials in excellent yields, but requires water-free conditions and protection of nucleophilic moieties [125].

Another approach toward fluorophosphate diesters uses phosphoramidites of type **214**. Standard P(III) couplings with a nucleoside followed by oxidation with Se₈ yield interesting mononucleoside intermediates of type **215**, which are reacted with a second appropriately protected nucleoside and then with MeI to yield protected 3'–5' thymidine dimers **216** (Scheme 39b). Fluoride-mediated substitution of the methylselenoate and subsequent deprotection affords 3'–5' dithymidine-fluorophosphate dimers and their thio-analogs **217** [126]. P(III) chemistry also allows for preparation of fluoride-bearing reagents of type **218**; activation of the diisopropylamine moiety with 1*H*-tetrazole followed by substitution with an alcohol (Scheme 39c) affords the protected fluorophosphite **219**. Oxidation by TBHP or S₈ [intermediate (**W**)] and subsequent deprotection provides materials of type **220**. Multistep reagent synthesis and use of protecting groups are a drawback of this method; however, the routes are convergent, the yields are nearly quantitative, and all materials are stable to chromatography [127].



Scheme 40 Synthesis of fluorophosphates using P(V) chemistry

8.2 P(V) Chemistry

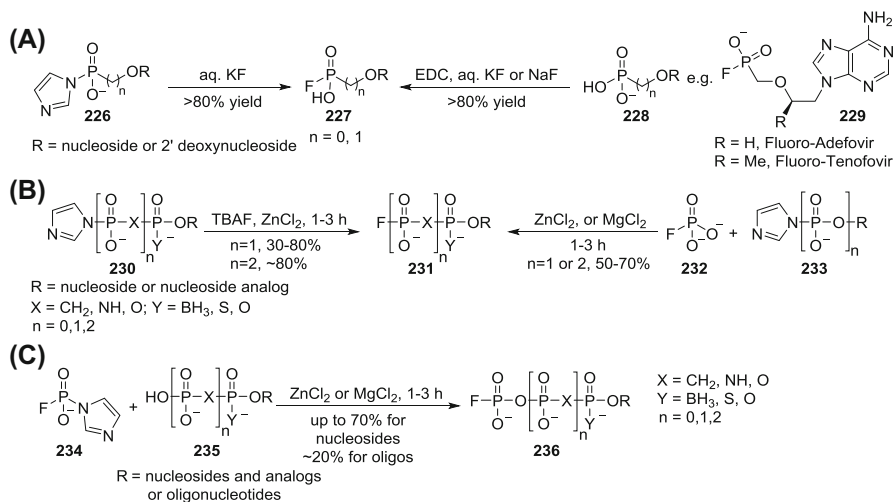
P(V) fluorophosphate intermediates are typically derived from phosphorus oxychloride and require use of protecting groups much like the aforementioned P(III) chemistry. In one approach, alcohols in the presence of DMAP are reacted with diphenylchlorophosphate **221** to yield the protected P(V) triesters **222** (Scheme 40a). The triesters are then fluorinated and deprotected in a single step to provide TBA salts **223**, which may be ion exchanged for H^+ using a sulfonic acid resin [128]. The tested scope of this method is relatively narrow but can likely be expanded to many appropriately protected substrates.

$P(O)Cl_2F$ is another example of a phosphorus oxychloride-derived reagent and is particularly useful for preparation of cyclic fluorophosphates as shown by Rüedi and coworkers (Scheme 40b). Using the 8 precursor stereoisomers of **224**, they prepared all 16 possible stereoisomers of **225** and used them in subsequent enzyme inhibition studies [129]. Difficult synthesis of $P(O)Cl_2F$ [130], use of a glove box, and potential for high toxicity of reagents and intermediates are the main drawbacks of this method.

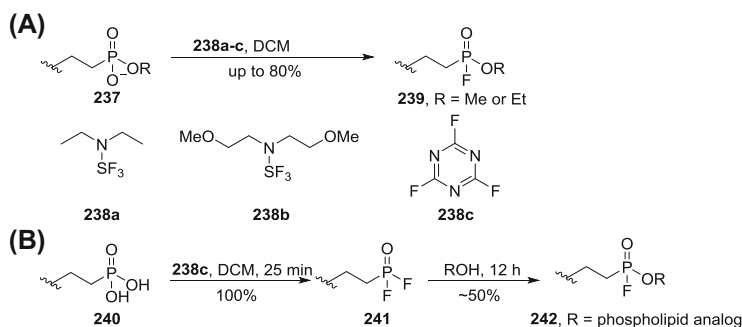
8.3 Late-Stage Fluorination of Biomolecules and Their Analogs

Direct fluorination is an attractive strategy for synthesis of fluorophosphates when the desired precursors are available. Preparation of nucleoside 5'-fluorophosphates or phosphonates of general structure **227** ($n = 0$ and 1 represents phosphate and phosphonate, respectively) from their activated imidazolides **226** was reported to proceed in aqueous solutions of KF (Scheme 41a). An updated method from the same group showed that the imidazolides may be replaced with EDC activation of unmodified nucleoside 5'-phosphates or phosphonates of type **228**. This method was applied successfully to convert the phosphonate-containing HIV drugs adefovir and tenofovir to their respective fluorophosphonate **229**. The ease of this method is notable; however, scalability is unclear because of scant experimental details of these reports [131, 132].

A similar method employed nucleoside imidazolides of type **230** in direct fluorinations using TBAF with $ZnCl_2$ (Scheme 41b); non-natural and non-hydrolyzable analogs were also prepared. A minor method adjustment demonstrated incorporation of an additional phosphate unit using fluorophosphate **232** in the



Scheme 41 Late-stage fluorination of nucleoside derivatives



Scheme 42 Late-stage fluorination of phosphonates

presence of zinc or magnesium chlorides. This process variation converts natural nucleoside monophosphates **233** into β -fluorodiphosphates **231** ($n = 1$), but also holds potential for synthesis of nucleoside γ -fluorophosphates **231** ($n = 2$). For cases where prior synthesis of nucleoside imidazolides is impractical, the authors developed fluorophosphate imidazole **234**, which adds a fluorophosphate unit onto un-activated nucleosides and derivatives **235** (Scheme 41c). Products of type **236** are prepared from a number of natural and unnatural analogs, as well as oligonucleotides, demonstrating the wide scope of reagent **234** [123]. These methods can likely be extended to substrates outside of the nucleoside family as they display high functional group tolerance.

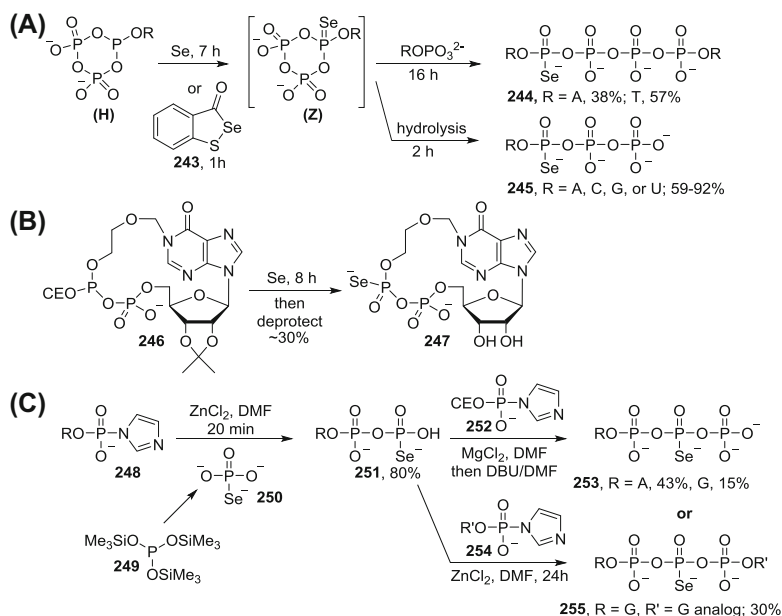
Fluorophosphonate diesters have utility in activity-based protein profiling [133]. These products are most commonly accessed through late-stage fluorination of phosphonate diesters of type **237** to yield fluorophosphonates **239** (Scheme 42a). Amino-sulfur trifluoride reagent **238a** (DAST) or Deoxo-Fluor **238b** [134] are

commonly employed for these transformations, with cyanuric fluoride **238c** [135, 136] also finding recent utility. For compounds with incompatible functionality to reagents **238a–c**, an alternate synthetic strategy may be employed. In one example, reagent **238c** is used to prepare a difluorophosphonate **241** from its precursor phosphonic acid **240** (Scheme 42b). One of the fluorides is then replaced by an alcohol of a phospholipid analog yielding the target material **242**. The more sensitive phospholipid analog portion **242** is never exposed to strong fluorinating agents using this approach.

9 Selenophosphates

Selenophosphate is a rare motif represented by only a handful of recent works; however, the use of Se in X-ray phasing makes selenophosphate biomimetics intriguing synthetic targets [137–139]. Selenium is introduced onto a phosphate nearly exclusively through P(III) intermediates, typically en route to nucleotide derivatives. Note that O-to-Se substitution in non-symmetrical phosphodiester creates a stereocenter, resulting in diastereomeric mixtures. For simplicity, all subsequent yields are reported as ~1:1 mixtures of diastereomers.

Cyclic mixed P(V)–P(III) intermediates (**H**) can be oxidized with either elemental selenium or the commercially available reagent **243** to yield



Scheme 43 Selenophosphates

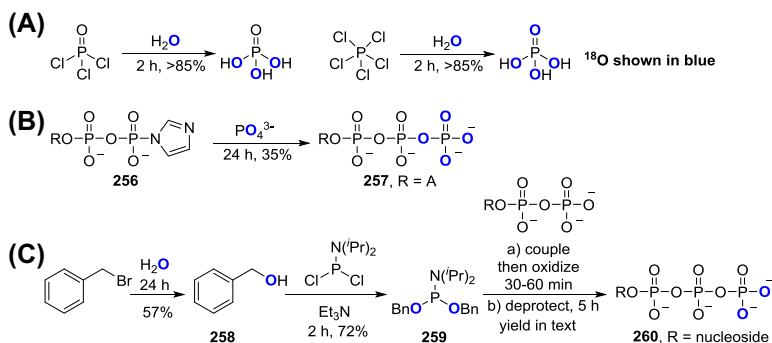
selenophosphates (**Z**) (Scheme 43a). Type (**Z**) intermediates undergo nucleophilic ring opening with a nucleoside phosphate to yield nucleoside homodimers of type **244** or with water to yield nucleoside triphosphates **245** with a selenium on the α -phosphate [140, 141]. A similar P(III)-to-P(V) selenium oxidation was used in the synthesis macrocycle **247** from its protected precursor **246** (Scheme 43b) [142].

Selenium may be installed on the β -phosphate of NTPs using selenophosphate **250**, which is derived from the P(III) species **249** via a Se_8 oxidation (Scheme 43c). **250** is used as a nucleophile with activated NMP-imidazolides **248** to prepare seleno-NDPs **251**. The nucleophilic attack can then be repeated on imidazolide **252** to yield NTP analogs **253** with a microwave-assisted reaction time of 5 min or 2 h unassisted. **251** may also be reacted with an NMP **254** to afford an Npp(Se)pN heterodimer **255**; in either case, an excess of MgCl_2 or ZnCl_2 is necessary to promote the reaction (Scheme 43c) [143, 144]. The often modest yields and a dearth of selenium-containing reagents highlight the challenging preparation of selenophosphate compounds.

10 ^{18}O -Labeled Phosphates

Oxygen-18 is a useful MS handle for studying kinase activity [145–147], which makes ^{18}O a desirable phosphate label. Notably, the ATP ^{18}O content can also be followed by ^{31}P -NMR, allowing for real-time ^{18}O monitoring [148]. ^{18}O -phosphate derivatives are rare as biological probes because of a dearth of synthetic methods.

^{18}O is typically purchased in the form of H_2^{18}O , which acts as a starting point for most syntheses. Tri- or tetra- ^{18}O -labeled phosphate can be obtained via hydrolysis of $\text{P}(\text{O})\text{Cl}_3$ or PCl_5 in >85% yield (Scheme 44a) [149]. The labeled $^{18}\text{O}_4$ -phosphoric acid can then be enzymatically incorporated into tetra-labeled γ - $^{18}\text{O}_4$ -ATP or γ - $^{18}\text{O}_4$ -GTP from the precursor NDPs on a 50-nmol scale [150]. For larger preparations, activated ADP-imidazolide **256** is reacted with $\text{P}^{18}\text{O}_4^{3-}$ tetrabutylammonium salt to yield γ - $^{18}\text{O}_4$ -ATP **257** on a 50-mg scale



Scheme 44 Synthesis of ^{18}O -labeled phosphates

(Scheme 44b) [151]. Another approach uses benzyl- ^{18}O -alcohol **258**, which is derived from benzylbromide, and is then incorporated into the phosphoramidite reagent **259** (Scheme 44c). Activation of **259** with 5-ethylthio-1*H*-tetrazole in the presence of NDPs followed by oxidation and deprotection yields the bis- ^{18}O -labeled NTPs **260** in 95% yield and 80% purity with precursor nucleoside diphosphates as the main contaminants. Upon further purification, pure γ -[$^{18}\text{O}_2$]-ATP, γ -[$^{18}\text{O}_2$]-UTP, and γ -[$^{18}\text{O}_2$]-5,6-dihydro-UTP were obtained on up to a 10-mg scale in 44–52% yield [39].

Acknowledgements Funding was provided by Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (Grant No. SNF, PP00P2_157607) and HFSP Organization (Grant No. RGP0025/2016).

References

- Elliott TS, Slowey A, Ye Y, Conway SJ (2012) The use of phosphate bioisosteres in medicinal chemistry and chemical biology. *Med Chem Commun* 3:735–751. doi:10.1039/c2md20079
- Staab HA (1962) New methods of preparative organic chemistry IV. Syntheses using heterocyclic amides (azolides). *Angew Chem Int Ed* 1:351–367
- Russell MA, Laws AP, Atherton JH, Page MI (2008) The mechanism of the phosphoramidite synthesis of polynucleotides. *Org Biomol Chem* 6:3270–3275. doi:10.1039/b808999j
- Westheimer FH (1987) Why nature chose phosphates. *Science* 235:1173–1178
- Rao F, Cha J, Xu J et al (2014) Inositol pyrophosphates mediate the DNA-PK/ATM-p53 cell death pathway by regulating CK2 phosphorylation of Tti1/Tel2. *Mol Cell* 54:119–132. doi:10.1016/j.molcel.2014.02.020
- Prasad A, Jia Y, Chakraborty A et al (2011) Inositol hexakisphosphate kinase 1 regulates neutrophil function in innate immunity by inhibiting phosphatidylinositol-(3,4,5)-trisphosphate signaling. *Nat Immunol* 12:752–760. doi:10.1038/ni.2052
- Pulloor NK, Nair S, Kostic AD et al (2014) Human genome-wide RNAi screen identifies an essential role for inositol pyrophosphates in type-I interferon response. *PLoS Pathog* 10:e1003981. doi:10.1371/journal.ppat.1003981.s006
- Choi K, Mollapour E, Choi JH, Shears SB (2008) Cellular energetic status supervises the synthesis of bis-diphosphoinositol tetrakisphosphate independently of AMP-activated protein kinase. *Mol Pharmacol* 74:527–536. doi:10.1124/mol.107.044628
- Szjgyarto Z, Garedew A, Azevedo C, Saiardi A (2011) Influence of inositol pyrophosphates on cellular energy dynamics. *Science* 334:802–805. doi:10.1126/science.1207306
- Brown NW, Marmelstein AM, Fiedler D (2016) Chemical tools for interrogating inositol pyrophosphate structure and function. *Chem Soc Rev* 45:6311–6326. doi:10.1039/C6CS00193A
- Riley AM, Wang H, Shears SB, Potter BVL (2015) *Chem Commun* 51:12605–12608. doi:10.1039/C5CC05017K
- Antczak MI, Montchamp J-L (2009) Reactions of α -boranophosphorus compounds with electrophiles: alkylation, acylation, and other reactions. *J Org Chem* 74:3758–3766. doi:10.1021/jo900300c
- Antczak MI, Montchamp J-L (2008) Synthesis of 1,1-bis-phosphorus compounds from organoboranes. *Tetrahedron Lett* 49:5909–5913. doi:10.1016/j.tetlet.2008.07.144
- Gavara L, Petit C, Montchamp J-L (2012) ChemInform abstract: DBU-promoted alkylation of alkyl phosphinates and H-phosphonates. *Tetrahedron Lett* 53:5000–5003. doi:10.1016/j.tetlet.2012.07.019
- Gelat F, Lacomme C, Berger O et al (2015) Synthesis of (phosphonomethyl)phosphinate pyrophosphate analogues via the phospho-Claisen condensation. *Org Biomol Chem* 13:825–833. doi:10.1039/C4OB02007C
- Taylor SD, Mirzaei F, Bearne SL (2006) An unsymmetrical approach to the synthesis of bis-methylene triphosphate analogues. *Org Lett* 8:4243–4246. doi:10.1021/ol0615432

17. Medved TY, Polikarpov YM, Pisareva SA (1968) Phosphine oxides and phosphorus acids, containing several = P(O)CH₂ groups in the molecule. *Russ Chem Bull* 17:1959–1965
18. Hauryliuk V, Atkinson GC, Murakami KS et al (2015) Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Publ Group* 13:298–309. doi:[10.1038/nrmicro3448](https://doi.org/10.1038/nrmicro3448)
19. Wexselblatt E, Katzhendler J, Saleem-Batcha R et al (2010) ppGpp analogues inhibit synthetase activity of Rel proteins from Gram-negative and Gram-positive bacteria. *Bioorg Med Chem* 18:4485–4497. doi:[10.1016/j.bmc.2010.04.064](https://doi.org/10.1016/j.bmc.2010.04.064)
20. Engelsma SB, Meeuwenoord NJ, Overkleef HS et al (2017) Combined phosphoramidite-phosphodiester reagents for the synthesis of methylene bisphosphonates. *Angew Chem Int Ed* 56:2955–2959. doi:[10.1002/ange.201611878](https://doi.org/10.1002/ange.201611878)
21. Oka N, Shimizu M, Saigo K, Wada T (2006) 1,3-Dimethyl-2-(3-nitro-1,2,4-triazol-1-yl)-2-pyrrolidin-1-yl-1,3,2-diazaphospholidinium hexafluorophosphate (MNTP): a powerful condensing reagent for phosphate and phosphonate esters. *Tetrahedron* 62:3667–3673. doi:[10.1016/j.tet.2006.01.084](https://doi.org/10.1016/j.tet.2006.01.084)
22. Prakash G, Zibinsky M, Upton TG et al (2010) Synthesis and biological evaluation of fluorinated deoxynucleotide analogs based on bis-(difluoromethylene) triphosphoric acid. *Proc Natl Acad Sci USA* 107:15693–15698. doi:[10.1073/pnas.1007430107](https://doi.org/10.1073/pnas.1007430107)
23. Chunikhin KS, Kadyrov AA, Pasternak PV, Chkanikov ND (2010) Difluoromethylenephosphonates: synthesis and transformations. *Russ Chem Rev* 79:371–396. doi:[10.1070/RC2010v079n05ABEH003883](https://doi.org/10.1070/RC2010v079n05ABEH003883)
24. Ivanova MV, Bayle A, Besset T et al (2016) New prospects toward the synthesis of difluoromethylated phosphate mimics. *Chem Eur J* 22:10284–10293. doi:[10.1002/chem.201601310](https://doi.org/10.1002/chem.201601310)
25. Batra VK, Pedersen LC, Beard WA et al (2010) Halogenated β, γ-methylene- and ethylidene-dGTP-DNA ternary complexes with DNA polymerase β: structural evidence for stereospecific binding of the fluoromethylene analogues. *J Am Chem Soc* 132:7617–7625. doi:[10.1021/ja909370k](https://doi.org/10.1021/ja909370k)
26. Evans DA, Britton TC, Ellman JA, Dorow RL (1990) The asymmetric synthesis of α-amino acids. Electrophilic azidation of chiral imide enolates, a practical approach to the synthesis of (R)- and (S)-α-azido carboxylic acids. *J Am Chem Soc* 112:4011–4030
27. Benati L, Nanni D, Spagnolo P (1999) Reactions of benzocyclic β-keto esters with sulfonyl azides. 2. further insight into the influence of azide structure and solvent on the reaction course. *J Org Chem* 64:5132–5138. doi:[10.1021/jo9901541](https://doi.org/10.1021/jo9901541)
28. Wurz RP, Lin W, Charette AB (2003) Trifluoromethanesulfonyl azide: an efficient reagent for the preparation of α-cyano-α-diazo carbonyls and an α-sulfonyl-α-diazo carbonyl. *Tetrahedron Lett* 44:8845–8848. doi:[10.1016/j.tetlet.2003.09.197](https://doi.org/10.1016/j.tetlet.2003.09.197)
29. Chamberlain BT, Upton TG, Kashemirov BA, McKenna CE (2011) α-Azido bisphosphonates: synthesis and nucleotide analogues. *J Org Chem* 76:5132–5136. doi:[10.1021/jo200045a](https://doi.org/10.1021/jo200045a)
30. Upton TG, Kashemirov BA, McKenna CE et al (2009) α, β-Difluoromethylene deoxynucleoside 5'-triphosphates: a convenient synthesis of useful probes for DNA polymerase β structure and function. *Org Lett* 11:1883–1886. doi:[10.1021/ol701755k](https://doi.org/10.1021/ol701755k)
31. Wu Y, Zakharova VM, Kashemirov BA et al (2012) β, γ-CHF- and β, γ-CHCl-dGTP diastereomers: synthesis, discrete 31P NMR Signatures, and absolute configurations of new stereochemical probes for DNA polymerases. *J Am Chem Soc* 134:8734–8737. doi:[10.1021/ja300218x](https://doi.org/10.1021/ja300218x)
32. Kim SM, Kim HR, Kim DY (2005) Catalytic enantioselective fluorination and amination of β-keto phosphonates catalyzed by chiral palladium complexes. *Org Lett* 7:2309–2311. doi:[10.1021/ol050413a](https://doi.org/10.1021/ol050413a)
33. Hamashima Y, Suzuki T, Shimura Y et al (2005) An efficient catalytic enantioselective fluorination of β-ketophosphonates using chiral palladium complexes. *Tetrahedron Lett* 46:1447–1450. doi:[10.1016/j.tetlet.2005.01.018](https://doi.org/10.1016/j.tetlet.2005.01.018)
34. Kang Y, Cho M, Kim S, Kim D (2007) Asymmetric electrophilic fluorination of α-cyanoalkylphosphonates-catalyzed by chiral palladium complexes. *Synlett* 2007:1135–1138. doi:[10.1055/s-2007-977436](https://doi.org/10.1055/s-2007-977436)
35. Moriya K-I, Hamashima Y, Sodeoka M (2007) Pd(II)-catalyzed asymmetric fluorination of α-Aryl-α-cyanophosphonates with the aid of 2,6-lutidine. *Synlett* 2007:1139–1142. doi:[10.1055/s-2007-977437](https://doi.org/10.1055/s-2007-977437)
36. Oliveira FM, Barbosa LCA, Ismail FMD (2014) The diverse pharmacology and medicinal chemistry of phosphoramidates—a review. *RSC Adv* 4:18998–19012. doi:[10.1039/c4ra01454e](https://doi.org/10.1039/c4ra01454e)

37. Paul S, Caruthers MH (2016) Synthesis of phosphorodiamidate morpholino oligonucleotides and their chimeras using phosphoramidite chemistry. *J Am Chem Soc* 138:15663–15672. doi:[10.1021/jacs.6b08854](https://doi.org/10.1021/jacs.6b08854)
38. Guga P, Koziolkiewicz M (2011) Phosphorothioate nucleotides and oligonucleotides—recent progress in synthesis and application. *Chem Biodivers* 8:1642–1681. doi:[10.1002/cbdv.201100130](https://doi.org/10.1002/cbdv.201100130)
39. Hofer A, Cremosnik GS, Müller AC et al (2015) A modular synthesis of modified phosphoanhydrides. *Chem Eur J* 21:10116–10122. doi:[10.1002/chem.201500838](https://doi.org/10.1002/chem.201500838)
40. Cremosnik GS, Hofer A, Jessen HJ (2013) Iterative synthesis of nucleoside oligophosphates with phosphoramidites. *Angew Chem Int Ed* 53:286–289. doi:[10.1002/anie.201306265](https://doi.org/10.1002/anie.201306265)
41. Hofer A, Marques E, Kieliger N et al (2016) Chemoselective dimerization of phosphates. *Org Lett* 18:3222–3225. doi:[10.1021/acs.orglett.6b01466](https://doi.org/10.1021/acs.orglett.6b01466)
42. Nadel Y, Lecka J, Gilad Y et al (2014) Highly potent and selective ectonucleotide pyrophosphatase/phosphodiesterase I inhibitors based on an adenosine 5'-(α or γ)-thio-(α , β - or β , γ)-methylene-triphosphate scaffold. *J Med Chem* 57:4677–4691. doi:[10.1021/jm500196c](https://doi.org/10.1021/jm500196c)
43. Thillier Y, Sallamand C, Baraguey C et al (2014) Solid-phase synthesis of oligonucleotide 5'-(α -P-Thio)triphosphates and 5'-(α -P-thio)(β , γ -methylene)triphosphates. *Eur J Org Chem* 2015:302–308. doi:[10.1002/ejoc.201403381](https://doi.org/10.1002/ejoc.201403381)
44. Iyer RP, Egan W, Regan JB, Beaucage SL (1990) 3H-1, 2-Benzodithiole-3-one 1, 1-dioxide as an improved sulfurylating reagent in the solid-phase synthesis of oligodeoxyribonucleoside phosphorothioates. *J Am Chem Soc* 112:1254–1255
45. Stec WJ, Grajkowski A, Koziolkiewicz M, Uznanski B (1991) Novel route to oligo(deoxyribonucleoside phosphorothioates). Stereocontrolled synthesis of P-chiral oligo(deoxyribonucleoside phosphorothioates). *Nucl Acids Res* 19:5883–5888
46. Oka N, Wada T, Saigo K (2003) An oxazaphospholidine approach for the stereocontrolled synthesis of oligonucleoside phosphorothioates. *J Am Chem Soc* 125:8307–8317. doi:[10.1021/ja034502z](https://doi.org/10.1021/ja034502z)
47. Nukaga Y, Yamada K, Ogata T et al (2012) Stereocontrolled solid-phase synthesis of phosphorothioate oligoribonucleotides using 2'-O-(2-cyanoethoxymethyl)-nucleoside 3'-O-oxazaphospholidine monomers. *J Org Chem* 77:7913–7922. doi:[10.1021/jo301052v](https://doi.org/10.1021/jo301052v)
48. Jahns H, Roos M, Imig J et al (2015) Stereochemical bias introduced during RNA synthesis modulates the activity of phosphorothioate siRNAs. *Nat Commun* 6:1–9. doi:[10.1038/ncomms7317](https://doi.org/10.1038/ncomms7317)
49. Krakowiak A, Pęcherzewska R, Kaczmarek R et al (2011) Bioorganic & medicinal chemistry. *Bioorg Med Chem* 19:5053–5060. doi:[10.1016/j.bmc.2011.06.028](https://doi.org/10.1016/j.bmc.2011.06.028)
50. Kaczmarek R, Krakowiak A, Korczyński D et al (2016) Bioorganic & medicinal chemistry. *Bioorg Med Chem* 24:5068–5075. doi:[10.1016/j.bmc.2016.08.027](https://doi.org/10.1016/j.bmc.2016.08.027)
51. Yang X, Mierzejewski E (2010) Synthesis of nucleoside and oligonucleoside dithiophosphates. *New J Chem* 34:805. doi:[10.1039/b9nj00618d](https://doi.org/10.1039/b9nj00618d)
52. Li N-S, Frederiksen JK, Piccirilli JA (2012) Automated solid-phase synthesis of RNA oligonucleotides containing a nonbridging phosphorodithioate linkage via phosphorothioamidites. *J Org Chem* 77:9889–9892. doi:[10.1021/jo301834p](https://doi.org/10.1021/jo301834p)
53. Yang X, Sierant M, Janicka M et al (2012) Gene silencing activity of siRNA molecules containing phosphorodithioate substitutions. *ACS Chem Biol* 7:1214–1220. doi:[10.1021/cb300078e](https://doi.org/10.1021/cb300078e)
54. Li N-S, Frederiksen JK, Piccirilli JA (2011) Synthesis, properties, and applications of oligonucleotides containing an RNA dinucleotide phosphorothioate linkage. *Acc Chem Res* 44:1257–1269. doi:[10.1021/ar200131t](https://doi.org/10.1021/ar200131t)
55. Eguagie O, Cooke LA, Martin PML et al (2016) Synthesis of novel pyrophosphorothiolate-linked dinucleoside cap analogues in a ball mill. *Org Biomol Chem* 14:1201–1205. doi:[10.1039/c5ob02061a](https://doi.org/10.1039/c5ob02061a)
56. Meltzer D, Nadel Y, Lecka J et al (2013) Nucleoside-(5' \rightarrow P) methylenebisphosphonodithioate analogues: synthesis and chemical properties. *J Org Chem* 78:8320–8329. doi:[10.1021/jo400931n](https://doi.org/10.1021/jo400931n)
57. Amir A, Sayer AH, Zagalsky R et al (2013) O, O'-Diester methylenediphosphonotetrathioate: synthesis, characterization, and potential applications. *J Org Chem* 78:270–277. doi:[10.1021/jo301786m](https://doi.org/10.1021/jo301786m)
58. Gryaznov SM (2010) Oligonucleotide N3' \rightarrow P5' phosphoramidates and thio-phosphoramidates as potential therapeutic agents. *Chem Biodivers* 7:477–493. doi:[10.1002/cbdv.200900187](https://doi.org/10.1002/cbdv.200900187)
59. Pongracz K, Gryaznov S (1999) Oligonucleotide N3' \rightarrow P5' thiophosphoramidates: synthesis and properties. *Tetrahedron Lett* 40:7661–7664
60. Herbert B-S, Pongracz K, Shay JW et al (2002) Oligonucleotide N3' \rightarrow P5' phosphoramidates as efficient telomerase inhibitors. *Oncogene* 21:638–642. doi:[10.1038/sj.onc.1205064](https://doi.org/10.1038/sj.onc.1205064)

61. Wagner GK, Pesnot T, Field RA (2009) A survey of chemical methods for sugar-nucleotide synthesis. *Nat Prod Rep* 26:1172–1194. doi:[10.1039/b909621n](https://doi.org/10.1039/b909621n)
62. Trmčić M, Hodgson DRW (2011) Synthesis of thiophosphoramidates in water: click chemistry for phosphates. *Chem Commun* 47:6156–6158. doi:[10.1039/c1cc11586c](https://doi.org/10.1039/c1cc11586c)
63. Jessen HJ, Ahmed N, Hofer A (2014) Phosphate esters and anhydrides—recent strategies targeting nature’s favoured modifications. *Org Biomol Chem* 12:3526–3530. doi:[10.1039/c4ob00478g](https://doi.org/10.1039/c4ob00478g)
64. Trmčić M, Chadbourn FL, Brear PM et al (2013) Aqueous synthesis of N, S-dialkylthiophosphoramidates: design, optimisation and application to library construction and antileishmanial testing. *Org Biomol Chem* 11:2660. doi:[10.1039/c3ob27448a](https://doi.org/10.1039/c3ob27448a)
65. Conway LP, Delley RJ, Neville J et al (2014) The aqueous N-phosphorylation and N-thiophosphorylation of aminonucleosides. *RSC Adv* 4:38663–38671. doi:[10.1039/C4RA08317B](https://doi.org/10.1039/C4RA08317B)
66. Conway LP, Mikkola S, O’Donoghue AC, Hodgson DRW (2016) The synthesis, conformation and hydrolytic stability of an N, S-bridging thiophosphoramidate analogue of thymidyl-3’,5’-thymidine. *Org Biomol Chem* 14:7361–7367. doi:[10.1039/C6OB01270A](https://doi.org/10.1039/C6OB01270A)
67. Korlach J, Bibillo A, Wegener J et al (2008) Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides Nucl Acids* 27:1072–1082. doi:[10.1080/15257770802260741](https://doi.org/10.1080/15257770802260741)
68. Kumar S, Sood A, Wegener J et al (2005) Terminal phosphate labeled nucleotides: synthesis, applications, and linker effect on incorporation by dna polymerases. *Nucleosides Nucleotides Nucl Acids* 24:401–408. doi:[10.1081/NCN-200059823](https://doi.org/10.1081/NCN-200059823)
69. Johnson SA, Hunter T (2005) Kinomics: methods for deciphering the kinome. *Nat Meth* 2:17–25. doi:[10.1038/nmeth731](https://doi.org/10.1038/nmeth731)
70. Satake W, Nakabayashi Y, Mizuta I et al (2009) Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson’s disease. *Nat Genet* 41:1303–1307. doi:[10.1038/ng.485](https://doi.org/10.1038/ng.485)
71. Cohen P (2002) Protein kinases—the major drug targets of the twenty-first century? *Nat Rev Drug Discov* 1:309–315. doi:[10.1038/nrd773](https://doi.org/10.1038/nrd773)
72. Manning G, Whyte DB, Martinez R, Hunter T (2002) The protein kinase complement of the human genome. *Science* 298:1912–1934
73. Statsuk AV, Maly DJ, Seeliger MA et al (2008) Tuning a three-component reaction for trapping kinase substrate complexes. *J Am Chem Soc* 130:17568–17574. doi:[10.1021/ja807066f](https://doi.org/10.1021/ja807066f)
74. Blethrow JD, Glavy JS, Morgan DO, Shokat KM (2008) Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc Natl Acad Sci USA* 105:1442–1447. doi:[10.1073/pnas.0708966105](https://doi.org/10.1073/pnas.0708966105)
75. Suwal S, Pflum MKH (2010) Phosphorylation-dependent kinase-substrate cross-linking. *Angew Chem Int Ed* 49:1627–1630. doi:[10.1002/anie.200905244](https://doi.org/10.1002/anie.200905244)
76. Hacker SM, Mex M, Marx A (2012) Synthesis and stability of phosphate modified ATP analogues. *J Org Chem* 77:10450–10454. doi:[10.1021/jo301923p](https://doi.org/10.1021/jo301923p)
77. Lee SE, Elphick LM, Anderson AA et al (2009) Bioorganic & medicinal chemistry letters. *Bioorg Med Chem Lett* 19:3804–3807. doi:[10.1016/j.bmcl.2009.04.028](https://doi.org/10.1016/j.bmcl.2009.04.028)
78. Ratnakar SJ, Alexander V (2005) Synthesis and relaxivity studies of a gadolinium(III) complex of ATP-conjugated DO3A as a contrast enhancing agent for MRI. *Eur J Inorg Chem* 2005:3918–3927. doi:[10.1002/ejic.200401018](https://doi.org/10.1002/ejic.200401018)
79. Martić S, Labib M, Freeman D, Kraatz H-B (2011) Probing the role of the linker in ferrocene-ATP conjugates: monitoring protein kinase catalyzed phosphorylations electrochemically. *Chem Eur J* 17:6744–6752. doi:[10.1002/chem.201003535](https://doi.org/10.1002/chem.201003535)
80. Song H, Kerman K, Kraatz H-B (2008) Electrochemical detection of kinase-catalyzed phosphorylation using ferrocene-conjugated ATP. *Chem Commun*. doi:[10.1039/B714383D](https://doi.org/10.1039/B714383D)
81. Green KD, Pflum MKH (2007) Kinase-catalyzed biotinylation for phosphoprotein detection. *J Am Chem Soc* 129:10–11. doi:[10.1021/ja066828o](https://doi.org/10.1021/ja066828o)
82. Parang K, Kohn JA, Saldanha SA, Cole PA (2002) Development of photo-crosslinking reagents for protein kinase–substrate interactions. *FEBS Lett* 520:156–160
83. Korhonen HJ, Conway LP, Hodgson DR (2014) ScienceDirectPhosphate analogues in the dissection of mechanism. *Curr Opin Chem Biol* 21:63–72. doi:[10.1016/j.cbpa.2014.05.001](https://doi.org/10.1016/j.cbpa.2014.05.001)
84. Gao X, Schutz-Geschwender A, Hardwidge PR (2008) Near-infrared fluorescence detection of ATP-biotin-mediated phosphoprotein labeling. *Biotechnol Lett* 31:113–117. doi:[10.1007/s10529-008-9824-0](https://doi.org/10.1007/s10529-008-9824-0)

85. Senevirathne C, Pflum MKH (2013) Biotinylated phosphoproteins from kinase-catalyzed biotinylation are stable to phosphatases: implications for phosphoproteomics. *ChemBioChem* 14:381–387. doi:[10.1002/cbic.201200626](https://doi.org/10.1002/cbic.201200626)
86. Dunn JD, Reid GE, Bruening ML (2009) Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. *Mass Spectrom Rev* 29:29–54. doi:[10.1002/mas.20219](https://doi.org/10.1002/mas.20219)
87. Fouda AE, Pflum MKH (2015) A cell-permeable ATP analogue for kinase-catalyzed biotinylation. *Angew Chem* 127:9754–9757. doi:[10.1002/ange.201503041](https://doi.org/10.1002/ange.201503041)
88. Eid J, Fehr A, Gray J et al (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138. doi:[10.1126/science.1162986](https://doi.org/10.1126/science.1162986)
89. Fuller CW, Kumar S, Porel M et al (2016) Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. *Proc Natl Acad Sci USA* 113:5233–5238. doi:[10.1073/pnas.1601782113](https://doi.org/10.1073/pnas.1601782113)
90. Hornung V, Ellegast J, Kim S et al (2006) 5'-Triphosphate RNA is the ligand for RIG-I. *Science* 314:994–997. doi:[10.1126/science.1132505](https://doi.org/10.1126/science.1132505)
91. Schlee M (2013) Master sensors of pathogenic RNA – RIG-I like receptors. *Immunobiology* 218:1322–1335. doi:[10.1016/j.imbio.2013.06.007](https://doi.org/10.1016/j.imbio.2013.06.007)
92. Martinez-Gil L, Goff PH, Hai R et al (2013) A Sendai virus-derived RNA agonist of RIG-I as a virus vaccine adjuvant. *J Virol* 87:1290–1300. doi:[10.1128/JVI.02338-12](https://doi.org/10.1128/JVI.02338-12)
93. Kolakofsky D, Kowalinski E, Cusack S (2012) A structure-based model of RIG-I activation. *RNA* 18:2118–2127. doi:[10.1261/rna.035949.112](https://doi.org/10.1261/rna.035949.112)
94. Poeck H, Besch R, Maihofer C et al (2008) 5'-Triphosphate-siRNA: turning gene silencing and Rig-I activation against melanoma. *Nat Med* 14:1256–1263. doi:[10.1038/nm.1887](https://doi.org/10.1038/nm.1887)
95. Burgess K, Cook D (2000) Syntheses of nucleoside triphosphates. *Chem Rev* 100:2047–2060. doi:[10.1021/cr990045m](https://doi.org/10.1021/cr990045m)
96. Roy B, Depaix A, Périgaud C, Peyrottes S (2016) Recent trends in nucleotide synthesis. *Chem Rev* 116:7854–7897. doi:[10.1021/acs.chemrev.6b00174](https://doi.org/10.1021/acs.chemrev.6b00174)
97. Sun Q, Edathil JP, Wu R et al (2008) One-pot synthesis of nucleoside 5'-triphosphates from nucleoside 5'- H-phosphonates. *Org Lett* 10:1703–1706. doi:[10.1021/ol8003029](https://doi.org/10.1021/ol8003029)
98. Sproat BS, Rupp T, Menhardt N, Keane D (1999) Fast and simple purification of chemically modified hammerhead ribozymes using a lipophilic capture tag. *Nucl Acids Res* 27:1950–1955
99. Goldeck M, Tuschl T, Hartmann G, Ludwig J (2014) Efficient solid-phase synthesis of pppRNA by using product-specific labeling. *Angew Chem Int Ed* 53:4694–4698. doi:[10.1002/anie.201400672](https://doi.org/10.1002/anie.201400672)
100. Sarac I, Meier C (2015) Efficient automated solid-phase synthesis of DNA and RNA 5'-triphosphates. *Chem Eur J* 21:16421–16426. doi:[10.1002/chem.201502844](https://doi.org/10.1002/chem.201502844)
101. Merino P, Weinschenk L, Meier C (2013) Chemical syntheses of nucleoside triphosphates. In: Merino P (ed) *Chemical synthesis of nucleoside analogues*. Wiley, Hoboken. doi: [10.1002/9781118498088.ch5](https://doi.org/10.1002/9781118498088.ch5)
102. Fardet L, Fève B (2014) Systemic glucocorticoid therapy: a review of its metabolic and cardiovascular adverse events. *Drugs* 74:1731–1745. doi:[10.1007/s40265-014-0282-9](https://doi.org/10.1007/s40265-014-0282-9)
103. Kern JC, Cancilla M, Dooney D et al (2016) Discovery of pyrophosphate diesters as tunable, soluble, and bioorthogonal linkers for site-specific antibody–drug conjugates. *J Am Chem Soc* 138:1430–1445. doi:[10.1021/jacs.5b12547](https://doi.org/10.1021/jacs.5b12547)
104. Wu M, Chong LS, Perlman DH et al (2016) Inositol polyphosphates intersect with signaling and metabolic networks via two distinct mechanisms. *Proc Natl Acad Sci USA* 113:E6757–E6765. doi:[10.1073/pnas.1606853113](https://doi.org/10.1073/pnas.1606853113)
105. Wu M, Dul BE, Trevisan AJ, Fiedler D (2013) Synthesis and characterization of non-hydrolysable diphosphoinositol polyphosphate messengers. *Chem Sci* 4:405–410. doi:[10.1039/C2SC21553E](https://doi.org/10.1039/C2SC21553E)
106. Li P, Sergueeva ZA, Dobrikov M, Shaw BR (2007) Nucleoside and oligonucleoside boranophosphates: chemistry and properties. *Chem Rev* 107:4746–4796. doi:[10.1021/cr050009p](https://doi.org/10.1021/cr050009p)
107. Alayrac C, Lakhdar S, Abdellah I et al (2015) Recent advances in synthesis of P-BH₃ compounds. *Top Curr Chem* 361:1–82. doi:[10.1007/128_2014_565](https://doi.org/10.1007/128_2014_565)
108. Cheek MA, Sharaf ML, Dobrikov MI, Shaw BR (2013) Inhibition of hepatitis C viral RNA-dependent RNA polymerase by alpha-P-boranophosphate nucleotides: exploring a potential strategy for mechanism-based HCV drug design. *Antiviral Res* 98:144–152. doi:[10.1016/j.antiviral.2013.02.014](https://doi.org/10.1016/j.antiviral.2013.02.014)
109. Li P, Xu Z, Liu H et al (2005) Synthesis of α -P-modified nucleoside diphosphates with ethylenediamine. *J Am Chem Soc* 127:16782–16783. doi:[10.1021/ja055179y](https://doi.org/10.1021/ja055179y)

110. Ginsburg-Shmuel T, Haas M, Grbic D et al (2012) UDP made a highly promising stable, potent, and selective P2Y6-receptor agonist upon introduction of a boranophosphate moiety. *Bioorg Med Chem* 20:5483–5495. doi:[10.1016/j.bmc.2012.07.042](https://doi.org/10.1016/j.bmc.2012.07.042)
111. Yelovitch S, Camden J, Weisman GA, Fischer B (2012) Boranophosphate isoster controls P2Y-receptor subtype selectivity and metabolic stability of dinucleoside polyphosphate analogues. *J Med Chem* 55:437–448. doi:[10.1021/jm2013198](https://doi.org/10.1021/jm2013198)
112. Xu Z, Ramsay B (2015) Synthesis, hydrolysis, and protonation-promoted intramolecular reductive breakdown of potential NRTIs: stavudine α -P-borano- γ -P-N-l-tryptophanyltriphosphates. *Molecules* 20:18808–18826. doi:[10.3390/molecules201018808](https://doi.org/10.3390/molecules201018808)
113. Azran S, Förster D, Danino O et al (2013) Highly efficient biocompatible neuroprotectants with dual activity as antioxidants and P2Y receptor agonists. *J Med Chem* 56:4938–4952. doi:[10.1021/jm400197m](https://doi.org/10.1021/jm400197m)
114. Fujita S, Oka N, Matsumura F, Wada T (2011) Synthesis of oligo(α -D-glycosyl phosphate) derivatives by a phosphoramidite method via boranophosphate intermediates. *J Org Chem* 76:2648–2659. doi:[10.1021/jo102584g](https://doi.org/10.1021/jo102584g)
115. Ferry A, Guinchar X, Retailleau P, Crich D (2012) Synthesis, characterization, and coupling reactions of six-membered cyclic P-chiral ammonium phosphonite-boranes; reactive H-phosphinate equivalents for the stereoselective synthesis of glycomimetics. *J Am Chem Soc* 134:12289–12301. doi:[10.1021/ja305104b](https://doi.org/10.1021/ja305104b)
116. Xu Z, Sergueeva ZA, Shaw BR (2013) Synthesis and hydrolytic properties of thymidine boranomonophosphate. *Tetrahedron Lett* 54:2882–2885. doi:[10.1016/j.tetlet.2013.03.110](https://doi.org/10.1016/j.tetlet.2013.03.110)
117. Nahum V, Fischer B (2004) Boranophosphate salts as an excellent mimic of phosphate salts: preparation, characterization, and properties. *Eur J Inorg Chem* 2004:4124–4131. doi:[10.1002/ejic.200400142](https://doi.org/10.1002/ejic.200400142)
118. Kowalska J, Wypijewska del Nogal A, Darzynkiewicz ZM et al (2014) Synthesis, properties, and biological activity of boranophosphate analogs of the mRNA cap: versatile tools for manipulation of therapeutically relevant cap-dependent processes. *Nucl Acids Res* 42:10245–10264. doi:[10.1093/nar/gku757](https://doi.org/10.1093/nar/gku757)
119. Belabassi Y, Antczak MI, Tellez J, Montchamp J-L (2008) Borane complexes of the H₃PO₂ P(III) tautomer: useful phosphinate equivalents. *Tetrahedron* 64:9181–9190. doi:[10.1016/j.tet.2008.07.054](https://doi.org/10.1016/j.tet.2008.07.054)
120. Ferry A, Malik G, Retailleau P et al (2013) Alternative synthesis of P-chiral phosphonite-borane complexes: application to the synthesis of phosphonite–phosphonite dimers. *J Org Chem* 78:6858–6867. doi:[10.1021/jo400864s](https://doi.org/10.1021/jo400864s)
121. Higashida R, Oka N, Kawanaka T, Wada T (2009) Nucleoside H-boranophosphonates: a new class of boron-containing nucleotide analogues. *Chem Commun*. doi:[10.1039/b901045a](https://doi.org/10.1039/b901045a)
122. Oka N, Takayama Y, Ando K, Wada T (2012) Synthesis of nucleoside 50-boranophosphorothioate derivatives using an H-boranophosphonate monoester as a precursor. *Bioorg Med Chem Lett* 22:4571–4574. doi:[10.1016/j.bmcl.2012.05.093](https://doi.org/10.1016/j.bmcl.2012.05.093)
123. Baranowski MR, Nowicka A, Rydzik AM et al (2015) Synthesis of fluorophosphate nucleotide analogues and their characterization as tools for 19F NMR studies. *J Org Chem* 80:3982–3997. doi:[10.1021/acs.joc.5b00337](https://doi.org/10.1021/acs.joc.5b00337)
124. Bollmark M, Stawinski J (1998) Nucleotide analogues containing the P–F bond. An overview of the synthetic methods. *Nucleosides Nucleotides Nucl Acids* 17:663–680. doi:[10.1080/07328319808005208](https://doi.org/10.1080/07328319808005208)
125. Dąbkowski W, Tworowska I, Michalski J, Cramer F (2000) New efficient synthesis of thymidine cyclic 3', 5'-phosphorofluoridate and its sulfur analogue via the phosphoroamidite route. *Nucleosides Nucleotides Nucl Acids* 19:1779–1785. doi:[10.1016/0040-4039\(94\)02403-X](https://doi.org/10.1016/0040-4039(94)02403-X)
126. Misiura K, Szymanowicz D, Kuśnierczyk H (2001) Synthesis, chemical and enzymatic reactivity, and toxicity of dithymidylyl-3',5'-phosphorofluoridate and -phosphorothiofluoridate. *Bioorg Med Chem* 9:1525–1532
127. Dąbkowski W, Tworowska I (2001) Novel phosphitylating reagents containing a phosphorus–fluorine bond and their application in efficient synthesis of phosphorofluoridates and phosphorofluoridothionates. *J Chem Soc Perkin Trans 1*:2462–2469. doi:[10.1039/b103082p](https://doi.org/10.1039/b103082p)
128. Murai T, Tonomura Y, Takenaka T (2011) Phosphorofluoric acid ammonium salts and acids: synthesis, NMR properties, and application as acid catalysts. *Heteroatom Chem* 22:417–425. doi:[10.1002/hc.20700](https://doi.org/10.1002/hc.20700)

129. Wächter M, Rüedi P (2012) Synthesis and characterization of enantiomerically pure cis- and trans-3-fluoro-2,4-dioxo-7-aza-3-phosphadecalin 3-oxides as acetylcholine mimetics and inhibitors of acetylcholinesterase. *Helv Chim Acta* 95:716–736
130. Rovnaník P, Žák Z, Černík M (2006) Syntheses of phosphoryl chloro- and bromofluorides and crystal structures of POFCl₂ and POF₂Cl. *Z Anorg Allg Chem* 632:1356–1362. doi:10.1002/zaac.200500510
131. Aldersley MF, Joshi PC, Schwartz HM, Kirby AJ (2014) The reaction of activated RNA species with aqueous fluoride ion: a convenient synthesis of nucleotide 5'-phosphorofluoridates and a note on the mechanism. *Tetrahedron Lett* 55:1464–1466. doi:10.1016/j.tetlet.2014.01.051
132. Aldersley MF, Joshi PC, Ott EL et al (2015) The introduction of P–F bonds using aqueous fluoride ion and a water soluble carbodiimide: a convenient alternative synthesis of phosphorofluoridates and phosphonofluoridates. *Tetrahedron Lett* 56:5272–5274. doi:10.1016/j.tetlet.2015.07.036
133. Cravatt BF, Wright AT, Kozarich JW (2008) Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu Rev Biochem* 77:383–414. doi:10.1146/annurev.biochem.75.101304.124125
134. Singh RP, Jean'ne MS (2002) Recent advances in nucleophilic fluorination reactions of organic compounds using deoxofluor and DAST. *Synthesis* 17:2561–2578
135. Saltmarsh JR, Boyd AE, Rodríguez OP et al (2000) Synthesis of fluorescent probes directed to the active site gorge of acetylcholinesterase. *Bioorg Med Chem Lett* 10:1523–1526
136. Guo L, Suarez AI, Braden MR et al (2010) Inhibition of acetylcholinesterase by chromophore-linked fluorophosphonates. *Bioorg Med Chem Lett* 20:1194–1197. doi:10.1016/j.bmcl.2009.12.007
137. Pallan PS, Egli M (2007) Selenium modification of nucleic acids: preparation of oligonucleotides with incorporated 2'-SeMe-uridine for crystallographic phasing of nucleic acid structures. *Nat Protoc* 2:647–651. doi:10.1038/nprot.2007.75
138. Salon J, Sheng J, Jiang J et al (2007) Oxygen replacement with selenium at the thymidine 4-position for the Se base pairing and crystal structure studies. *J Am Chem Soc* 129:4862–4863. doi:10.1021/ja0680919
139. Sheng J, Huang Z (2010) Selenium derivatization of nucleic acids for X-ray crystal-structure and function studies. *Chem Biodivers* 7:753–785. doi:10.1002/cbdv.200900200
140. Han Q, Sarafianos SG, Arnold E et al (2009) Synthesis of boranoate, selenoate, and thioate analogs of AZTp₄A and Ap₄A. *Tetrahedron* 65:7915–7920. doi:10.1016/j.tet.2009.07.079
141. Lin L, Caton-Williams J, Kaur M et al (2011) Facile synthesis of nucleoside 5'-(alpha-P-seleno)-triphosphates and phosphoroselenoate RNA transcription. *RNA* 17:1932–1938. doi:10.1261/rna.2719311
142. Qi N, Jung K, Wang M et al (2011) A novel membrane-permeant cADPR antagonist modified in the pyrophosphate bridge. *Chem Commun* 47:9462–9464. doi:10.1039/c1cc13062e
143. Strenkowska M, Wanat P, Ziemniak M et al (2012) Preparation of synthetically challenging nucleotides using cyanoethyl P-imidazolides and microwaves. *Org Lett* 14:4782–4785. doi:10.1021/ol302071f
144. Kowalska J, Lukaszewicz M, Zuberek J et al (2009) Phosphoroselenoate dinucleotides for modification of mRNA 5' end. *ChemBioChem* 10:2469–2473. doi:10.1002/cbic.200900522
145. Müller AC, Giambro R, Weißer J, Májek P (2016) Identifying kinase substrates via a heavy ATP kinase assay and quantitative mass spectrometry. *Sci Rep* 6:1–10
146. Li Y, Cross FR, Chait BT (2014) Method for identifying phosphorylated substrates of specific cyclin/cyclin-dependent kinase complexes. *Proc Natl Acad Sci USA* 111:11323–11328. doi:10.1073/pnas.1409666111
147. Xue L, Wang P, Cao P et al (2014) Identification of extracellular signal-regulated kinase 1 (ERK1) direct substrates using stable isotope labeled kinase assay-linked phosphoproteomics. *Mol Cell Proteomics* 13:3199–3210. doi:10.1074/mcp.O114.038588
148. Scian M, Acchione M, Li M, Atkins WM (2014) Reaction dynamics of ATP hydrolysis catalyzed by P-glycoprotein. *Biochemistry* 53:991–1000. doi:10.1021/bi401280v
149. Melby ES, Soldat DJ, Barak P (2011) Synthesis and detection of oxygen-18 labeled phosphate. *PLoS ONE* 6:e18420. doi:10.1371/journal.pone.0018420.001
150. Kübler D, Schäfer M, Lehmann W-D, Seidel J (2011) Manufacture and usage of (γ -¹⁸O₃)ATP or [γ -¹⁸O₃]GTP. *WO* 2011/064289 A1, 1–23
151. Fu C, Zheng X, Jiang Y et al (2013) A universal and multiplex kinase assay using γ -[¹⁸O₄]-ATP. *Chem Commun* 49:2795–2797. doi:10.1039/c3cc38467e

Phosphate-Modified Nucleotides for Monitoring Enzyme Activity

Susanne Ermert¹ · Andreas Marx¹ · Stephan M. Hacker¹

Received: 2 November 2016 / Accepted: 30 January 2017 / Published online: 1 March 2017
© Springer International Publishing Switzerland 2017

Abstract Nucleotides modified at the terminal phosphate position have been proven to be interesting entities to study the activity of a variety of different protein classes. In this chapter, we present various types of modifications that were attached as reporter molecules to the phosphate chain of nucleotides and briefly describe the chemical reactions that are frequently used to synthesize them. Furthermore, we discuss a variety of applications of these molecules. Kinase activity, for instance, was studied by transfer of a phosphate modified with a reporter group to the target proteins. This allows not only studying the activity of kinases, but also identifying their target proteins. Moreover, kinases can also be directly labeled with a reporter at a conserved lysine using acyl-phosphate probes. Another important application for phosphate-modified nucleotides is the study of RNA and DNA polymerases. In this context, single-molecule sequencing is made possible using detection in zero-mode waveguides, nanopores or by a Förster resonance energy transfer (FRET)-based mechanism between the polymerase and a fluorophore-labeled nucleotide. Additionally, fluorogenic nucleotides that utilize an intramolecular interaction between a fluorophore and the nucleobase or an intramolecular FRET effect have been successfully developed to study a variety of different enzymes. Finally, also some novel techniques applying electron paramagnetic resonance (EPR)-based detection of nucleotide cleavage or the detection of the cleavage of fluorophosphates are discussed. Taken together, nucleotides modified at the terminal phosphate

This article is part of the Topical Collection “Phosphate Labeling and Sensing in Chemical Biology”; edited by Henning Jessen.

✉ Stephan M. Hacker
stephan.hacker@uni-konstanz.de

¹ Department of Chemistry and Konstanz Research School Chemical Biology, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany

position have been applied to study the activity of a large diversity of proteins and are valuable tools to enhance the knowledge of biological systems.

Keywords Modified nucleotides · Activity assays · Kinases · DNA sequencing · Fluorogenic substrates · FRET

1 Introduction

Nucleotide-dependent processes are fundamental to many biological pathways. ATP is the universal energy currency in every living cell and is utilized to couple endergonic and exergonic processes [1]. Nucleoside triphosphates (NTPs) are the building blocks for the synthesis of DNA and RNA and are, therefore, pivotal to information storage and usage in living systems [2]. Furthermore, nucleotide-based small molecules are messengers that are important for the adjustment to various conditions [3–6]. Studying the activity of enzymes involved in the turnover of nucleotides is, therefore, an important strategy to gain insights into biological systems.

Over the years, it has been shown that attachment of a reporter molecule to nucleotides is a useful strategy to obtain probes that are still efficiently turned over by the enzyme of interest, but show additional features that allow investigating this reaction. For this purpose, e.g., fluorophores [7, 8], affinity handles [9] or photocrosslinkers [10] are attached to nucleotides. Whereas a variety of nucleotides modified at the nucleoside can be used for different applications and especially to detect nucleotide binding [11], nucleotides terminally modified at the phosphate chain hold special promise as the modification will be separated from the nucleotide during the enzymatic reaction. This allows, for example, in the case of kinases, the transfer of a reporter to a target protein [12] or the enzyme itself [9]. In the case of nucleic acid polymerases, phosphate-modified nucleotides allow the detection of enzymatic activity, while synthesizing the natural polymer. This is especially advantageous if many nucleotides need to be incorporated in a subsequent fashion [13]. Furthermore, the fact that the phosphate modification is separated from the nucleotide after cleavage is used to construct fluorogenic probes for enzymatic activity [8, 14]. All these possible applications render nucleotides modified at the phosphate chain interesting entities to study enzymatic activity.

2 Tailoring Phosphate-Modified Nucleotides for Specific Applications

For the terminal attachment of a modification to the phosphate chain of nucleotides a variety of different connection modes can be used. This attachment, in many cases, needs to be optimized for the specific enzyme of interest. Factors that contribute to the selection of the modification include (1) tolerance of the enzyme towards the modification, (2) stability of the modification under the assay conditions and (3) synthetic accessibility of the specific molecule. In this section, we will present a selection of nucleotides modified with a reporter at the terminal phosphate and review the respective synthesis routes as well as their stability. A more

comprehensive report of modern synthesis strategies towards these challenging molecules, including their modular synthesis using P(III) chemistry [15], is given by H. Jessen in this book.

Many terminally phosphate-modified nucleotides have a length of the phosphate chain of three phosphoanhydride units as found in the most common NTPs. In the easiest case, one of the atoms of the γ -phosphate can be exchanged with another atom (Fig. 1). This has, e.g., been done for the exchange of one oxygen to sulfur (ATP γ S, **1**) [16] or fluorine (ATP γ F, **2**) [17]. However, many systems require the installation of larger reporter molecules. These modifications are obtained with various chemistries on the terminal phosphate (Fig. 2). Molecules with nitrogen (**3**) [10, 18–23], oxygen (**4**) [20, 22, 23], sulfur (**5**) [24] or carbon (**6**) [22, 25] directly attached to the phosphor atom have been reported. For synthesizing modified nucleotides based on strong nucleophiles like amines for nitrogen-linked nucleotides [21] or phenols for oxygen-linked variants [26, 27], the unmodified NTP can be directly activated with, e.g., a carbodiimide-based reagent to give the cyclic trimetaphosphate [28]. This intermediate is then attacked by the nucleophile to give the modified nucleotide. In cases in which the nucleophile cannot be used to do this (aliphatic alcohols for oxygen-linked nucleotides [29] or in general for carbon-linked nucleotides [22, 25]), a strategy was used that is based on the coupling of a modified monophosphate building block (*O*-alkyl phosphate [29] or alkyl phosphonate [22, 25]) with the nucleoside diphosphate. In the case of oxygen-linked [22, 30] and sulfur-linked molecules [24], the direct alkylation of the γ -phosphate of the NTP or NTP γ S was also employed. A special class of oxygen-linked nucleotides are acyl-phosphates (**7**) [9]. These contain an anhydride between the terminal phosphate and a carboxylic acid. These molecules were synthesized by activation of a carboxylic acid and subsequent coupling to the NTP.

For several applications, especially for DNA and RNA polymerases [26], but also for the ubiquitin-activating enzyme UBA1 [31], it has been shown that longer phosphate chains are beneficial for the acceptance of a modification at the terminal phosphate. Therefore, oxygen- [22], nitrogen- [32] and carbon-linked [22] tetraphosphates have been synthesized. For the oxygen- and carbon-linked molecules [22], the NTP is activated as mentioned above to give the trimetaphosphate and coupled to an *O*-alkyl phosphate or alkyl phosphonate. For the nitrogen-linked analogs [32], a nucleoside tetraphosphate is activated with a carbodiimide and coupled to an amine. In the cases of DNA polymerases, even longer phosphate chains, i.e. penta- or hexaphosphates [13, 32, 33], have proven superior with regard to enzymatic turnover. These molecules are synthesized by coupling a di- or triphosphate containing the modification with the NTP.

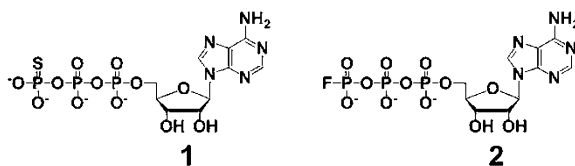


Fig. 1 Structures of ATP γ S (**1**) [16] and ATP γ F (**2**) [17]

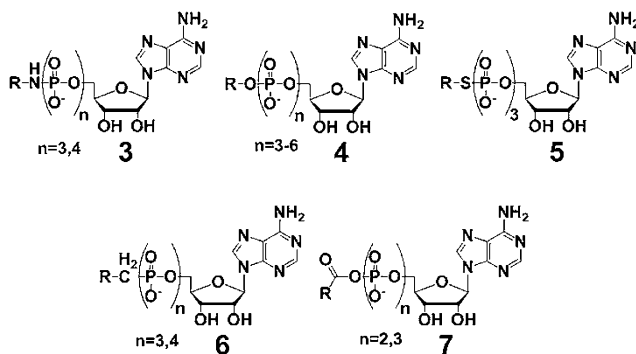


Fig. 2 Structures of adenosine nucleotide analogs with different types of modifications at the terminal phosphate

Besides the acceptance by the enzyme of interest, the chemical stability of the molecules towards the assay conditions is an important prerequisite. To address the question of the stability of various modes of nucleotide modification, nitrogen-, oxygen- and carbon-linked NTPs were investigated in regards to their stability under different pH conditions [22]. Whereas nitrogen-linked variants are hydrolyzed at the P–N bond in solutions with a pH value below 6, oxygen- and carbon-linked molecules are stable over a broad pH range from 2 to 12. The same stability is found for oxygen- and carbon-linked tetraphosphates. This shows that for nitrogen-linked analogs, the assay conditions have to be carefully chosen in order to exclude chemical hydrolysis, whereas oxygen- and carbon-linked nucleotides are stable over a wide range of conditions.

3 Kinase-dependent Labeling of Substrates

3.1 Transfer of Thiophosphate to Substrates

Besides radioactive variants of the γ -phosphate, the smallest modification that can be made to the γ -phosphate of a nucleotide is the exchange of an oxygen of the phosphate to another atom like sulfur (ATP γ S, **1**). In the case of kinases, ATP γ S was used to study enzymatic activity [34]. In this assay (Fig. 3), the thiophosphate is

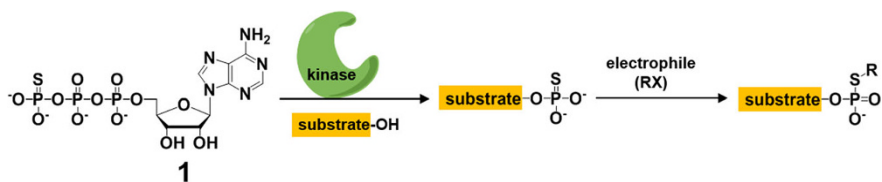


Fig. 3 Schematic depiction of an activity assay for kinases based on ATP γ S (**1**) [34]. The kinase transfers the thiophosphate from **1** to its substrate. The sulfur of this moiety is reacted with an electrophile containing a label that is used for detection

transferred to a substrate protein. Due to the exceptionally high nucleophilicity of the sulfur atom, the transferred thiophosphate can be reacted with an electrophile that bears a reporter molecule. One obvious challenge of the approach is that the natural amino acid cysteine also contains a free thiol group that is prone to react with electrophiles. It has been shown that at low pH values the thiophosphate is labeled specifically over cysteine making it possible to detect this modification [34]. Additionally, other strategies have been developed to discriminate cysteine labeling from thiophosphate labeling. In one approach, cysteines and thiophosphates are labeled with iodoacetamide biotin and both are affinity-enriched [35]. Under oxidative conditions, the thiophosphate modification is cleaved due to oxidation of the sulfur, whereas the modification on the oxidized cysteine remains stable. This is used to specifically elute thiophosphorylated peptides. In another approach, a *p*-nitrobenzyl group is transferred to cysteines and thiophosphates and the thiophosphate modification is enriched with a specific antibody [16, 36]. Using the “bump-and-hole” strategy pioneered by Shokat [37, 38], the targets of a specific kinase in the background of all other kinases in the proteome can be identified [16, 35].

3.2 Transfer of Biotin to Substrates

A versatile reporter that can be used to detect peptide modification by kinases is biotin. Thus, an ATP analog bearing biotin at the γ -phosphate through a nitrogen linkage has been developed (Fig. 4, 8) [12]. This substrate is utilized by kinases in place of ATP and results in the transfer of biotin to the target peptide [12, 23]. It has recently been shown that at least 25 kinases in the human proteome accept this molecule as substrate giving this approach some generality [40]. The modification of the peptide with this moiety is relatively resistant to phosphatase cleavage, but can be cleaved off chemically using acidic conditions [41]. The biotinylated peptides were detected in various ways. They were directly visualized after gel electrophoresis using biotin-directed antibodies or streptavidin conjugates [12, 39–41]. Furthermore, functionalized nanoparticles were used to detect the modified peptides using electrochemical [42], colorimetric [43] or resonance light scattering techniques [44]. Recently, it has been shown that biotinylated ATP analogs (9) can be designed in a way that they become cell-permeable [39]. This

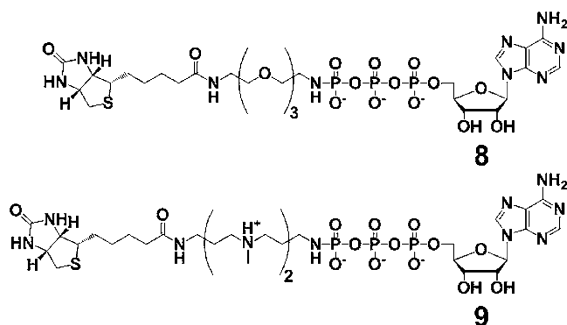


Fig. 4 Structures of biotin-labeled ATP analogs that are used for detecting kinase activity *in vitro* (8) [12] and *in cellulo* (9) [39]

allows the performance of whole cell-based experiments and, therefore, a way to monitor phosphorylation in the cellular context. Besides protein kinases also the T4 polynucleotide kinase was studied using biotinylated ATP as a substrate [45].

3.3 Transfer of Ferrocene to Substrates

Another reporter group that was enzymatically transferred to target peptides is the ferrocene moiety (Fig. 5). Due to the specific electrochemical properties of the iron in ferrocene, it was used to detect peptide modification in an electrochemical fashion. In this assay, the substrate peptide is immobilized on an electrode surface that was employed for voltammetry [18, 46]. After kinase-catalyzed transfer of the ferrocene of nucleotide **10** to the substrate, its oxidation and reduction can be followed using electrochemical methods and used to quantify the amount of ferrocene that was transferred. By optimizing the structure of the ferrocene-modified ATP analog, high efficiency for the utilization of this molecule by kinases could be achieved [19, 47]. It has been shown that the ferrocene moiety is also applicable to perform immunodetection of the labeling event using ferrocene-specific antibodies [48]. Furthermore, an ATP analog containing a ferrocene and an alkyne moiety for subsequent click chemistry (**11**) has been reported to further broaden the applicability of this type of modification [49].

3.4 Transfer of Photoreactive Groups to Substrates

Several ATP analogs were developed in order to identify their binding partners by photoaffinity labeling. For this purpose, an ATP analog containing an aromatic azide (**12**) has been reported (Fig. 6) [10]. The modified phosphate is transferred to the substrate and introduces a crosslink between the kinase and its substrate upon irradiation. The substrate will be labeled covalently due to the transfer of the modified phosphate, whereas the kinase is attached to the formerly photoreactive group. In this way, pairs of kinases and respective substrate proteins can be identified. In another concept, ATP is modified with two crosslinkers (**13**) [50], one at the nucleoside and one at the phosphate chain. This analog is designed in a way that it binds to the kinase in complex with the protein substrate, but is resistant to

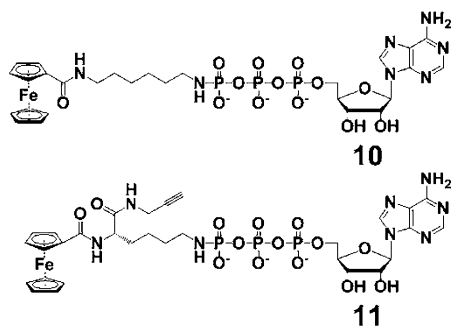


Fig. 5 Structures of ferrocene-labeled ATP analogs

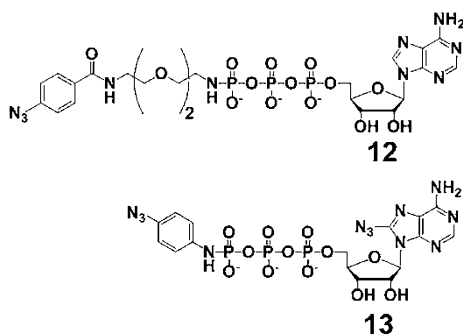


Fig. 6 Structures of ATP analogs used to photocrosslink kinases and their substrates

phosphate transfer. Upon irradiation, the nucleoside part will react with the kinase, whereas the photoreactive group on the phosphate chain will react with the substrate and both will be crosslinked.

Another application of photoreactive groups along these lines is to enzymatically label the substrate and use this new photoreactive entity to look for its interaction partners [51]. This has been exemplarily shown for T4 polynucleotide kinase and DNA [51]. The DNA substrate was rendered photoreactive by transferring a phosphate modified with an aromatic azide to its 5'-end. This DNA is added to a sample containing a potential interactor and if the two actually interact, they are crosslinked upon irradiation. The DNA binding of the proteins RPA and FEN-1 has been studied using this approach [51].

3.5 Transfer of Fluorophores to Substrates

Several ATP analogs with different dyes attached to the γ -phosphate were reported and used to label kinase substrates (e.g. analogs **14**–**16**, Fig. 7). Even the substrates of histidine kinases, which are usually difficult to label because of the instability of the phosphate histidine adduct, could be modified using a sulfur-linked dye-modified ATP analog (**16**) [53]. One main application of this technology is the design of FRET-based sensors to detect kinase activity (Fig. 7a). In these assays, a substrate peptide of a kinase of interest is labeled with a small molecule fluorophore [52] or a fluorescent quantum dot [54]. This substrate is incubated with the kinase and an ATP analog bearing another fluorophore at the γ -phosphate. After transfer of this second fluorophore to the substrate, the two dyes are able to undergo FRET, resulting in a change in the fluorescent signal. This signal change is used to detect the activity of the kinase. In a related application, biotinylated ATP was used for the same procedure [55]. In this case, after transfer of the biotin moiety to the substrate, an antibody conjugated to a second fluorophore binds to biotin and enables FRET.

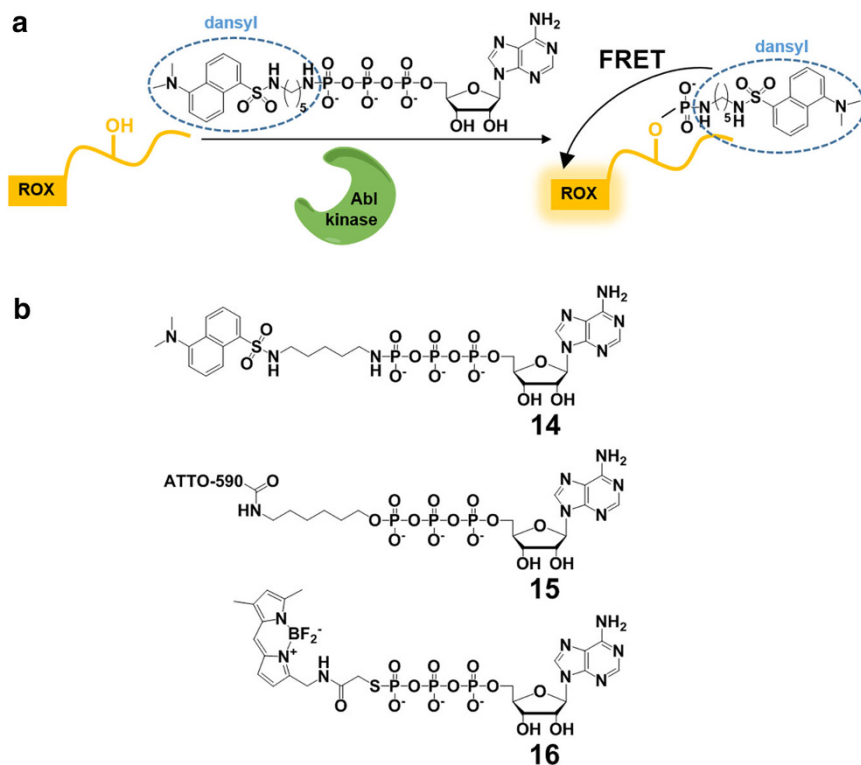


Fig. 7 **a** Concept for the detection of kinase activity using a FRET-based technique [52]. The kinase substrate is labeled with a fluorophore [e.g. 5(6)-carboxy-X-rhodamine, ROX] and the kinase is incubated with a nucleotide labeled with a second fluorophore (e.g. dansyl). Upon transfer of the modified phosphate, FRET between the two dyes occurs, allowing the detection of enzymatic activity; **b** structures of representative dye-labeled nucleotides

4 Detection of Nucleotide-Binding Sites with Acyl-Phosphates

Acyl-phosphates are special among terminally modified nucleotides as the modification in this case is attached via an anhydride between a phosphate and a carboxylic acid making it especially reactive [9]. In this way, the modification was used to covalently label lysine residues in proteins that interact with the nucleotide and thereby identify and profile nucleotide-binding pockets (Fig. 8a). Although this is not strictly speaking a tool to monitor the activity of the protein, it is a method that provides important information on nucleotide-binding proteins using the terminally labeled nucleotides.

The utilization of acyl-phosphates usually relies on the attachment of a biotin or desthiobiotin modification for enrichment (Fig. 8b, 17) [9]. Upon treatment of the sample with the acyl-phosphate probes, proteins that interact with the nucleotides will be labeled with these affinity handles, which can then be used to detect or enrich the proteins that initially reacted with the nucleotide. This allows the proteome-wide identification of modified proteins using mass spectrometry-based techniques. One

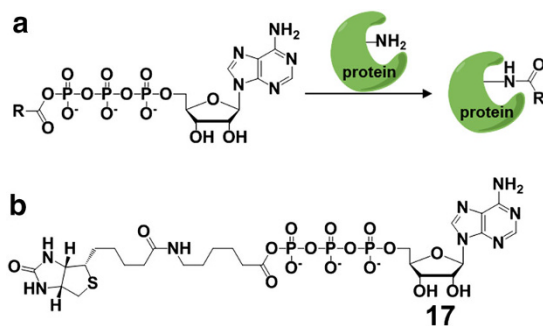


Fig. 8 **a** Concept for the labeling of nucleotide-binding sites with acyl-phosphates [9]. Upon binding of a nucleotide to the protein, the acyl-phosphate is able to react with a nearby lysine, resulting in the covalent labeling of the protein; **b** structure of a representative acyl-phosphate probe containing biotin as an affinity handle

main advantage of this technology is that the modification is transferred to result in a stable amide on the protein. In this way, after tryptic digest, the peptide that is directly modified with the affinity handle is identified [9]. Therefore, the method also allows the identification of the exact binding site on the protein. Acyl-phosphate probes have been designed to study the binding of adenosine [9, 56, 57] and guanosine nucleotides [56], but also for more specific nucleotides like 6-thioguanosine triphosphate [58]. Profiling of the nucleotide interactome has been performed for human [57], but also for plant [59] and mycobacterial proteomes [60].

One challenge of the method is that the acyl-phosphate probes are quite reactive and will also label some protein pockets that are not bona fide high-occupancy binding pockets for nucleotides. To address this, different protocols have been established. One option is to identify the sites for which the natural substrate competes labeling with the acyl-phosphate probe [61, 62]. Only pockets that quantitatively interact with the natural substrate should, in this case, be blocked from reactivity with the probe. In this way, specific interactions are identified. Another possibility is to perform the experiment at two different concentrations of the probe [63]. At high concentration, the probe will interact with many different pockets in the proteome. In contrast, at low concentration, only the pockets that have a high binding affinity to the nucleotide will be quantitatively labeled by the probe. The relative labeling between the two reactions is usually read out using isotopically coded probes in order to be able to quantify the two labeling reactions in the same mass spectrometric experiment [63].

Besides the ability to identify nucleotide-binding pockets, this methodology is also used to assess target engagement of small molecules and perform an off-target analysis. Most kinases possess at least one of two conserved lysines in the active site and this protein family is, therefore, amenable to be broadly studied using this technology [64]. In a pioneering study, acyl-phosphates have been used to profile the kinome-wide reactivity of staurosporine [9]. This study identified the cellular targets of staurosporine and also identified the concentration dependence of this binding event for many kinases in parallel. Additionally, this method has also been used to study the on- and off-target effects of a variety of drug candidates towards ATP binding sites [64–67].

5 Studying Nucleic Acid Polymerase Activity and Next-Generation Sequencing

5.1 Nucleotide Design for Polymerase Probes

Modified nucleotides, e.g. carrying a fluorescent dye at the terminal phosphate, are valuable tools to study the enzymatic activity of nucleic acid polymerases and are used to sequence DNA and RNA. For sequencing applications, many nucleotides have to be incorporated into the growing nucleic acid strand in a subsequent fashion. Therefore, their incorporation must be possible with high efficiency. Studies of DNA polymerases and their action on phosphate-modified triphosphates revealed that the catalytic activity of the enzymes is impaired by the modifications [26]. This made it impossible to synthesize long stretches of nucleic acids solely from modified building blocks. Subsequently, it could be shown that elongating the phosphate chain results in nucleotide analogs that are significantly better accepted by polymerases. While tetraphosphates are already better substrates [26], penta- and hexaphosphates seem to be ideal substrates for polymerases, enabling the consecutive incorporation of many building blocks into the growing strand exclusively from modified building blocks [13, 26].

5.2 Fluorogenic and Chromogenic Substrates for Polymerases

Fluorogenic and chromogenic substrates for nucleic acid polymerases typically rely on dyes whose absorbance or fluorescence characteristics require a free phenol group. Such a behavior is found in certain fluoresceins [7, 68] and coumarins [27] as well as nitrophenols [69]. This property was harnessed (Fig. 9) by attaching the dye through the phenolic hydroxyl to the terminal phosphate of the nucleotide, lowering its absorbance or fluorescence intensity. Upon incorporation of the nucleoside monophosphate into DNA the dye with a phosphate chain is released. In contrast to the intact, terminally modified nucleotide, this molecule is a substrate for phosphatases that will dephosphorylate the dye to the free phenol and result in an optical signal. This assay has been used on a chromogenic basis using *p*-nitrophenyl-substituted nucleotides (18) to monitor RNA polymerase activity [69]. Using the fluorogenic properties of coumarins (19, 20) it has also been applied to study various DNA [27, 70] and RNA polymerases [71].

More recent studies show that this concept can be explored to sequence DNA using fluorescein-labeled deoxynucleoside tetraphosphates (e.g. 21, Fig. 10) [7, 68]. For this purpose, the DNA that should be sequenced and a suitable primer are immobilized in wells of microreactors, allowing sequencing of many different DNA populations in parallel. The microreactor is then filled with one of the four dye-labeled nucleotides (21), the polymerase and a phosphatase. If the nucleotide is the canonical one, it will be incorporated and after phosphatase cleavage, the active fluorophore will be released. In this way, all wells that incorporated the nucleotide will light up and, therefore, indicate the identity of the base in the sequence. If there are several incorporation events for this nucleotide, as there is a homopolymeric

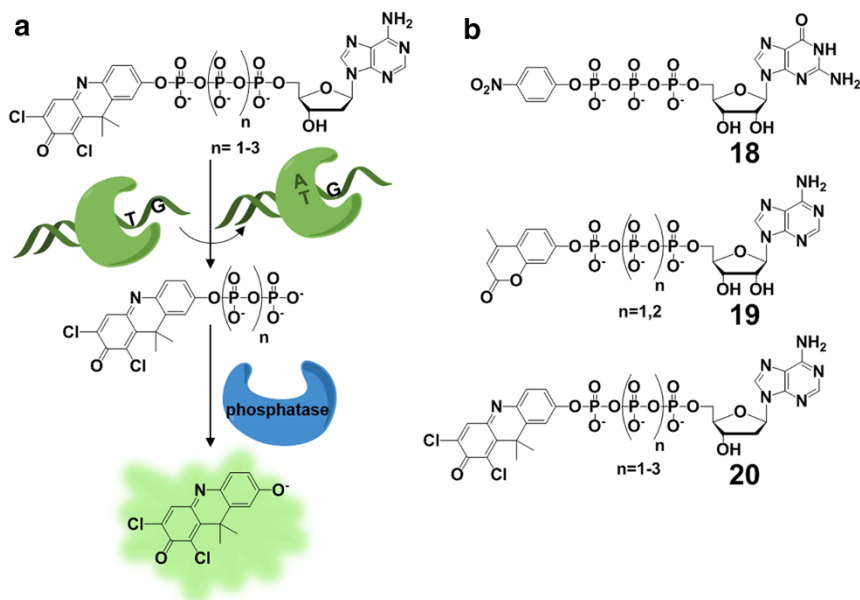


Fig. 9 **a** Concept of chromogenic and fluorogenic substrates for DNA and RNA polymerases [7]. Upon incorporation of the nucleotide into the DNA, the dye with an attached phosphate chain is released. This molecule is dephosphorylated by a phosphatase, resulting in a free phenolic hydroxyl group and an optical signal; **b** structures of representative chromogenic and fluorogenic polymerase substrates

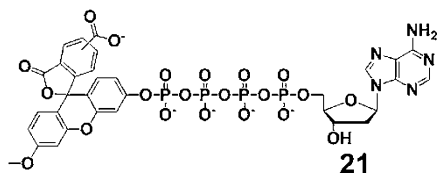


Fig. 10 Structure of a fluorogenic DNA polymerase substrate that can be used for sequencing

stretch in the DNA, the number of incorporation events can be read out by the intensity of the fluorescence signal. The reaction mixture is washed out and the procedure is repeated for the next nucleotide. Cycling through all nucleotides allows consecutive sequencing of the DNA strand [7, 68].

5.3 Fluorescently Labeled Nucleotides in Zero-mode Waveguides

One of the major challenges in next-generation sequencing of DNA is the ability to sequence a single DNA molecule. Nucleotides modified at the terminal phosphate were successfully applied to address this challenge. The most advanced technology in this context is commercialized through Pacific Bioscience (Fig. 11) [33, 72]. In this concept, a single DNA polymerase is immobilized in one well of a zero-mode waveguide [73]. It incorporates nucleotides that are modified with fluorophores at the terminal phosphate into the DNA strand. Due to the properties of the waveguide,

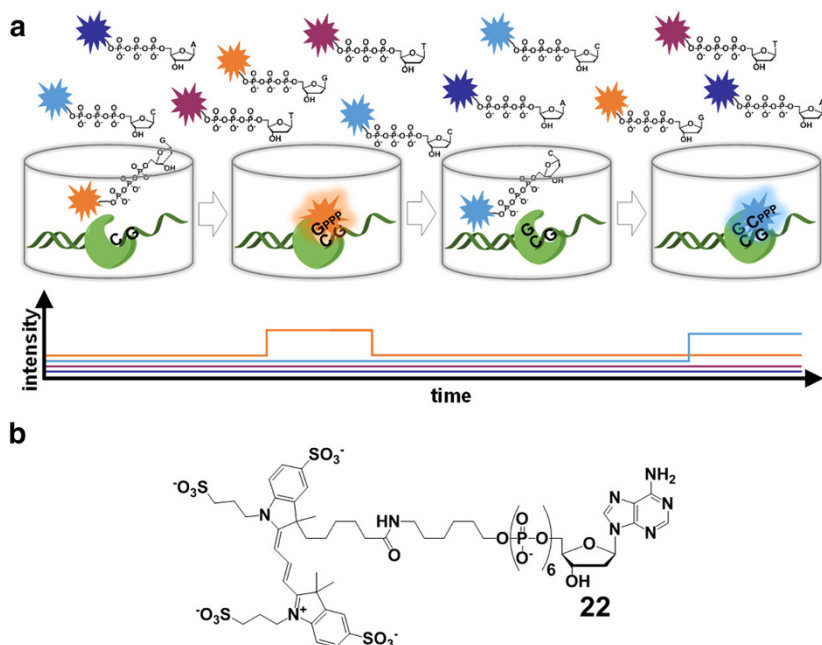


Fig. 11 **a** Concept for sequencing of DNA in zero-mode waveguides [33]. The DNA polymerase is immobilized on the bottom of a well of the waveguide and the DNA primer complex is bound. Only when a fluorescent nucleotide is incorporated into the DNA, it stays close to the polymerase for a substantial amount of time and its fluorescence signal is detected. After the incorporation, the fluorophore is cleaved off, terminating the signal; **b** structure of one representative example of a nucleoside hexaphosphate used in this approach [33]

only fluorescent molecules that are currently in close proximity of the polymerase are excited and give a fluorescent signal that can be detected over the high background concentration of all four nucleotides in solution. As only the nucleotide that is successfully incorporated into the growing DNA strand will stay close to the polymerase for a significant time, a longer fluorescent signal is directly assigned to an incorporation event. All four nucleotides are labeled with different fluorophores. The color of the fluorescence signal is used to sequence the DNA. In order to achieve high incorporation efficiencies, modified hexaphosphates of the nucleotides (e.g. **22**) are used, allowing the sequencing of more than 10,000 nucleotides in each well. In order to obtain even higher accuracy from a single molecule of DNA, the sequencing reaction is usually performed on circular DNA prepared from the initial linear DNA by the SMRTbell concept [74]. Using this approach, the sequence of the DNA is read multiple times and a consensus sequence is derived with very low error rates. Although generation of SMRTbell libraries is beneficial for sequencing accuracy, it has been shown that sequencing bacterial genomes can be performed in zero-mode waveguides without former library preparation [75].

One important application of single-molecule sequencing is the ability to generate data not only on the sequence of the DNA, but also on DNA base modifications like 5-methylcytosine, 6-methyladenosine and 5-hydroxymethylcytosine. 5-methylcytosine

occurrence can be analyzed with the described technology using bisulfite sequencing and subsequently comparing the native sequencing read with the bisulfite sequencing read, in which cytosine will be read as thymidine and 5-methylcytosine will be read as cytosine, allowing their discrimination [76]. This method, therefore, gives information on the position of 5-methylcytosine in the DNA. Nevertheless, it requires the chemical step of bisulfite treatment and subsequent DNA amplification. It has also been shown that DNA methylation can be directly detected during DNA sequencing in zero-mode waveguides [77–79]. This is based on the fact that after incorporation of a nucleotide opposite a methylated DNA base, the DNA polymerase will have a longer delay to incorporate the next nucleotide. Therefore, the kinetics of the incorporation are utilized to discriminate methylated and non-methylated cytosine and adenosine. This methodology can also be applied to 5-hydroxymethylcytosine [80]. Here, enzymatic tagging is used where β -glucosyltransferase transfers an azido-modified sugar to the hydroxyl group of 5-hydroxymethylcytosine. This sugar is modified with an affinity handle using copper-catalyzed azide alkyne cycloaddition (CuAAC) and the sequences containing 5-hydroxymethylcytosine are specifically enriched. Furthermore, in the sequencing using the described technology, this modified base will give a longer delay between incorporations, allowing the identification of the exact modification site.

Besides sequencing of DNA, also the sequencing of mRNA and, therefore, the transcriptome is of high interest. The described technology can also be applied to perform this task. On the one hand, mRNA can be reverse-transcribed into DNA and analyzed using the standard sequencing workflow [81, 82]. On the other hand, RNA can also be sequenced directly starting from single RNA molecules [83, 84]. This is possible by the immobilization of a reverse transcriptase instead of a DNA polymerase in the zero-mode waveguide. In this way not only the sequence of mRNA is detected, but also base modifications, like 6-methyladenosine, are identified on the single-molecule level for RNA.

5.4 FRET-based Sequencing of DNA

One main advantage of the technology mentioned above is that the zero-mode waveguide allows the detection of a single fluorophore-modified nucleotide over the background of a high concentration of fluorophores in the bulk solution. Another way to do this is to utilize the distance dependence of the FRET effect (Fig. 12) [85, 86]. In this method, the primer-DNA complex is immobilized on a glass slide. The DNA polymerase is modified with a quantum dot that serves as fluorescence donor for all four fluorophores that are attached to the different nucleotides. Only in the progress of incorporation, FRET between the quantum dot and the respective fluorophore occurs for a sustained time and is read out by fluorescence detection. After the incorporation event the fluorophore is released and the signal terminated. Each nucleotide is labeled with a specific dye that signals the exact sequence of the DNA strand.

5.5 Phosphate-Labeled Nucleotides in Nanopores

Another concept that successfully allows single-molecule sequencing using terminally phosphate-modified nucleotides is based on the usage of nanopores and

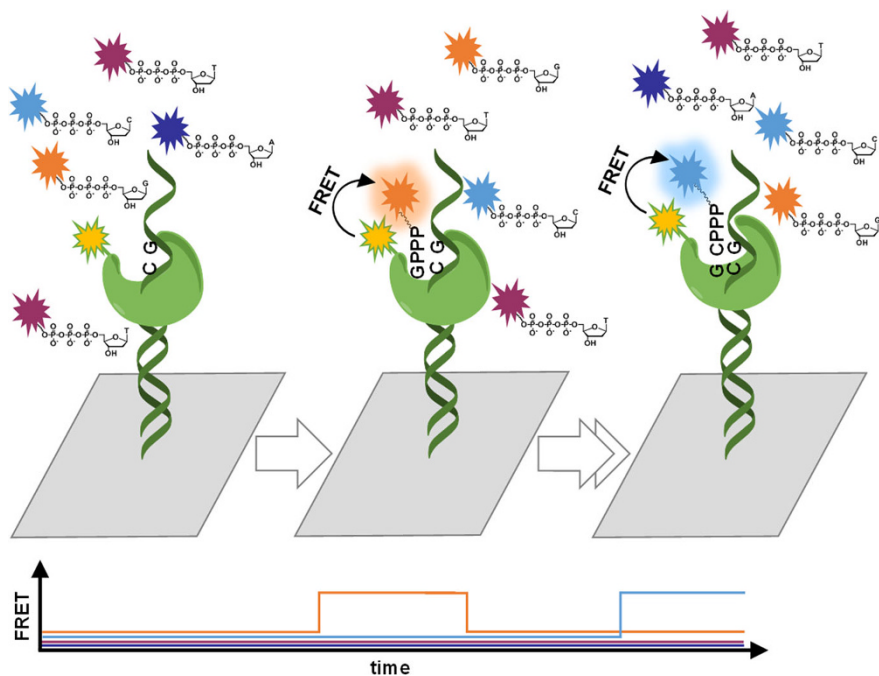


Fig. 12 Concept for FRET-based sequencing of DNA [85, 86]. The polymerase is labeled with a quantum dot and bound to a DNA primer complex immobilized on a glass slide. Upon incorporation of a fluorescently labeled nucleotide, the quantum dot serves as fluorescence donor allowing detection of the incorporation event

measuring the flow of electric current through these pores (Fig. 13) [32, 87]. For this purpose, a DNA polymerase is immobilized on one side of an α -hemolysin pore. This DNA polymerase is bound to a DNA primer complex and the DNA synthesis is initialized by addition of the four nucleotides modified with a tag at the terminal phosphate. Upon incorporation of the nucleotide into the DNA, the specific tag will be released and dragged through the nanopore. For the time the tag is in the nanopore, the current through the pore will be partly blocked giving an electric signal allowing detection of the incorporation event. To discriminate which nucleotide has been incorporated and thus to allow sequencing, each nucleotide is labeled with a different tag that will show different characteristics of blocking the pore. In initial studies, nucleoside tetraphosphates (e.g. **23**) were used that contained differently sized polyethylene glycol molecules as tags that block the pore for different times and with different magnitudes [32]. In a recent study, discrimination was obtained using different DNA sequences attached to the terminal phosphate of nucleoside hexaphosphates (e.g. **24**) [87]. This setup was successfully applied to sequence DNA from single DNA molecules using an electronic signal as readout.

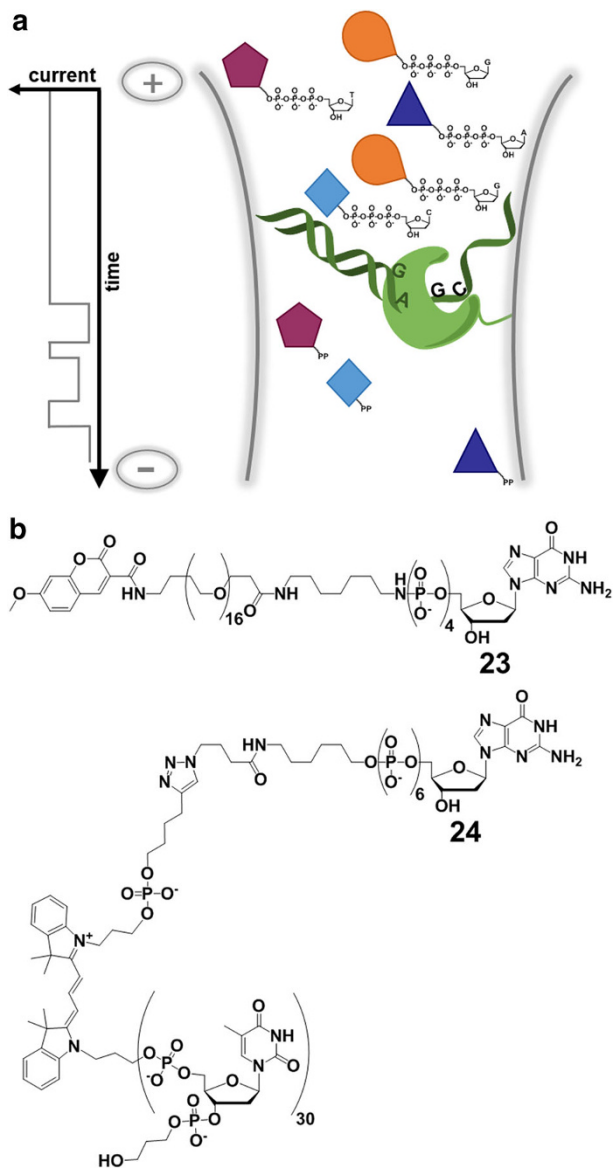


Fig. 13 a Concept for sequencing of DNA using nanopores [32, 87]. The DNA polymerase is immobilized on one side of an α -hemolysin pore and bound to a DNA primer complex. Upon incorporation of nucleotides with a tag at the terminal phosphate, this tag is released and dragged through the pore blocking the electric current through the pore with a specific pattern of duration and magnitude; **b** structures of nucleotides used in this concept harboring polyethylene glycol (**23**) or DNA (**24**) as tags

6 Fluorogenic and Chromogenic Probes

In order to study enzymatic cleavage of nucleotides on a broad basis, it is desirable to have nucleotides at hand that change their fluorescence characteristics upon hydrolysis and whose enzymatic turnover can, therefore, easily be measured following fluorescence intensity (Fig. 14). As the nucleobases are themselves electron-deficient aromatics, they are able to interact with certain fluorophores and alter their fluorescence characteristics. If the fluorophore is attached to the terminal phosphate of the nucleotide, upon enzymatic cleavage, the dye and the nucleobase are separated, resulting in a change of the fluorescence signal. It has been shown that the fluorescence spectrum of 1-aminonaphthalene-5-sulphonic acid is shifted when it is coupled to the γ -phosphate of ATP (γ -AmNS-ATP, **25**) [88]. Therefore, the hydrolysis of AmNS-ATP can be studied by monitoring its fluorescence spectra. This has been used to analyze the activity of the DNA-dependent RNA polymerase of *E. coli* and wheat germ, as well as the valyl-tRNA-synthetase of *E. coli*, snake venom phosphodiesterase (SVPD) of *Crotalus adamanteus* and potato apyrase. In the case of a 7-methyl-GTP analog that bears a pyrene moiety at the γ -phosphate, a strong decrease in its fluorescence intensity is observed upon enzymatic cleavage [89]. This was used to study the decapping enzyme DcpS.

The best sensitivity of detection is obtained for fluorogenic probes that increase in fluorescence intensity upon cleavage. This was realized for an uridine triphosphate (UTP) analog bearing 1-aminonaphthalene-5-sulphonic acid at the γ -phosphate (γ -AmNS-UTP, **26**) [14]. γ -AmNS-UTP is quenched in the intact state due to stacking between the dye and the nucleobase. The stacking effect is relieved upon cleavage of the phosphoanhydride bond and, therefore, results in a rise of fluorescence signal that was used to study the activity of RNA polymerases [90]. Using this assay, a screen for inhibitors of RNA polymerases from a 10,000-compound library was performed [91]. Another fluorogenic analog (**27**) is based on GTP γ S and harbors a BODIPY-FL fluorophore at the γ -phosphate [24]. The fluorescence of this dye is quenched by the guanine base. This analog has been used to study the activity of the human diadenosine triphosphate hydrolase FHIT [24], adenylyl cyclases [92] and GTPases [93].

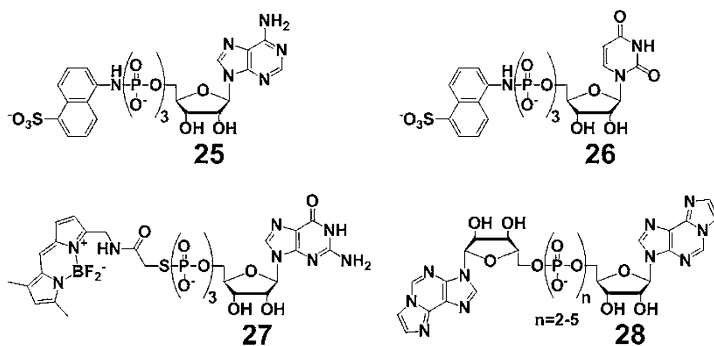


Fig. 14 Structures of different fluorogenic nucleotides used to study enzymatic activity

Fluorogenic nucleotide analogs are also used to study the cleavage of dinucleoside polyphosphates (Np_nNs) which are important signaling molecules in cells [4]. In these cases, as two nucleobases are present, the interaction of these nucleobases with each other can be used to measure enzyme activity. For this purpose, the nucleobase itself has to be rendered fluorescent. One nucleobase that is used for this is ϵ -adenine. It shows a high fluorescence quantum yield, which is significantly reduced if two ϵ -adenosine moieties are present in the same molecule as in $\epsilon Ap_n \epsilon As$ (**28**) [94]. Upon cleavage, the high fluorescence of the free ϵ -adenosine phosphates is restored, resulting in an increase of the fluorescence signal. This was used to study the activity of FHIT [94], NPP1 and NPP2 [95, 96], potato nucleotide pyrophosphatase [97], SVPD [98], spleen phosphodiesterase [98] and the RNases A, T1 and T2 [98]. It has also been used to identify the cellular components that are responsible for dinucleoside polyphosphate turnover from various cell lysates [99, 100].

7 FRET-based Nucleotide Probes

Using substrates that are labeled with two different fluorophores that are able to undergo FRET is a powerful technique to study enzymatic activity as it allows detection of the activity of an enzyme without the necessity to use further reagents or supplementary enzymes. Therefore, this technology was applied in complex biological systems like cell lysates or even intact cells. Furthermore, if fluorescence microscopy is used, enzymatic activity was studied with high temporal and spatial resolution. In the case of nucleotides, this concept (Fig. 15) is realized when a nucleotide is labeled with two fluorophores, one at the nucleoside part of the molecule and the other one at the terminal phosphate [31]. In this way, in the non-cleaved state the two fluorophores are able to interact through FRET. Upon excitation of the fluorescence donor, it will transfer the excitation energy to the acceptor, whose emission is detected. In contrast, no or only low direct emission of the donor is detectable. After enzymatic cleavage, FRET is no longer possible and the emission of the donor is detected, accompanied by a reduction in acceptor fluorescence. The ratio of the two fluorescence intensities, therefore, gives a concentration-independent measure for the cleavage of the nucleotide that can be detected by fluorescence spectroscopy or microscopy. To come up with a suitable nucleotide-based probe that fulfills this function, various parameters need to be optimized. On the one hand, the optical properties of the fluorophores need to be tailored for high changes in the fluorescence spectra. On the other hand, the nucleotides need to be optimized for acceptance by the enzyme of interest.

To study the properties of various fluorophores in the context of doubly modified nucleotides, a series of $O2'$ - and γ -modified ATP analogs (e.g. **29**) has been synthesized [101]. These molecules are cleaved by SVPD and can be investigated in regards to their fluorescence characteristics before and after cleavage. Intriguingly, a FRET pair consisting of the dyes Sulfo-Cy3 and Sulfo-Cy5 results in a more than 100-fold change of the ratio of the donor and acceptor fluorescence intensities upon cleavage. Furthermore, using the non-fluorescent acceptor Eclipse in combination

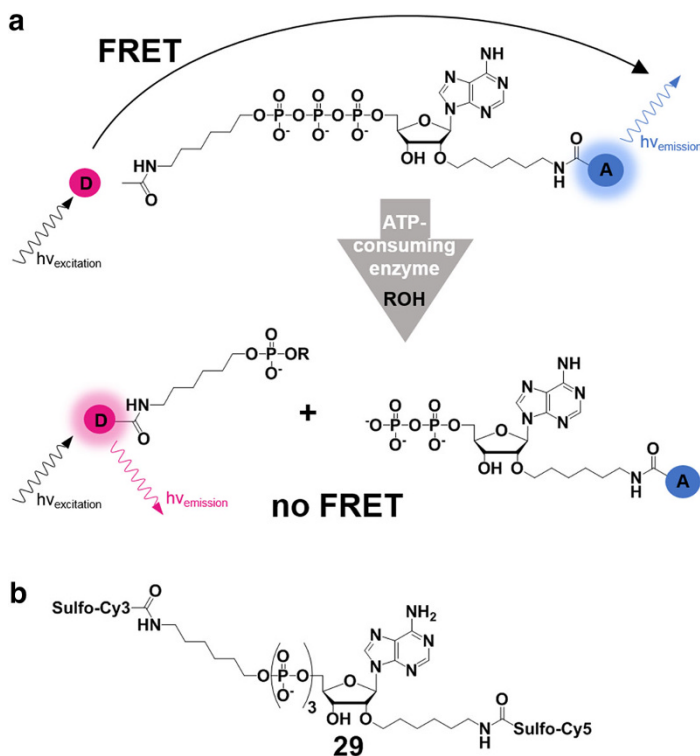


Fig. 15 **a** Concept of FRET-based nucleotides as enzyme substrates [31, 101]. In the intact state, the two fluorophores attached to the nucleotide are able to undergo FRET, resulting in fluorescence of the acceptor. After enzymatic turnover, FRET is no longer possible and direct emission of the donor is detected; **b** representative structure of a doubly fluorophore-labeled nucleotide used in this FRET-based assay

with Sulfo-Cy3 a fluorogenic probe can be generated, whose fluorescence increases more than 50-fold upon cleavage. Thus, these combinations of dyes are suitable to construct FRET-based nucleotide probes [101].

For optimizing the attachment of the dyes to the nucleotide, a two-step protocol has been developed that tests the two modifications separately. The modification at the nucleoside can be optimized by screening a variety of nucleoside-modified ATP analogs with a free phosphate chain [102]. For the modification at the phosphate chain, ATP analogs only modified at the terminal phosphate can separately be tested. In the case of UBA1, it has been shown that this enzyme accepts modifications at the δ -phosphate of an adenosine tetraphosphate analog [31]. UBA1 is able to tolerate modifications at the *N*6-position and the δ -phosphate simultaneously. If this scaffold is equipped with the two fluorophores, a probe is obtained that is able to directly detect UBA1

activity without the need of using any additional enzymes or reagent. This assay could be extended to the activating enzymes of Nedd8 and SUMO and was applied to study a small library of compounds in regards to their inhibition of UBA1 activity, identifying β -lapachone as an inhibitor of this enzyme [31]. One of the main advantages of this kind of assay is that it works in complex environments. As a model study, an $O2'$ - and γ -phosphate-modified ATP analog and an $O2'$ - and δ -phosphate-modified Ap_4 analog were studied in cell lysates of human cells [103]. It was observed that the ATP analog is stable in these lysates, whereas the Ap_4 analog is quickly turned over. Utilizing different chemical inhibitors and siRNA experiments, it was shown that the human diadenosine tetraphosphate hydrolase NUDT2 is responsible for this activity, making this compound an attractive tool to study NUDT2 activity in cell lysates. Going one step further, the analogs were also internalized into human cells using electroporation. Upon fixation and using an acceptor-photobleaching protocol, the turnover of these analogs can also be studied in a cellular environment. These studies showed that both nucleotide-based probes are rapidly turned over, whereas a non-hydrolyzable control that does not contain a nucleotide element is stable as indicated by constant high FRET efficiencies. In this way, a first proof was found that these analogs are actually usable to study nucleotide turnover directly in a cellular environment. In another study [104], a fluorogenic variant of ATP that contained a donor fluorophore and a dark quencher as acceptor at the γ -phosphate and the $C2$ -position, respectively, was used. This analog could be utilized to clarify the involvement of ATP turnover in a novel carbonylation pathway in lysates of the anaerobic bacterium *Desulfococcus biacutus*.

Using a diadenosine triphosphate (Fig. 16, 30) as the scaffold, this technology could also be efficiently transferred to monitoring the activity of the human diadenosine triphosphate hydrolase FHIT [8]. The two fluorophores are introduced to the two $N6$ -positions of Ap_3A , retaining good substrate characteristics for FHIT. Therefore, the activity of FHIT can be efficiently studied in vitro and the effect of inhibitors on this activity can be investigated. Furthermore, it could be shown that the activity of endogenous FHIT can be studied in lysates of human cells using this tool.

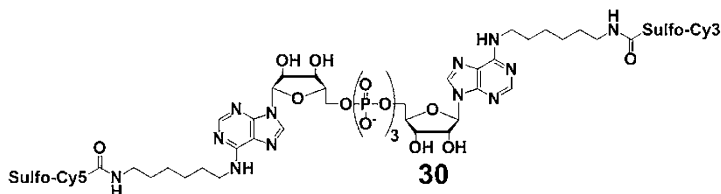


Fig. 16 Structure of the FRET-based Ap_3A analog used to study FHIT activity [8]

8 EPR-based Detection of Nucleotide Turnover

Based on the doubly modified design that is used to construct FRET-based fluorogenic enzyme substrates, nucleotides can be synthesized that allow to detect their turnover using EPR-based techniques (Fig. 17) [105]. In this case, instead of the two fluorophores, two spin-labels that can be detected by EPR are attached to the nucleotide (**31**). In the intact state, the two spin labels are interacting due to dipolar coupling. This coupling is lost upon enzymatic cleavage. Therefore, the extent of dipolar coupling was used to study enzymatic activity. Investigating an adenosine tetraphosphate analog with two nitroxyl radicals attached to the *N6*-position and the phosphate chain using pulsed EPR revealed that in most of the molecules the spin labels reside within a distance of less than 1.5 nm from each other. In this way, EPR was used to gain further information of the structure of the probes used in addition to studying their turnover. Furthermore, this short distance between the two spin labels allows monitoring the changes in dipolar coupling by the shape of continuous-wave (cw)-EPR spectra of the probe. cw-EPR allows much quicker measurement of the EPR spectra with less expensive equipment in comparison to pulsed EPR and is, therefore, a promising method to utilize these probes for detection of enzymatic activity. Using this EPR-based setup, the activity of SVPD was detected.

9 Fluorophosphates for Studying Nucleotide Turnover

Another possibility to monitor nucleotide turnover that can be done with only minimal modification of the nucleotide is using fluorophosphates as probe molecules (Fig. 18). In this case, one of the hydroxyl groups of the terminal phosphate is replaced by a fluorine. These molecules are easily synthesized by coupling phosphoimidazolides of nucleotides with fluoride from TBAF or with fluorophosphate or by activating fluorophosphate to the imidazolide and coupling it to nucleotides [17]. The cleavage of these molecules can be monitored by ^{19}F NMR spectroscopy [17]. This enabled the detection of the activity of SVPD with ATP γ F (**2**), of DcpS with 7-methyl-GDP β F (**32**) and 7-methyl-GMPF (**33**) and RNase T2 with cPAPPF (**34**). An important extension of this methodology is the ability to detect the turnover of fluorophosphate-modified nucleotides using fluorescence detection. This is possible if fluoride is liberated from the nucleotide as for the turnover of AMPF (**35**) with SVPD and of 7-methyl-GMPF (**33**) with DcpS [106].

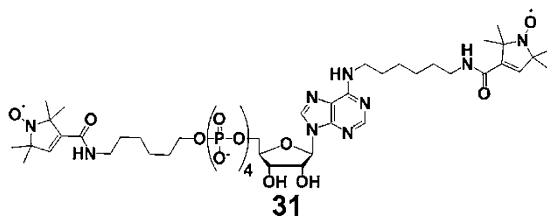


Fig. 17 Structure of an ATP analog modified with two spin labels used to detect enzymatic activity by EPR

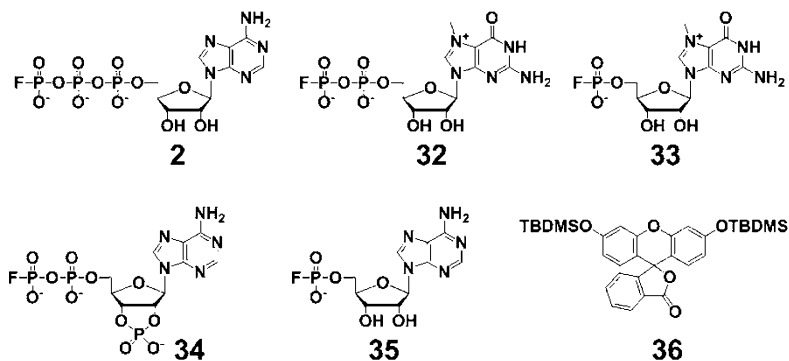


Fig. 18 Structures of fluorophosphate analogs of nucleotides (2, 32–35) that can be used to study enzymatic activity by ^{19}F -NMR [17] or using a fluoride sensor (36) [106]

The released fluoride is reacted with TBDMS-FL (36), a variant of fluorescein that has a TBDMS group on both phenolic hydroxyl groups and is non-fluorescent. In the presence of fluoride, the TBDMS groups are cleaved off, resulting in the high fluorescence of the unmodified fluorescein. This fluorescence signal is proportional to the amount of fluoride liberated in the first step and was used for measuring enzymatic activity. In this way, a fluorogenic signal is obtained from these minimally modified nucleotide analogs.

10 Conclusion

Taken together, nucleotides modified at the terminal position of the phosphate chain are powerful tools to study nucleotide-dependent processes. These methods usually utilize the fact that the modification is separated from the nucleoside moiety of the nucleotide after the enzymatic reaction. In this way, probes were developed that allow the detection of enzymatic activity using electrochemical [18, 46], affinity [9], chromogenic [69], fluorogenic [7, 31], NMR-based [17] and EPR-based technologies [105]. The applicability of these molecules spans a wide variety of different enzymes, including ATP-cleaving enzymes [31], GTPases [93], decapping enzymes [106], diadenosine polyphosphate hydrolases [8], DNA and RNA polymerases [33] and kinases [12]. They were utilized to study not only the activity of enzymes, but also obtain further important information on biological systems. In the case of kinases, the substrates can be labeled with the modified phosphate and the substrates of specific kinases were identified [36]. For DNA polymerases, the application of phosphate-modified dNTP analogs is used to establish next-generation sequencing techniques [33]. These allow sequencing of the genetic [33] and epigenetic [77, 83] information of DNA on a single-molecule level. In the light of these already established methods to utilize terminally phosphate-modified nucleotides, we strongly believe that a large number of novel applications will be made possible by these molecules in the future.

Acknowledgements Financial support by the Deutsche Forschungsgemeinschaft (Grant SFB 969) and the Konstanz Research School Chemical Biology is gratefully acknowledged. S.M.H. also acknowledges the Deutsche Forschungsgemeinschaft, the Studienstiftung des deutschen Volkes and the Zukunftskolleg of the University of Konstanz for stipends.

Compliance with Ethical Standards

Conflict of interest The authors declare that there is no conflict of interest regarding this publication.

References

1. Lipmann F (1941) Metabolic generation and utilization of phosphate bond energy. *Adv Enzymol Rel S Bi* 1:99–162
2. Masai H et al (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annu Rev Biochem* 79:89–130
3. Kalia D et al (2013) Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis. *Chem Soc Rev* 42:305–341
4. McLennan AG (2000) Dinucleoside polyphosphates—friend or foe? *Pharmacol Ther* 87:73–89
5. Robinson GA, Butcher RW, Sutherland EW (1968) Cyclic AMP. *Annu Rev Biochem* 37:149–174
6. Corrigan RM, Grundling A (2013) Cyclic di-AMP: another second messenger enters the fray. *Nat Rev Microbiol* 11:513–524
7. Sims PA, Greenleaf WJ, Duan H, Xie XS (2011) Fluorogenic DNA sequencing in PDMS microreactors. *Nat Methods* 8:575–580
8. Hacker SM, Mortensen F, Scheffner M, Marx A (2014) Selective monitoring of the enzymatic activity of the tumor suppressor Fhit. *Angew Chem Int Ed* 53:10247–10250
9. Patricelli MP et al (2007) Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochem US* 46:350–358
10. Suwal S, Pflum MKH (2010) Phosphorylation-dependent kinase-substrate cross-linking. *Angew Chem Int Ed* 49:1627–1630
11. Hiratsuka T (1983) New ribose-modified fluorescent analogs of adenine and guanine nucleotides available as substrates for various enzymes. *Biochim Biophys Acta* 742:496–508
12. Green KD, Pflum MKH (2007) Kinase-catalyzed biotinylation for phosphoprotein detection. *J Am Chem Soc* 129:10–11
13. Korlach J et al (2008) Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides Nucleic Acids* 27:1072–1083
14. Dhar G, Bhaduri A (1999) Synthesis and characterization of stacked and quenched uridine nucleotide fluorophores. *J Biol Chem* 274:14568–14572
15. Hofer A et al (2015) A modular synthesis of modified phosphoanhydrides. *Chem Eur J* 21:10116–10122
16. Allen JJ et al (2007) A semisynthetic epitope for kinase substrates. *Nat Methods* 4:511–516
17. Baranowski MR et al (2015) Synthesis of fluorophosphate nucleotide analogues and their characterization as tools for 19F NMR studies. *J Org Chem* 80:3982–3997
18. Song H, Kerman K, Kraatz H-B (2008) Electrochemical detection of kinase-catalyzed phosphorylation using ferrocene-conjugated ATP. *Chem Commun* 4:502–504
19. Martić S, Labib M, Freeman D, Kraatz H-B (2011) Probing the role of the linker in ferrocene-ATP conjugates: monitoring protein kinase catalyzed phosphorylations electrochemically. *Chem Eur J* 17:6744–6752
20. Lee SE et al (2009) Synthesis and reactivity of novel γ -phosphate modified ATP analogues. *Bioorg Med Chem Lett* 19:3804–3807
21. Serdjukow S et al (2014) Synthesis of γ -labeled nucleoside 5'-triphosphates using click chemistry. *Chem Commun* 50:1861–1863
22. Hacker SM, Mex M, Marx A (2012) Synthesis and stability of phosphate modified atp analogues. *J Org Chem* 77:10450–10454
23. Ermert S et al (2016) Different enzymatic processing of gamma-phosphoramidate and gamma-phosphoester-modified ATP analogues. *ChemBioChem*. doi:10.1002/cbic.201600590

24. Draganescu A, Hodawadekar SC, Gee KR, Brenner C (2000) Fhit-nucleotide specificity probed with novel fluorescent and fluorogenic substrates. *J Biol Chem* 275:4555–4560
25. Wanat P et al (2015) Ethynyl, 2-propynyl, and 3-butyryl C-phosphonate analogues of nucleoside di- and triphosphates: synthesis and reactivity in CuAAC. *Org Lett* 17:3062–3065
26. Kumar S et al (2005) Terminal phosphate labeled nucleotides: synthesis, applications, and linker effect on incorporation by DNA polymerases. *Nucleosides Nucleotides Nucleic Acids* 24:401–408
27. Sood A et al (2005) Terminal phosphate-labeled nucleotides with improved substrate properties for homogeneous nucleic acid assays. *J Am Chem Soc* 127:2394–2395
28. Hampton A, Kappler F, Picker D (1982) Species- or isozyme-specific enzyme inhibitors. 4. Design of a two-site inhibitor of adenylate kinase with isozyme selectivity. *J Med Chem* 25:638–644
29. Johnson TB, Coward JK (1987) Synthesis of oligophosphopeptides and related ATP γ -peptide esters as probes for cAMP-dependent protein kinase. *J Org Chem* 52:1771–1779
30. Ratnakar SJ, Alexander V (2005) Synthesis and relaxivity studies of a gadolinium(III) complex of ATP-conjugated DO3A as a contrast enhancing agent for MRI. *Eur J Inorg Chem* 2005:3918–3927
31. Hacker SM et al (2013) Fluorogenic ATP analogues for online monitoring of ATP consumption: observing ubiquitin activation in real time. *Angew Chem Int Ed* 52:11916–11919
32. Kumar S et al (2012) PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. *Sci Rep* 2:684
33. Eid J et al (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138
34. Kwon SW et al (2003) Selective enrichment of thiophosphorylated polypeptides as a tool for the analysis of protein phosphorylation. *Mol Cell Proteom* 2:242–247
35. Blethrow JD, Glavy JS, Morgan DO, Shokat KM (2008) Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc Natl Acad Sci USA* 105:1442–1447
36. Allen JJ, Lazerwith SE, Shokat KM (2005) Bio-orthogonal affinity purification of direct kinase substrates. *J Am Chem Soc* 127:5288–5289
37. Bishop AC et al (2000) A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* 407:395–401
38. Liu Y et al (1998) Engineering Src family protein kinases with unnatural nucleotide specificity. *Chem Biol* 5:91–101
39. Fouada AE, Pflum MKH (2015) A cell-permeable ATP analogue for kinase-catalyzed biotinylation. *Angew Chem Int Ed* 54:9618–9621
40. Senevirathne C et al (2016) The generality of kinase-catalyzed biotinylation. *Bioorg Med Chem* 24:12–19
41. Senevirathne C, Pflum MKH (2013) Biotinylated phosphoproteins from kinase-catalyzed biotinylation are stable to phosphatases: implications for phosphoproteomics. *ChemBioChem* 14:381–387
42. Kerman K, Chikae M, Yamamura S, Tamiya E (2007) Gold nanoparticle-based electrochemical detection of protein phosphorylation. *Anal Chim Acta* 588:26–33
43. Wang Z, Lévy R, Fernig DG, Brust M (2006) Kinase-catalyzed modification of gold nanoparticles: a new approach to colorimetric kinase activity screening. *J Am Chem Soc* 128:2214–2215
44. Wang Z, Lee J, Cossins AR, Brust M (2005) Microarray-based detection of protein binding and functionality by gold nanoparticle probes. *Anal Chem* 77:5770–5774
45. Ma C, Yeung ES (2010) Highly sensitive detection of DNA phosphorylation by counting single nanoparticles. *Anal Bioanal Chem* 397:2279–2284
46. Kerman K et al (2008) Peptide biosensors for the electrochemical measurement of protein kinase activity. *Anal Chem* 80:9395–9401
47. Martić S, Rains MK, Freeman D, Kraatz H-B (2011) Use of 5'- γ -ferrocenyl adenosine triphosphate (Fc-ATP) bioconjugates having poly(ethylene glycol) spacers in kinase-catalyzed phosphorylations. *Bioconjugate Chem* 22:1663–1672
48. Martić S et al (2012) Versatile strategy for biochemical, electrochemical and immunoarray detection of protein phosphorylations. *J Am Chem Soc* 134:17036–17045
49. Wang N et al (2015) Clickable 5'- γ -ferrocenyl adenosine triphosphate bioconjugates in kinase-catalyzed phosphorylations. *Chem Eur J* 21:4988–4999
50. Parang K, Kohn JA, Saldanha SA, Cole PA (2002) Development of photo-crosslinking reagents for protein kinase–substrate interactions. *FEBS Lett* 520:156–160
51. Petrousseva IO et al (2003) A new approach to the synthesis of the 5'-end substituted oligonucleotides using T4 polynucleotide kinase and γ -amides of ATP bearing photoreactive groups. *Dokl Biochem Biophys* 389:114–117

52. Green KD, Pflum MKH (2009) Exploring kinase cosubstrate promiscuity: monitoring kinase activity through dansylation. *ChemBioChem* 10:234–237
53. Wilke KE, Francis S, Carlson EE (2012) Activity-based probe for histidine kinase signaling. *J Am Chem Soc* 134:9150–9153
54. Freeman R, Finder T, Gill R, Willner I (2010) Probing protein kinase (CK2) and alkaline phosphatase with CdSe/ZnS quantum dots. *Nano Lett* 10:2192–2196
55. Wang L-J, Yang Y, Zhang C-Y (2015) Phosphorylation-directed assembly of a single quantum dot based nanosensor for protein kinase assay. *Anal Chem* 87:4696–4703
56. Xiao Y, Guo L, Jiang X, Wang Y (2013) Proteome-wide discovery and characterizations of nucleotide-binding proteins with affinity-labeled chemical probes. *Anal Chem* 85:3198–3206
57. Qiu H, Wang Y (2007) Probing adenosine nucleotide-binding proteins with an affinity-labeled nucleotide probe and mass spectrometry. *Anal Chem* 79:5547–5556
58. Xiao Y, Ji D, Guo L, Wang Y (2014) Comprehensive characterization of SGTP-binding proteins by orthogonal quantitative SGTP-affinity profiling and SGTP/GTP competition assays. *Anal Chem* 86:4550–4558
59. Villamor JG et al (2013) Profiling protein kinases and other atp binding proteins in arabidopsis using Acyl-ATP probes. *Mol Cell Proteomics* 12:2481–2496
60. Wolfe LM et al (2013) A chemical proteomics approach to profiling the ATP-binding Proteome of *Mycobacterium tuberculosis*. *Mol Cell Proteom* 12:1644–1660
61. Adachi J et al (2014) Proteome-wide discovery of unknown ATP-binding proteins and kinase inhibitor target proteins using an ATP probe. *J Proteome Res* 13:5461–5470
62. Nordin BE et al (2015) ATP acyl phosphate reactivity reveals native conformations of hsp90 paralogs and inhibitor target engagement. *Biochem US* 54:3024–3036
63. Xiao Y, Guo L, Wang Y (2013) Isotope-coded atp probe for quantitative affinity profiling of ATP-binding proteins. *Anal Chem* 85:7478–7486
64. Patricelli MP et al (2011) In situ kinase profiling reveals functionally relevant properties of native kinases. *Chem Biol* 18:699–710
65. Kwiatkowski N et al (2014) Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* 511:616–620
66. Deng X et al (2011) Characterization of a selective inhibitor of the parkinson's disease kinase LRRK2. *Nat Chem Biol* 7:203–205
67. Xie T et al (2014) Pharmacological targeting of the pseudokinase Her3. *Nat Chem Biol* 10:1006–1012
68. Chen Z et al (2015) Fluorogenic sequencing using halogen-fluorescein-labeled nucleotides. *ChemBioChem* 16:1153–1157
69. Vassiliou W et al (2000) Exploiting polymerase promiscuity: a simple colorimetric RNA polymerase assay. *Virology* 274:429–437
70. Kore AR, Yang B, Srinivasan B (2014) Efficient synthesis of terminal 4-methylumbelliferyl labeled 5-fluoro-2'-deoxyuridine-5'-O-tetraphosphate (Um-PPPP-FdU): a potential probe for homogenous fluorescent assay. *Tetrahedron Lett* 55:4822–4825
71. Kozlov M et al (2005) Homogeneous fluorescent assay for RNA polymerase. *Anal Biochem* 342:206–213
72. Korlach J et al. (2010) Chapter 20 - Real-Time DNA Sequencing from Single Polymerase Molecules. In: Nils GW (ed) *Method Enzymol*, vol 472. Academic Press, pp 431–455
73. Levene MJ et al (2003) Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299:682–686
74. Travers KJ et al (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* 38:e159
75. Coupland P et al (2012) Direct sequencing of small genomes on the pacific biosciences RS without library preparation. *Biotechniques* 53:365–372
76. Yang Y et al (2015) Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). *BMC Genom* 16:1–11
77. Flusberg BA et al (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7:461–465
78. Murray IA et al (2012) The methylomes of six bacteria. *Nucleic Acids Res* 40:11450–11462
79. Clark TA et al (2012) Characterization of DNA methyltransferase specificities using single-molecule. Real-time DNA sequencing. *Nucleic Acids Res* 40:e29

80. Song C-X et al (2012) Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat Methods* 9:75–77
81. Au KF et al (2013) Characterization of the human ESC transcriptome by hybrid sequencing. *Proc Natl Acad Sci USA* 110:E4821–E4830
82. Sharon D, Tilgner H, Grubert F, Snyder M (2013) A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol* 31:1009–1014
83. Vilfan ID et al (2013) Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J Nanobiotechnol* 11:8
84. Saletore Y et al (2012) The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol* 13:175
85. Pennisi E (2010) Semiconductors inspire new sequencing technologies. *Science* 327:1190
86. Munroe DJ, Harris TJR (2010) Third-generation sequencing fireworks at marco island. *Nat Biotechnol* 28:426–428
87. Fuller CW et al (2016) Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. *Proc Natl Acad Sci USA* 113:5233–5238
88. Yarbrough LR (1978) Synthesis and properties of a new fluorescent analog of ATP: adenosine-5'-triphospho- γ -1-(5-sulfonic acid) naphthylamide. *Biochem Biophys Res Commun* 81:35–41
89. Kasprzyk R et al (2016) Acetylpyrene-labelled 7-methylguanine nucleotides: unusual fluorescence properties and application to decapping scavenger activity monitoring. *Org Biomol Chem* 14:3863–3868
90. Schlageck JG, Baughman M, Yarbrough LR (1979) Spectroscopic techniques for study of phosphodiester bond formation by *Escherichia coli* RNA polymerase. *J Biol Chem* 254:12074–12077
91. Bhat J et al (2006) High-throughput screening of RNA polymerase inhibitors using a fluorescent UTP analog. *J Biomol Screen* 11:968–976
92. Vadakkadathmeethal K et al (2007) Fluorescence-based adenylyl cyclase assay adaptable to high throughput screening. *Comb Chem High Throughput Screen* 10:289–298
93. Jameson EE et al (2005) Real-time detection of basal and stimulated G protein GTPase activity using fluorescent GTP analogues. *J Biol Chem* 280:7712–7719
94. Asensio AC, Oaknin S, Rotllán P (1999) Fluorimetric detection of enzymatic activity associated with the human tumor suppressor Fhit protein. *BBA Protein Struct M* 1432:396–400
95. Asensio AC et al (2007) Biochemical analysis of ecto-nucleotide pyrophosphatase phosphodiesterase activity in brain membranes indicates involvement of NPP1 isoenzyme in extracellular hydrolysis of diadenosine polyphosphates in central nervous system. *Neurochem Int* 50:581–590
96. Oaknin S et al (2008) Binding of 5'-O-(2-Thiodiphosphate) to rat brain membranes is prevented by diadenosine tetraphosphate and correlates with ecto-nucleotide pyrophosphatase phosphodiesterase 1 (NPP1) activity. *Neurosci Lett* 432:25–29
97. Wierchowski J, Sierakowska H, Shugar D (1985) Continuous fluorimetric assay of nucleotide pyrophosphatase. kinetics, inhibitors, and extension to dinucleoside oligophosphatases. *BBA Protein Struct Mol Enzymol* 828:109–115
98. Tolman GL, Barrio JR, Leonard NJ (1974) Chloroacetaldehyde-modified dinucleoside phosphates. Dynamic fluorescence quenching and quenching due to intramolecular complexation. *Biochem US* 13:4869–4878
99. Ramos A, Rotllán P (1995) Specific dinucleoside polyphosphate cleaving enzymes from chromaffin cells: a fluorimetric study. *BBA Protein Struct Mol Enzymol* 1253:103–111
100. Rotllán P et al (1991) Di (1, N6-ethenoadenosine)5', 5'''-P1, P4-tetraphosphate, a fluorescent enzymatically active derivative of Ap4A. *FEBS Lett* 280:371–374
101. Hardt N, Hacker SM, Marx A (2013) Synthesis and fluorescence characteristics of ATP-based FRET probes. *Org Biomol Chem* 11:8298–8305
102. Hacker SM et al (2013) Fingerprinting differential active site constraints of ATPases. *Chem Sci* 4:1588–1596
103. Hacker SM, Buntz A, Zumbusch A, Marx A (2015) Direct monitoring of nucleotide turnover in human cell extracts and cells by fluorogenic ATP analogs. *ACS Chem Biol* 10:2544–2552
104. Gutiérrez Acosta OB et al (2014) Thiamine pyrophosphate stimulates acetone activation by *desulfococcus biacutus* as monitored by a fluorogenic ATP analogue. *ACS Chem Biol* 9:1263–1266
105. Hacker SM, Hintze C, Marx A, Drescher M (2014) Monitoring enzymatic ATP hydrolysis by EPR spectroscopy. *Chem Commun* 50:7262–7264
106. Baranowski MR, Nowicka A, Jemielity J, Kowalska J (2016) A fluorescent HTS assay for phosphohydrolases based on nucleoside 5'-fluorophosphates: its application in screening for inhibitors of mRNA decapping scavenger and PDE-I. *Org Biomol Chem* 14:4595–4604

Chemical Approaches to Studying Labile Amino Acid Phosphorylation

Alan M. Marmelstein^{1,2} · Javier Moreno¹ · Dorothea Fiedler¹

Received: 12 January 2017 / Accepted: 23 January 2017 / Published online: 6 February 2017
© Springer International Publishing Switzerland 2017

Abstract Phosphorylation of serine, threonine, and tyrosine residues is the archetypal posttranslational modification of proteins. While phosphorylation of these residues has become standard textbook knowledge, phosphorylation of other amino acid side chains is underappreciated and minimally characterized by comparison. This disparity is rooted in the relative instability of these chemically distinct amino acid side chain moieties, namely phosphoramidates, acyl phosphates, thiophosphates, and phosphoanhydrides. In the case of the *O*-phosphorylated amino acids, synthetic constructs were critical to assessing their stability and developing tools for their study. As the chemical biology community has become more aware of these alternative phosphorylation sites, methodology has been developed for the synthesis of well-characterized standards and close mimics of these phosphorylated amino acids as well. In this article, we review the synthetic chemistry that is a prerequisite to progress in this field.

Keywords Posttranslational modification · Protein phosphorylation · Phosphohistidine · Phosphoarginine · Phospholysine · Pyrophosphorylation

A. M. Marmelstein and J. Moreno contributed equally.

This article is part of the Topical Collection “Phosphate Labeling and Sensing in Chemical Biology”; edited by Henning Jessen.

✉ Dorothea Fiedler
fiedler@fmp-berlin.de

¹ Leibniz-Institut für Molekulare Pharmakologie, Robert-Rössle-Straße 10, 13125 Berlin, Germany

² Department of Chemistry, Princeton University, Washington Road, Princeton 08544, USA

1 Introduction

Protein phosphorylation is one of the most widespread posttranslational modifications (PTMs) and influences almost all aspects of cell biology [1]. The addition and removal of phosphoryl groups onto protein substrates has emerged as a key mechanism in signal transduction pathways, and is catalyzed by the large group of protein kinases and phosphatases. Misregulation of cellular information processing is associated with a number of diseases, and, consequently, many protein kinases have become much sought after drug targets [2].

Protein phosphorylation is known to occur on nine amino acids—Ser, Thr, Tyr, His, Lys, Arg, Cys, Asp, and Glu—but the major efforts to date have focused on just three modifications: phosphoserine (**1**, pSer), phosphothreonine (**2**, pThr), and phosphotyrosine (**3**, pTyr) (Fig. 1a) [3]. These phosphomonoesters are acid-stable and can withstand the acidic conditions typically used in phosphopeptide and phosphoprotein analysis. In contrast, phosphoramidates (Fig. 1b) [as encountered in phosphohistidine (**4**, pHis), phosphoarginine (**5**, pArg), and phospholysine (**6**, pLys)], acyl-phosphates [phosphoaspartate (**8**, pAsp) and phosphoglutamate (**9**, pGlu)], and thiophosphates [phosphocysteine (**7**, pCys)] are susceptible to hydrolysis under acidic conditions. Therefore, these modifications are often overlooked in proteomic analyses using conventional approaches.

A remarkable expansion of phosphorylation-based signaling was the discovery that phosphorylation may not be limited to the addition of just one phosphoryl group

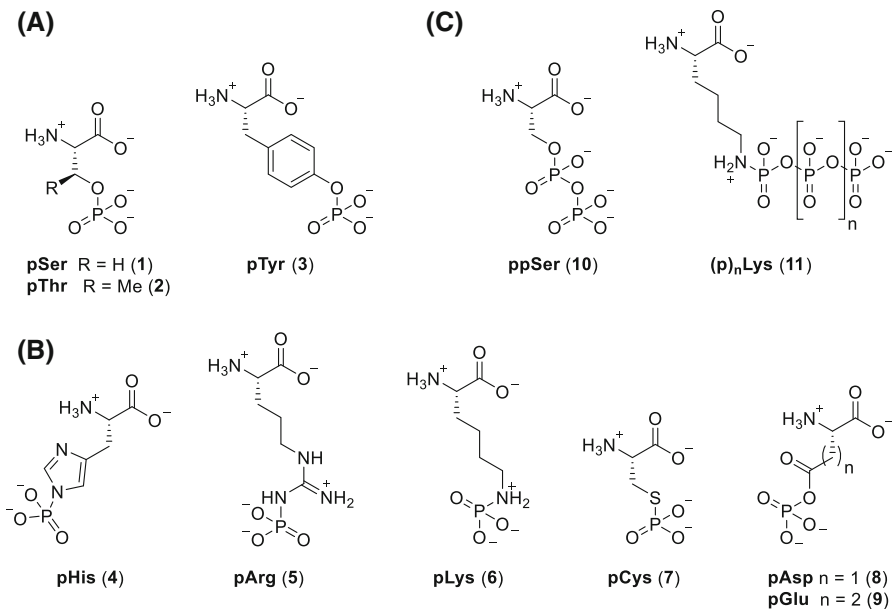


Fig. 1 **a** Phosphohydroxyacids, **b** phosphorylated amino acids with labile *P*-heteroatom bonds, and **c** pyrophospho- and polyphosphoamino acids. All phosphorylated amino acids are shown in their expected protonation states at physiological pH

per amino acid side chain (Fig. 1c). Phosphoserine residues were shown to undergo a PTM termed pyrophosphorylation, in which a β -phosphoryl group from an inositol pyrophosphate molecule can be transferred onto the pre-existing pSer to yield pyrophosphoserine (**10**, ppSer) [4]. Most recently, the covalent attachment of linear chains of inorganic polyphosphates to lysine residues was described (**11**, (P)_nLys) [5]. These unusual phosphorylation patterns provide an additional layer of complexity to the cell's phosphorylation network and pose a significant analytical challenge.

Because acid-labile phosphoryl groups and pyro- and polyphosphorylation sites have evaded standard phosphoproteomic detection techniques, chemical biologists have been hard pressed to develop new tools for the analysis of these modifications. Several excellent reviews exist, describing the initial syntheses and chemical properties of the acid-labile phosphorylated residues Arg, Lys, His, Asp, and Glu [6–8]. Considering how influential peptide and protein chemistry were in enabling the biological evaluation of pSer, pThr, and pTyr, this review focuses mainly on recent synthetic approaches to install labile phosphoryl-groups in the context of peptides and proteins. We will also discuss useful stabilized analogs of the labile modifications, and highlight the current and future applications of these tools.

2 Phosphohistidine

2.1 Background

Phosphorylation of histidine (**4**, pHis) is known to be critical for cellular signal transduction in prokaryotes, and reports of pHis in mammalian systems are steadily increasing. pHis was first discovered over 50 years ago in an alkaline digest of ³²P_i radiolabeled mitochondria isolated from bovine liver, though tools did not exist to characterize the biological significance of this modification [9]. Since then, it was found that phosphohistidine is utilized in the two-component signal transduction systems (TCSs) that typically monitor environmental conditions by sensing small molecule ligands in bacteria, fungi, and plants [10, 11]. Notable examples of TCSs include quorum sensing systems [12], which can modulate bacterial virulence [13] and antibiotic resistance induction pathways [14]. These systems usually consist of a dimeric membrane-bound histidine kinase (HK) sensor that auto-phosphorylates upon ligand binding, and a cytosolic response regulator (RR) that is phosphorylated on aspartate by the HK. Histidine phosphorylation is important for enabling the subsequent phosphoryl transfer to the RR aspartate residue, as phosphoryl transfer from Ser/Thr/Tyr residues would be thermodynamically difficult [10, 11]. In higher organisms, histidine phosphorylation has been observed on Histone H4, and has been associated with increased DNA synthesis [15, 16]. For other examples of histidine phosphorylated proteins and their functions, readers are referred to recent reviews by Perry et al. [10] and Kee et al. [17].

Phosphorylation on histidine is unique in that it can occur in two different isomeric forms: 1-pHis (**12**) and 3-pHis (**4**), also called π -pHis (for pro or “near”), and τ -pHis (for tele or “far”), respectively (Fig. 2). Crystal structures of proteins

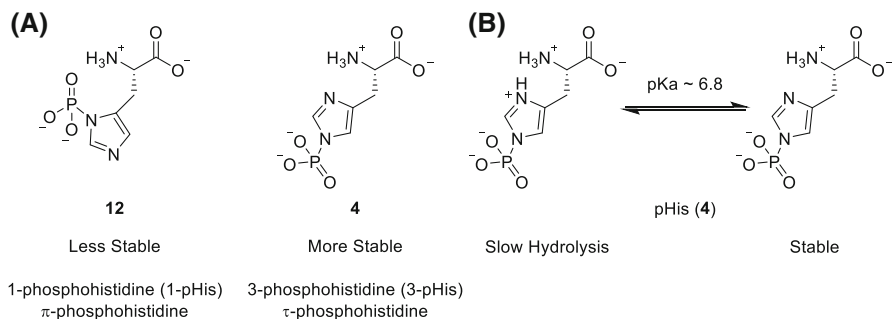


Fig. 2 **a** pHis isomers. 3-pHis is more hydrolytically stable than 1-pHis. **b** Below pH 6.8, the imidazole ring is protonated and hydrolysis of the P–N bond accelerates

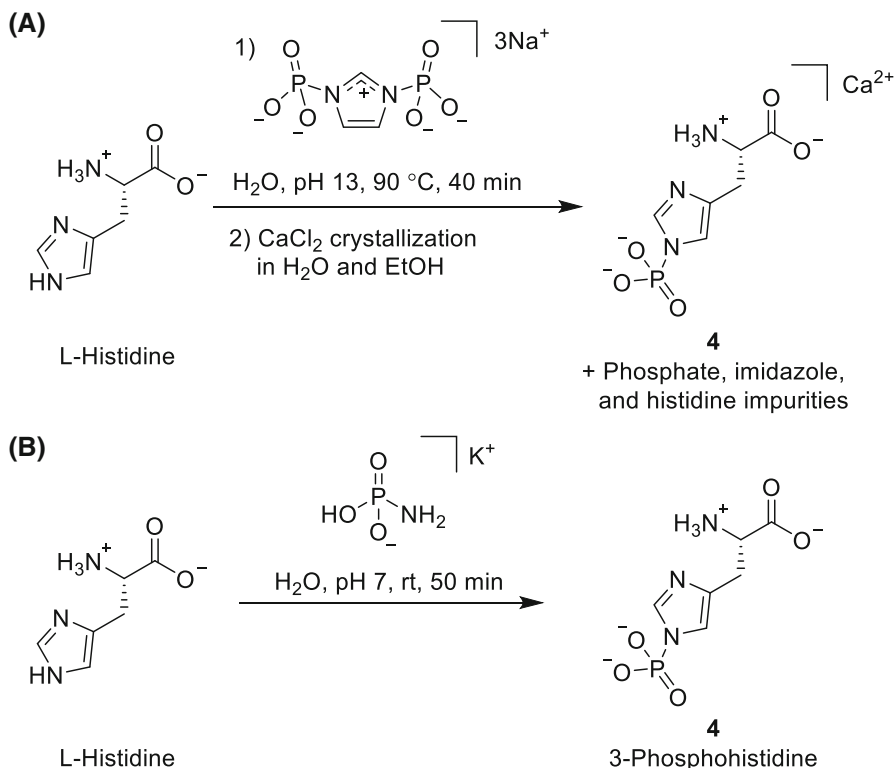
containing either 1-pHis or 3-pHis have been reported [18, 19], indicating that both isomers are biologically significant. It must be noted that the nomenclature of 1 and 3-pHis does not follow the proper IUPAC numbering of the imidazole ring, and this has given rise to discrepancies in the literature. However, the naming shown in Fig. 2 has become the de facto nomenclature for the field and will therefore be followed in this review.

The stability of the pHis isomers is highly pH-dependent. In strong acid, both isomers hydrolyze rapidly, with half-lives of only seconds in 1 M HCl at 49 °C [20] (compared to hours for hydroxyl-phosphorylated amino acids [21]). At high pH, however, pHis is stable [20] and able to survive harsh alkaline protein hydrolysis [9]. Between pH 4 and 6, both isomers are semi-stable, but 1-pHis hydrolyzes approximately ten times more rapidly than 3-pHis (Fig. 2a) [20]. The greater instability of 1-pHis is caused by its proximity to the positively charged alpha nitrogen, which enhances the electrophilicity of the phosphoryl group [6]. The rate of hydrolysis of 1-pHis between pH 6 and 9 roughly correlates to the proportion of pHis in which the imidazole ring is protonated ($pK_a \sim 6.8$ based on phosphorimidazole) (Fig. 2b) [22].

Stability is also a critical consideration when assessing the suitability of mass spectrometry for identifying pHis on protein substrates. Using a synthetic pHis-containing peptide (see next section, and [23]), it was found that electrospray ionization partially preserved phosphorylation, whereas matrix assisted laser desorption ionization (MALDI) and collision induced dissociation (CID) of electrospray-ionized peptides typically caused pHis degradation, preventing phosphorylation site assignment. Even so, the study was able to narrow down the phosphorylation site of synthetically histidine-phosphorylated peptides derived from human tyrosine phosphatase PTP-PEST using a combination of MS and ^1H NMR data [24, 25].

2.2 Synthesis

An early synthesis of pHis involved reacting the free amino acid with POCl_3 [26], though later reports indicated that this method was inefficient. Rosenberg [27]



Scheme 1a,b Syntheses of 3-pHis from L-histidine. **a** Phosphorylation with sodium diphosphorylimidazole in water at basic pH. **b** Phosphorylation with potassium phosphoramidate in water at neutral pH

reported the use of sodium diphosphorylimidazole to synthesize pHis (Scheme 1a), but purification and characterization of the product were incomplete. The observation that aromatic amines such as derivatives of pyridine and imidazole accelerated the degradation of electrophilic phosphorus species in aqueous solution [28, 29] informed the development of more efficient preparative syntheses and thorough characterization of pHis. Potassium phosphoramidate in neutral aqueous solution was able to provide pHis starting from un-protected L-histidine (Scheme 1b) [20, 30]. Over extended reaction times (ca. 36 h), 1-pHis formed in the early stages of the reaction degrades, leaving 3-pHis and small amounts of 1,3-diphosphohistidine [30]. An optimized version of this protocol was later published with variations to provide either 1-pHis, 3-pHis, or 1,3-diphosphohistidine as desired [31].

A somewhat more accessible procedure for synthesis of phosphohistidine calls for the phosphorylation of high molecular weight polyhistidine with POCl_3 , followed by alkaline hydrolysis of the peptide backbone to provide 3-pHis in ca. 18% yield and 1-pHis in ca. 3% yield [31]. The polymer's high ratio of histidine side chains to free N-termini renders the yield of α -N-phosphorylated histidine

byproducts negligible. However, the allure of this simpler procedure is diminished by racemization of the alpha position.

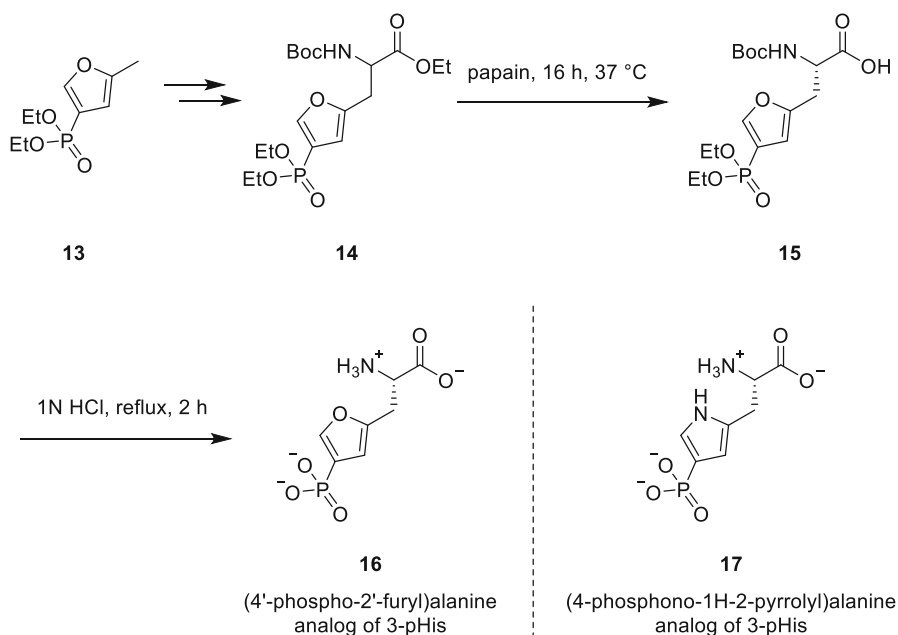
While no solid phase peptide synthesis (SPPS) method exists for site-selective incorporation of a phosphorylated histidine residue into a peptide, direct chemical phosphorylation of the histidine side chain has enabled the preparation of pHis peptides [24, 25] and proteins [23]. This method utilizes the observation that the potassium phosphoramidate reagent used to phosphorylate free histidine does not react with Ser, Thr, or Tyr side chains [30, 31]. The first synthetic pHis peptide, Ac-PLSFTNPL_pHSDDWH-CONH₂, was derived from the human tyrosine phosphatase PTP-PEST, a suspected substrate of histidine phosphorylation. After chemical phosphorylation, the sequence was used to test mass spectrometric detection methods (CID and MALDI) for pHis (described in the previous section) [24]. It should be noted that, although phosphorylated lysine residues were not detected on protein substrates treated with potassium phosphoramidate in a similar study [23], another report, in which the same phosphorylation protocol was used to synthesize phospholysine peptides [32], suggested that lysine phosphorylation can take place and that it is difficult to detect. Nevertheless, the potassium phosphoramidate phosphorylation protocol was later used on bovine serum albumin (BSA) to make a sequence-independent pHis protein standard for screening antisera for pan-specific pHis antibodies [33].

Muir and co-workers [34] were able to identify a unique pHis neutral loss pattern of $\Delta 80$ Da, $\Delta 98$ Da, and $\Delta 116$ Da using CID on synthetically phosphorylated histidine-containing peptides. Interestingly, $\Delta 98$ Da was the most prominent neutral loss, and an investigation of the fragmentation mechanism concluded that the phosphoryl group could first be transferred to nearby carboxylate residues or to the C-terminus before loss from the peptide ion. The neutral loss pattern was used successfully to target pHis-containing peptides identified with CID for MS/MS analysis using multistage activation (MSA). In conjunction with an immunoprecipitation enrichment step using a pan-specific pHis antibody (see below), this workflow found both previously characterized and novel pHis-containing proteins from an *Escherichia coli* cell lysate.

2.3 Stable Analogs of pHis

The synthesis of the phosphorylated amino acids and short peptides was crucial to the development of the pHis field. However, attempts to raise antibodies against pHis-containing antigens were unsuccessful, most likely due to rapid degradation of phosphohistidine in blood plasma [6]. Therefore, more recent synthetic efforts have focused on non-hydrolyzable C-phosphonate analogs of phosphohistidine for use as epitopes to raise antibodies.

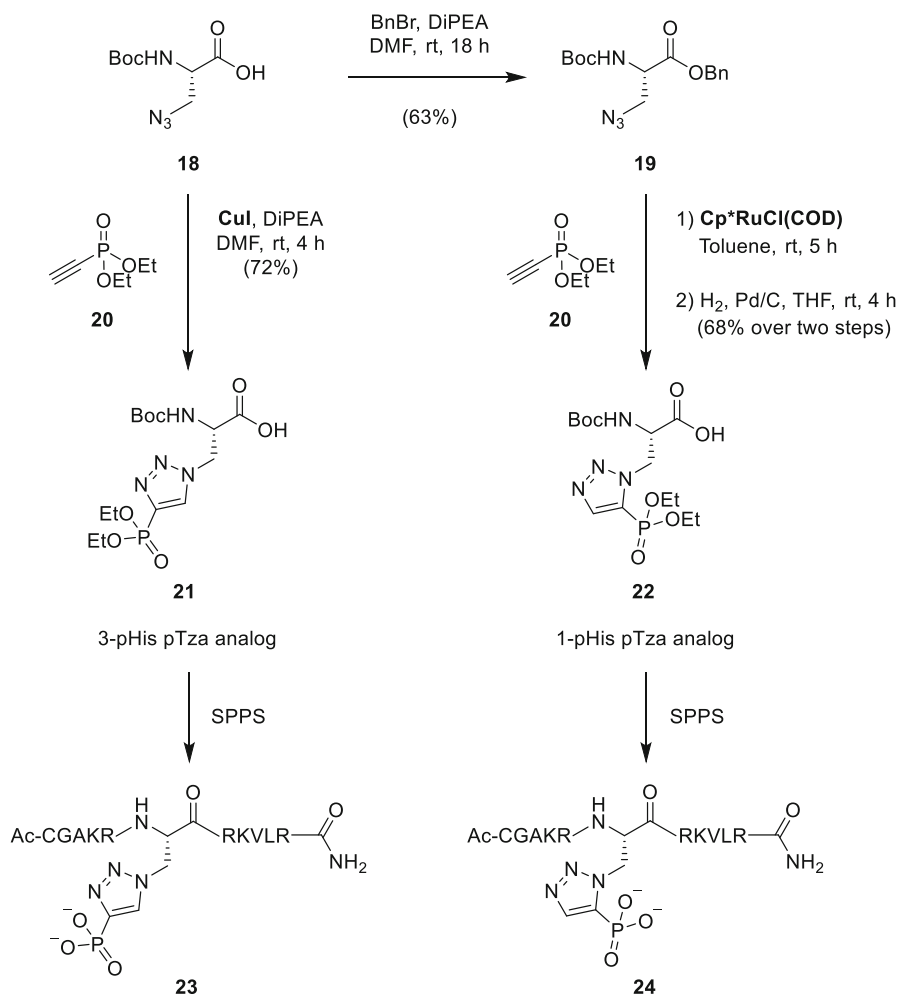
One of the first analog approaches was the synthesis of (4'-phospho-2'-furyl)alanine (**16**, Scheme 2) [35]. In order to install the unnatural side chain residue, the authors elaborated phosphorylated furan ring **13** to give racemic ethyl ester **14**. The chiral center was established using the esterase papain, which selectively hydrolyses (L)-amino acids. This approach provided carboxylic acid **15** with 40% isolated yield and 75% ee. Importantly, the (4'-phospho-2'-furyl)alanine



Scheme 2 Synthesis of non-hydrolyzable 3-pHis analog (4'-phospho-2'-furyl)alanine (**16**), which has a hydrogen bond acceptor (oxygen) in the same position as authentic 3-pHis. The (4'-phospho-2-pyrrolyl)alanine (**17**) 3-pHis analog did not raise 3-pHis antibodies [35]

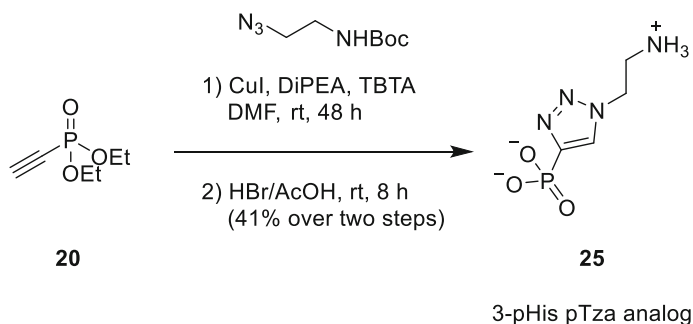
analog **16** places a hydrogen bond acceptor in the 2-position of the aromatic side chain, making it a closer electronic mimic of the authentic modification than the more obvious (4'-phospho-2'-pyrrolyl)alanine analog **17** [35]. Indeed, antibodies raised by analog **17** reacted only with the analog itself, and not with phosphohistidine [6]. Despite its apparent potential as a hapten, no report of using **16** to generate antibodies has surfaced to date.

A more rapid and versatile synthesis of a phosphohistidine analog was described by Muir and co-workers [16], who made use of [3 + 2] azide alkyne cycloadditions to create phosphoryltriazoalanine (pTza) analogs of pHis (Scheme 3). By utilizing either a copper or a ruthenium catalyst, the regioselectivity of the reaction could be switched to obtain either a 3-pHis pTza mimic (**21**) or a 1-pHis pTza mimic (**22**), respectively. Because of the route's brevity, efficiency, and reliance on readily available starting materials, sufficient quantities of the Boc-protected amino acids could be obtained for incorporation into peptides via Boc-mode SPPS. These constructs enabled the production of the first polyclonal antibodies against phosphohistidine in specific histone H4 peptide sequences. Significantly, the antibody was shown to be unreactive against pSer, pThr, pTyr, and non-phosphorylated histidine residues incorporated into the same position of the peptide sequence [16]. In order to make pTza-containing peptides more accessible, an Fmoc-protected version of the 3-pHis pTza amino acid analog **21** was later reported, though this required a more elaborate synthetic route [36].

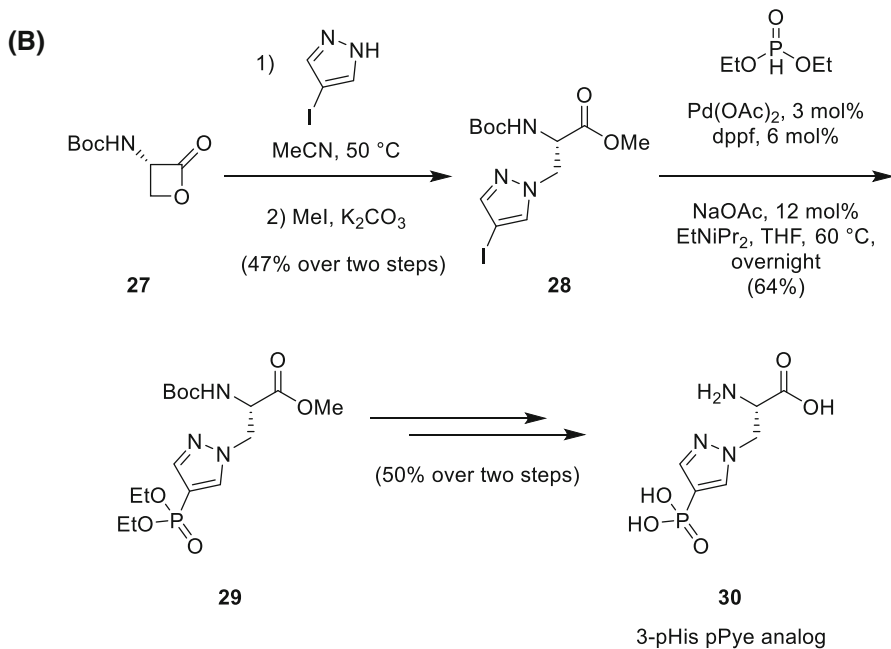
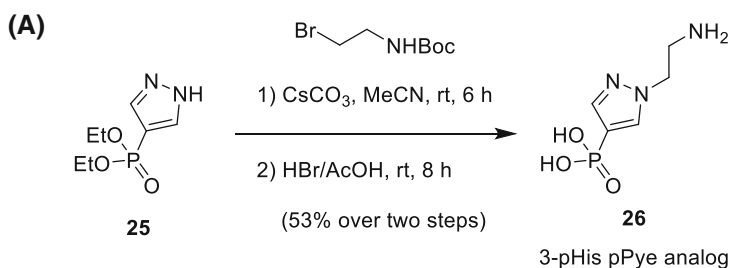


Scheme 3 Synthesis of phosphoryltriazolalanine (pTza) analogs of pHis and utilization in histone H4 tail peptide epitopes. Regiochemistry of the [3 + 2] cycloaddition can be switched by using either a copper catalyst (CuI for 3-pHis pTza analog) or a ruthenium catalyst [Cp*Ru(COD) for 1-pHis pTza analog]

A phosphohistamine version of the 3-pHis pTza analog (**25**, Scheme 4) was conjugated to keyhole limpet hemocyanin (KLH) and used to raise pan-specific polyclonal antibodies, though they have the drawback of some cross-reactivity with pTyr residues [33]. In an attempt to eliminate pTyr cross-reactivity, new pyrazole (pPye) pHis analogs **26** [37], and **30** [38] were synthesized that more closely mimic the phosphohistidine target (Scheme 5). While both constructs successfully raised polyclonal pHis antibodies with improved selectivity for pHis over pTyr, elimination of the carboxylate allowed the synthesis of pPye analog **26** to be expedited.



Scheme 4 Synthesis of 3-pHis pTza analog **25**



Scheme 5 Synthesis of pyrazole (pPye) 3-pHis analog **26** by Kee et al. [37], and **30** by Lilley et al. [38]

Hunter and co-workers [39] showed the full potential of pan-specific, monoclonal pHis antibodies by raising and utilizing them to identify previously unknown histidine-phosphorylated proteins. 1-pHis and 3p-His pTza analog epitopes were incorporated into 10-mer peptide sequences of random Ala and Gly residues. The peptides were conjugated to immunogenic KLH and rabbits were used to raise monoclonal antibodies capable of distinguishing between 3-pHis and 1-pHis. The sequence-independence of the pan-specific mAbs allowed pulldown and identification of both previously annotated and novel 1-pHis and 3-pHis substrates. Gene ontology analysis showed a preference for proteins involved in processing and modifying nucleic acids. Excitingly, antibody-staining of live cells showed a burst of 3-pHis in spindle poles in of dividing HeLa cells. Further insights into the nature of histidine phosphorylation in eukaryotes using these antibody-based tools and those like them are much anticipated.

3 Phosphoarginine

3.1 Background

Phosphoarginine (**5**, pArg) belongs to the family of phosphagens, which are molecules that possess a transferable phosphoryl group and buffer ATP concentrations in energy demanding tissues like muscle [40]. In invertebrates, pArg is the most common phosphagen, while, in vertebrates, this role is played by phospho-creatine. Arginine kinase catalyzes the transfer of the phosphoryl group between ATP and Arg to form ADP and pArg [40].

Protein arginine phosphorylation has not been studied extensively due to experimental difficulties in phosphoramidate research, caused by the acid lability of the P–N bond. Due to the sparse number of examples [41], pArg appeared to be scarcely present, and was originally found to occur on histone proteins [42]. The discovery of the first dedicated protein arginine kinase, McsB in bacteria represented a considerable breakthrough [43]. Subsequently more than 100 pArg modification sites in the Gram-positive model organism *Bacillus subtilis* were identified using an arginine phosphatase mutant strain [44, 45]. As this modification is more abundant than *O*-phosphorylation in *B. subtilis*, arginine phosphorylation represents a fundamental regulatory mechanism for cell physiology and survival [44]. McsB was shown to be involved in the heat-shock stress response [46], and to be decisive for the virulence of Gram-positive pathogens such as *Staphylococcus aureus* [47], stressing the importance of studying this protein modification. Most recently, it was found that pArg could play a critical role as a protein degradation tag, promoting the identification of aberrant proteins by the ClpC–ClpP protease, which is functionally analogous to the eukaryotic ubiquitin–proteasome system [48].

The most chemically stable form of pArg is found in the pH range from 4.4 to 9.6 (Fig. 3). While pArg suffers from extremely fast hydrolysis at low pH (1–3.5), it is stable at neutral pH and relatively stable at higher pH. Unlike pHis and pLys, it decomposes readily in hot alkali [49]. The biochemistry of phosphagens [40, 50] as well as the thermodynamic, kinetic and in vitro stability of pArg were reviewed in

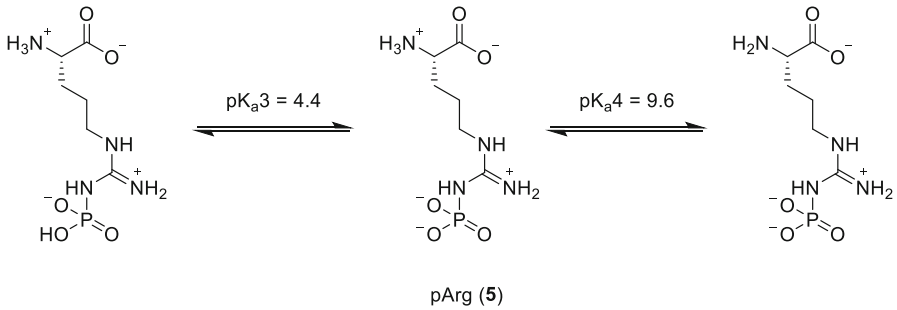
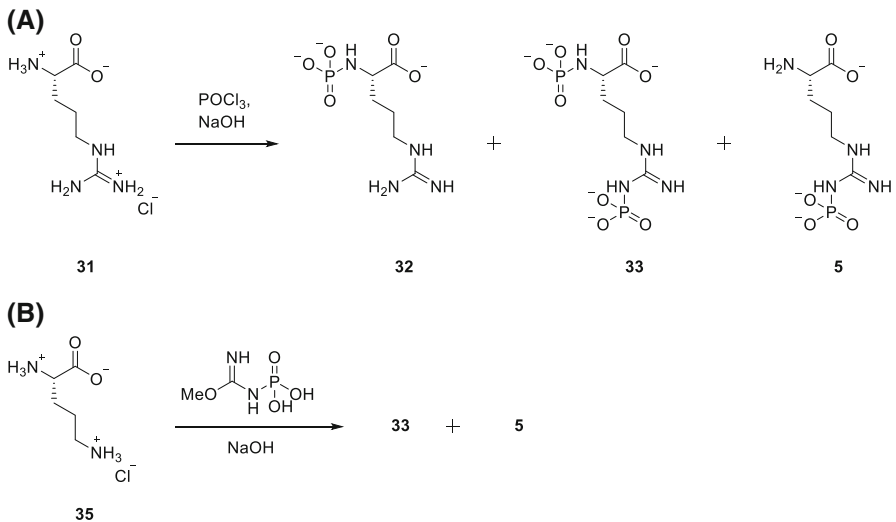


Fig. 3 Physiologically stable form of pArg undergoes sequential protonation/deprotonation outside the pH range of 4.4–9.6

greater depth elsewhere [7]. The stability of pArg in a protein context was studied by Clausen et al. [45], corroborating the rapid hydrolysis at low pH observed for free pArg.

Due to the inherent acid lability of the P–N bond in pArg, standard MS techniques have often failed to detect this modification. Consequently, acidic conditions must be specifically avoided [45] and selective enrichment of arginine-phosphorylated proteins is recommended prior to the phosphoproteomic analysis [51]. Thermal activation-based fragmentation techniques such as CID and higher-energy collisional dissociation (HCD) possess a higher tendency to produce false localizations, hence electron-transfer dissociation (ETD), a milder ionization technique, is the preferred method for LC–MS/MS detection of the pArg-sites in proteins and peptides [48, 52].



Scheme 6 Syntheses of pArg with **a** POCl_3 and **b** a phosphoramidate. pArg derivatives are shown in the fully deprotonated state present at high pH

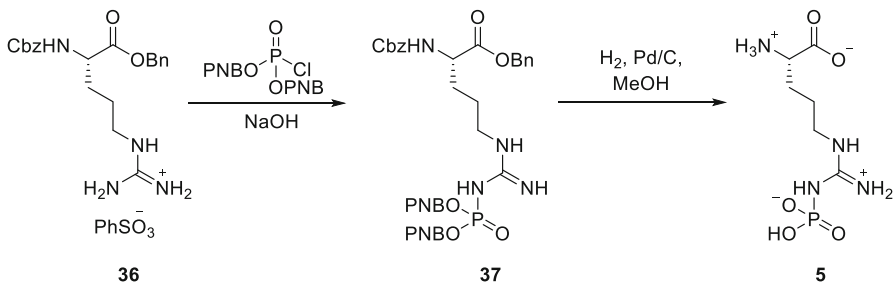
3.2 Synthesis

The first successful chemical synthesis of pArg (**5**) was achieved by treating arginine hydrochloride (**31**) with POCl₃ (Scheme 6a) [53]. However, *N*- α -phosphoarginine (**32**) and *N*- α , *N*- ω -diphosphoarginine (**33**) were also formed, and had to be removed by ion exchange chromatography. An alternative approach relied on the treatment of ornithine (**35**) with *O*-methyl-*N*-phosphorylisourea under basic conditions (Scheme 6b). This procedure efficiently produced pArg in 75% yield; however, 10% contamination of *N*- α , *N*- ω -diphosphoarginine (**33**) was generated, which was subsequently removed by fractional crystallization [54]. By introducing more robust *N*-protecting groups, such as the carboxybenzyl (Cbz) moiety (**36**), it was possible to obtain the protected pArg derivative **37** by treating with bis-(*p*-nitrobenzyl)chlorophosphate (Scheme 7). After deprotection by hydrogenolysis, pArg (**5**) was obtained in 73% yield [54]. While the above synthetic routes provided standards for chromatographic identification of this amino acid in protein digests, the strong alkaline conditions can cause racemization of the alpha position.

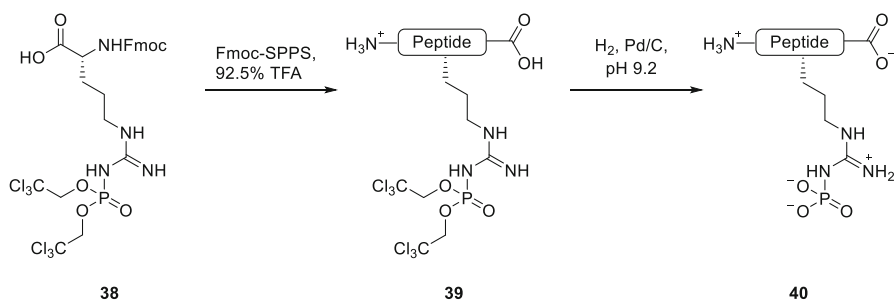
Initial methods for obtaining peptides containing pArg relied on treatment of the peptide with POCl₃, which unspecifically modified nucleophilic amino acid side chains [55]. SPPS using pre-phosphorylated arginine building blocks enabled the successful synthesis of tailor-made peptides bearing pArg-units [56]. For this building block approach, the choice of the trichloroethyl (Tc) protecting group for the phosphoguanidinium moiety was crucial. After acid-mediated global deprotection and cleavage of the peptide, the Tc group could be removed by hydrogenolysis at slightly alkaline pH (Scheme 8). While compatible with many peptide sequences, the use of a palladium catalyst during the Tc deprotection precludes the presence of cysteine in the final peptide sequence.

3.3 Stable Analogs of pArg

Since the P–N bond in pArg is extremely acid-labile, unstable in hot alkali, and sensitive to heat, the detection and isolation of pArg from biological sources has proven particularly difficult [7]. Furthermore, antigens containing pArg epitopes were dephosphorylated rapidly in serum, and were too unstable to directly produce



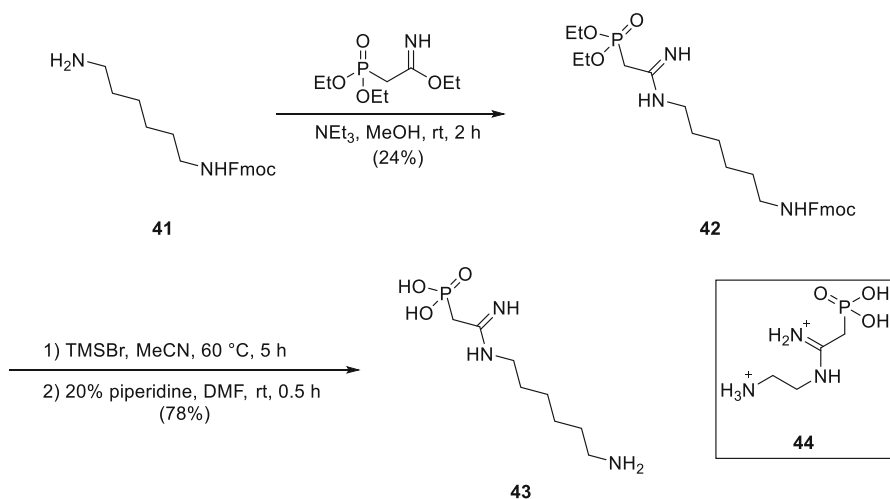
Scheme 7 Efficient protecting group strategy for the synthesis of pArg



Scheme 8 Solid phase peptide synthesis (SPPS) of pArg-bearing peptides

the corresponding antibody [51]. Clausen and co-workers [57] recently reported using *in vitro* phage-display as an alternative. While the antibody generated this way was useful for *in vitro* studies with recombinant proteins, the antibody's low affinity for pArg prevented its use in cellular studies. To circumvent these drawbacks and obtain a sequence-independent, high-affinity pArg antibody, acid-stable pArg mimics have been employed. Fuhrmann et al. [58] reported the synthesis of a non-hydrolyzable analog (**43**, Scheme 9) and its application for generating a pArg antibody in rabbits. Herein, the acid-labile P–N bond was replaced with a stable C-phosphorylated amidine functional group.

Fmoc protected 1,6-diaminohexane (**41**) was reacted with freshly prepared ethyl 2-(diethoxyphosphoryl)-acetimidate to generate compound **42**, which was subsequently deprotected over two steps, yielding the final PO₃-amidine (**43**). This hapten was conjugated to KLH carrier protein via the primary amine, and then injected into rabbits. After immunization, the resulting serum presented strong recognition of pArg, validating this analog as an effective mimic of pArg. By contrast, immunization of rabbits with an SO₃-amidine-KLH conjugate elicited only SO₃-



Scheme 9 Synthesis of non-hydrolyzable pArg analog **11** and related compound **12**

amidine specific antibodies that did not show any significant cross reactivity to pArg [58]. More recently, Zhao and co-workers [59] reported the gram-scale synthesis of a stable pArg analog, 2-((2-ammonioethyl)amino)-2-iminoethyl phosphonic acid (**44**), which produced the first mouse polyclonal antibodies specific to pArg containing proteins. The pArg antibody obtained with hapten **44** had an even higher affinity and lower detection limit than the antibody obtained with **43** [59].

4 Phospholysine

4.1 Background

Phospholysine (**6**, pLys) was first identified from rat liver cells by treating the non-denatured lysates with γ [^{32}P]ATP, performing an alkaline hydrolysis of the labeled proteins in strong base, and then analyzing the phosphorylated amino acid content by paper chromatography [60]. Although various kinases have been identified that are capable of mediating lysine phosphorylation, the lack of substrate specificity or clearly defined function has left the biological relevance of these phosphorylation events unclear [15, 61–65], highlighting the need for new methods to study the modification and tie it to specific biological functions [7].

Perhaps the main reason that phospholysine has been so minimally characterized is its extremely rapid hydrolysis in acidic conditions, which are routinely encountered when handling phosphorylated protein samples. Because a titration curve of pLys has not been reported, and because the phosphoramidate moiety in pLys is distal from the alpha amine and carboxylate, the pK_{a} s and hydrolysis rate constants for *n*-butyl phosphoramidate determined by Benkovic et al. [66] have been cited as models for pLys behavior by Besant et al. [7]. Readers are referred to their review for a detailed discussion of pLys stability, though some of the key points are highlighted below.

In contrast to carbonyl amides, in which the nitrogen lone pair electrons are substantially delocalized into the carbonyl's pi orbitals, phosphoramidates have mostly single P–N bond character [6, 67, 68]. Therefore, in contrast to amides, the pK_{a} s of phosphorylated amines are minimally perturbed, and remain protonated at

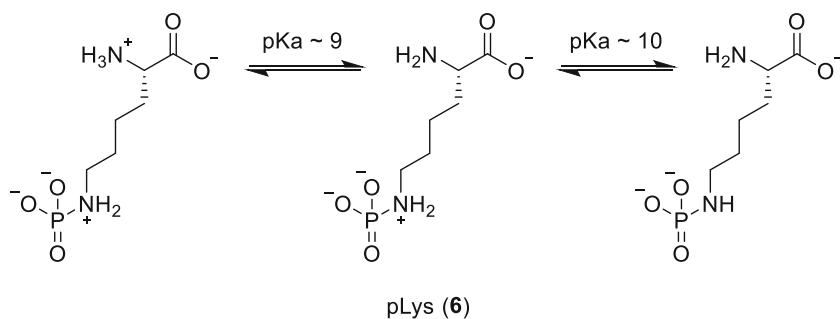


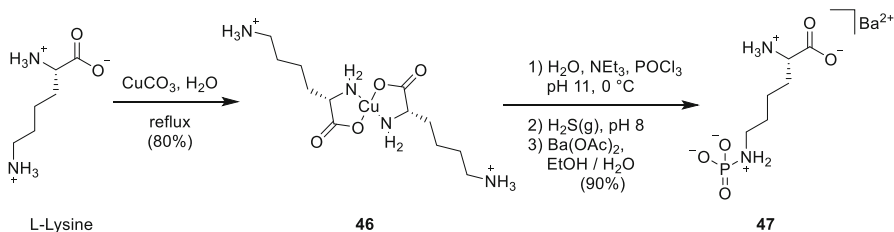
Fig. 4 Selected protonation states and approximate pK_{a} s of pLys (**6**)

physiological pH (Fig. 4). Below \sim pH 5.5, when the phosphoryl group starts to become protonated, hydrolysis is extremely rapid, but moderates between pH 5.5 and 9.0, where just the amidate nitrogen is protonated. Only when the amidate nitrogen is deprotonated above \sim pH 9.0 does this moiety become stable [66]. Indeed, in contrast to the evanescence of this modification at low and neutral pH, the high pK_a (\sim 40) of the deprotonated lysine side chain renders it an extremely poor leaving group, and therefore very base stable: in one study, only 11% of starting pLys degraded when heated to 100 °C in 9 M KOH for 9 h [7, 60]. Interestingly, a recently prepared pLys phosphorus diester (**51**, Scheme 11) in the context of a peptide sequence was found to have increased stability [69], raising questions about how the behavior of this modification on biological macromolecules could differ from that of single amino acid model compounds.

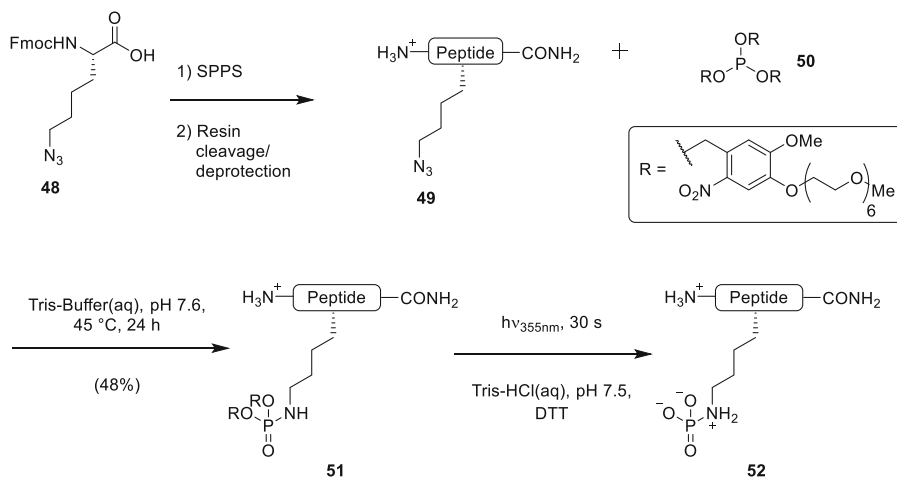
4.2 Synthesis

The first synthetic pLys standards were prepared by treating the free amino acid with POCl_3 , and separating the resulting *N*- ϵ -phosphorylated and *N*- α -phosphorylated products by ion exchange and preparative zone electrophoresis chromatography [60]. Subsequent improvements to the synthetic procedures for making pLys included applying the poly-lysine peptide phosphorylation and alkaline hydrolysis approach also used for pHis [31]. A much gentler method starting from the monomeric amino acid was used by Fujitaki et al. [70], who formed copper chelate **46** to protect the alpha nitrogen before phosphorylating the side chain with POCl_3 (Scheme 10). The copper was removed by precipitation with hydrogen sulfide gas, and the product was recrystallized as the barium salt to provide **47** in 72% overall yield. Unfortunately, a considerably lower yield was obtained when this protocol was attempted on arginine. Finally, Wei et al. [31] reported that the potassium phosphoramidate protocol for the preparation of pHis (described above in Scheme 1) could also be used to prepare pLys. This protocol was employed by Kowalewska et al. [32] to synthesize phosphorylated lysine peptides, and investigate their tractability to mass spectrometry analysis. Although their peptide substrates contained lysine and arginine residues, characterization by ^{31}P NMR was consistent with selective lysine phosphorylation.

The best characterized and most selective technique for synthesis of pLys-containing peptides was developed recently by Hackenberger and co-workers [69]. This technique takes advantage of the bio-orthogonal nature of the Staudinger-



Scheme 10 Copper chelate method of protecting alpha nitrogen during lysine phosphorylation



Scheme 11 Site-selective pLys peptide synthesis via a Staudinger-phosphite reaction

phosphite ligation and the stability of the azide coupling partner to the conditions of SPPS. Commercially available Fmoc-6-azidonorleucine (**48**) was incorporated into a peptide chain and, after resin cleavage and deprotection, coupled to protected phosphite **50** (Scheme 11). Both a base-labile 2-cyanoethyl group and a photo-labile and water soluble *o*-nitrobenzyl moiety (as in **50**) were tested and found to be effective protecting groups, site-selectively providing pLys-containing peptides in good yields. This protocol was later adapted to the solid phase using a base-labile resin [71]. Lysine phosphorylation of two synthetic peptides derived from the N-terminal sequence of histone H1 was confirmed by ^{31}P - ^1H NMR HMBC in which coupling of the phosphorus nucleus to the lysine aliphatic protons was observable. The site selectivity was also corroborated by MS/MS, using ETD to fragment the peptide backbone [69].

The synthetic availability of phosphorylated lysine peptides opens the door to the development of antibodies against this modification, as well as the more rigorous characterization of its stability in the context of various peptide sequences. Together with the mass spectrometry method that has been validated as a direct result of pLys peptide synthetic availability, these tools should enable the assessment of the true biological significance of lysine phosphorylation *in vivo*.

5 Phosphocysteine

5.1 Background

Phosphocysteine (**7**, pCys) functions as an intermediate in the phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS) [72]. Cysteine phosphorylation also occurs in cysteine-dependent protein phosphatases (CDPs), which catalyze the hydrolysis of phosphoester bonds via the formation of a pCys intermediate during

the dephosphorylation of pTyr residues by protein tyrosine phosphatases [73–75]. Additionally, pCys is involved in bacterial signaling and regulation [76–78]. The isolation and characterization of pCys peptides has been challenging due to the acid-lability of the P–S bond, which has prevented further identification of pCys sites using standard phosphoproteomic approaches [79]. In addition to its natural biological functions pCys has recently been employed as a mimetic for pSer in histone H3 [80] and pThr in protein kinase p38 α [81, 82].

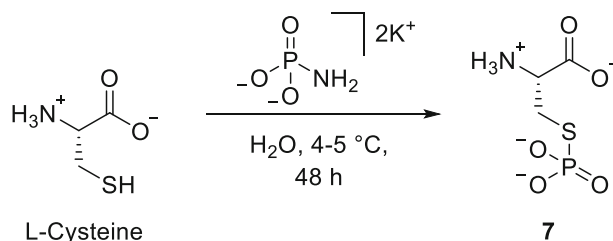
The stability of pCys is relatively high between pH 7 and pH 12 [83, 84], but the phosphothioester bond undergoes rapid hydrolysis at a pH of 2–4. Interestingly, at pH values lower than 2, the hydrolysis rate slows down [85]. Moreover, it was observed that *S*-phosphorylated cysteamine and *S*-phosphorylated mercaptopropionic acid undergo hydrolysis at a maximal rate at a pH of 3–4 (with a half-life of about 15 min at 37 °C) [84].

The P–S bond lability and the ease of hydrolysis at low pH values, have made the detection of pCys residues in complex samples a difficult task [79]. MS/MS techniques that rely on thermal CID and HCD do not offer satisfactory results. Using ETD with supplemental activation (known as EThcD), to improve the fragmentation of the charge-reduced precursor ion [86], enabled complete sequence coverage of peptides without neutral losses or phosphate rearrangements [87].

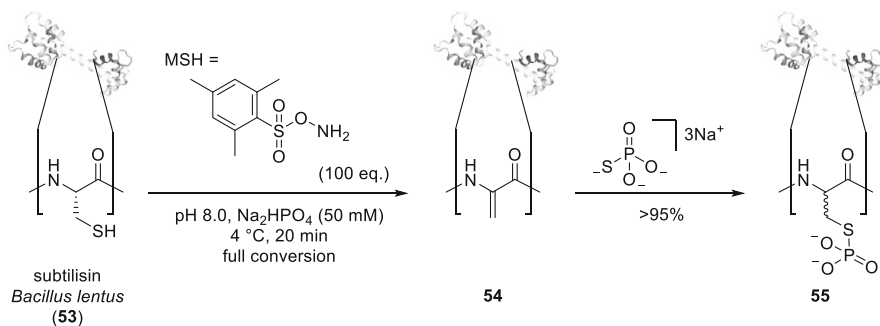
5.2 Synthesis

Chemical synthesis of pCys (**7**) was achieved by treatment of an aqueous solution of L-cysteine with potassium phosphoramidate (Scheme 12) [88]; unfortunately, no conversion or isolated yield was reported for this procedure.

Site-selective chemical cysteine phosphorylation of peptides and proteins, was first reported by Davis and co-workers [89] following a two-step procedure for installing pCys residues (Scheme 13). The method relies on generating dehydroalanine (Dha) followed by nucleophilic addition of sodium thiophosphate to produce the corresponding pCys residue. For example, a mutant (S156C) of the serine protease subtilisin from *Bacillus lentus* (**53**), which contains a single surface-exposed Cys, was treated with *O*-mesitylenesulfonylhydroxylamine (MSH), resulting in rapid and complete conversion of Cys156 to Dha156 (**54**) by an oxidative elimination mechanism. Subsequent addition of sodium thiophosphate enabled the synthesis of the pCys modified protein (**55**). However, this protocol delivered an



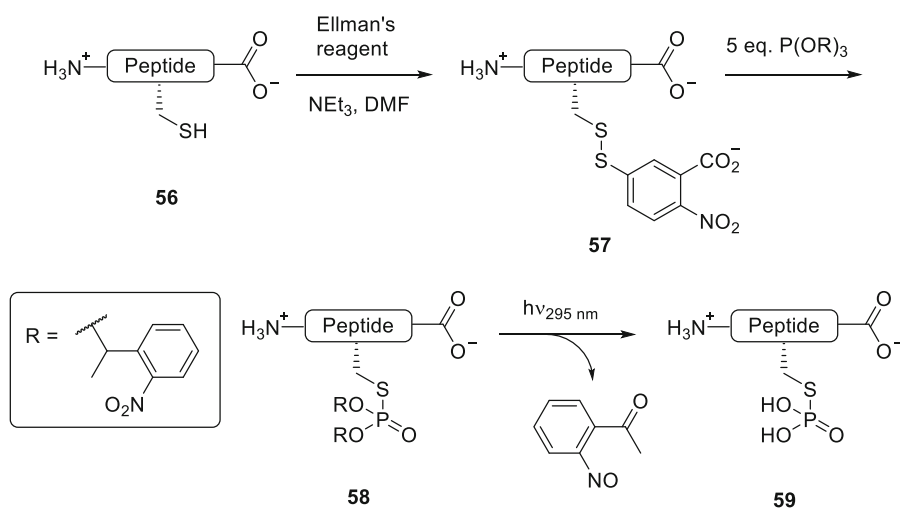
Scheme 12 Chemical synthesis of pCys **7**



Scheme 13 Synthesis of a pCys post-translationally modified protein

epimeric mixture of pCys proteins. Further optimization of the Dha generation conditions exploiting a bis-alkylation-elimination mechanism with a bis-halo compound [90], enabled the synthesis of a pCys-modified histone H3, which was shown to be useful as a pSer mimic [80].

Unfortunately, the elimination conditions required to prepare Dha residues, and the resulting lack of stereoselectivity, impose certain limitations on the general applicability of this chemical tool in the functional analysis of phosphorylated Cys residues [91]. A novel chemoselective and stereochemically defined phosphorylation strategy for Cys residues was recently reported by Hackenberger and co-workers [87]. Treatment of Cys residues with Ellman's reagent resulted in the formation of electron-poor electrophilic disulfides (57), which were subsequently treated with nucleophilic phosphites to obtain peptides bearing phosphorothiolate diesters (58). The photocleavable diesters were then deprotected by irradiation with 295 nm light to obtain free pCys residues (59, Scheme 14). The successful synthesis



Scheme 14 Chemoselective synthesis of pCys-containing peptides (59)

of the pCys-containing peptides was then used in a bottom-up proteomic approach to effectively identify and characterize an endogenous pCys peptide by means of tandem ESI-ETcD MS/MS from the glucose-specific transporter IICBGlc involved in the bacterial phosphotransferase system (PTS) [87].

6 Phosphoaspartate and Phosphoglutamate

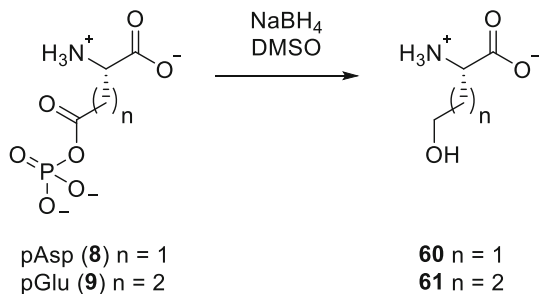
6.1 Background

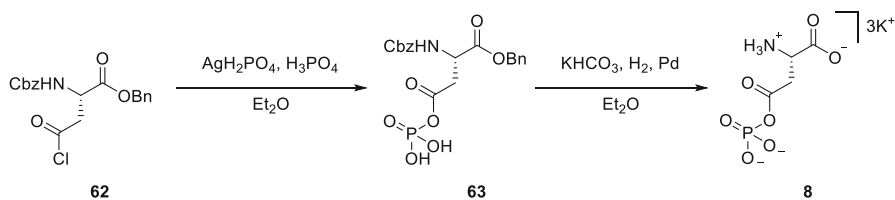
Phosphoaspartate (**8**, pAsp) and phosphoglutamate (**9**, pGlu) (Scheme 15) belong to the family of mixed phosphoanhydrides, which are reactive intermediates in the biosynthesis of aminoacyl-tRNAs. Acyl phosphates are nature's equivalent to the synthetic chemist's acyl chlorides and anhydrides: the side chain carboxyl groups of glutamate and aspartate are activated by phosphorylation, facilitating carbonyl substitution and reduction in a number of essential metabolic pathways, which include the biosynthesis of amino acids such as lysine, threonine and methionine from pAsp, and glutamine and proline from pGlu [8]. Studying the biochemistry of protein aspartate/glutamate phosphorylation has proven to be particularly difficult due to the chemical lability of the acyl phosphate functional group. pAsp occurs as an enzymatic intermediate in a number of enzyme-catalyzed reactions, such as those involving the haloacid dehalogenase superfamily [92, 93] as well as in two-component signaling systems [94–97]. While a review of the literature indicates that aspartate phosphorylation is a rare and transient event, rarer still is glutamate phosphorylation, which has only been reported to occur on prothymosin, a protein suggested to be involved in the production, processing, or export of RNA [98–100].

The phosphoanhydride groups of pAsp and pGlu are unstable under neutral, acidic and alkaline conditions [101]. At neutral pH, pAsp and pGlu hydrolyze readily, and react with organic bases such as hydroxylamine. In the pH range 4–10, 30% of pAsp is hydrolyzed after 30 min, and the hydrolysis rate increases exponentially at higher or lower pH.

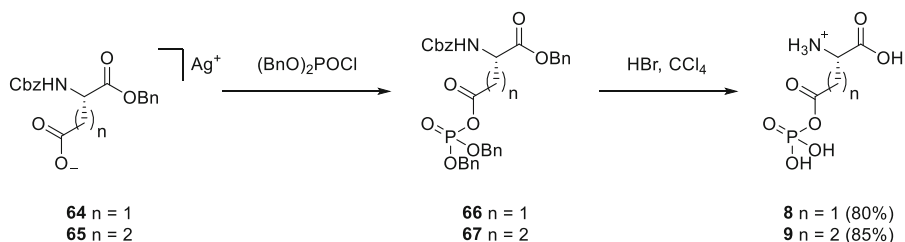
Due to their extreme lability, indirect detection methods are employed for the MS-detection of free or protein-bound pAsp and pGlu, which involve the reductive

Scheme 15 pAsp and pGlu are commonly reduced to **60** or **61** respectively for their MS-detection





Scheme 16 First chemical synthesis of pAsp (**8**)



Scheme 17 Synthesis of pAsp (**8**) and pGlu (**9**)

cleavage of the phosphoamino acids with NaBH_4 to give homoserine (**60**) and 4-hydroxy-1-aminovaleric acid (**61**), respectively (Scheme 15) [102]. Notably, under these conditions, aspartate and glutamate are resistant to reduction. Direct determination of pAsp and pGlu currently relies on Fourier transform infrared spectroscopy [103] or ^{31}P -NMR spectroscopy [104, 105].

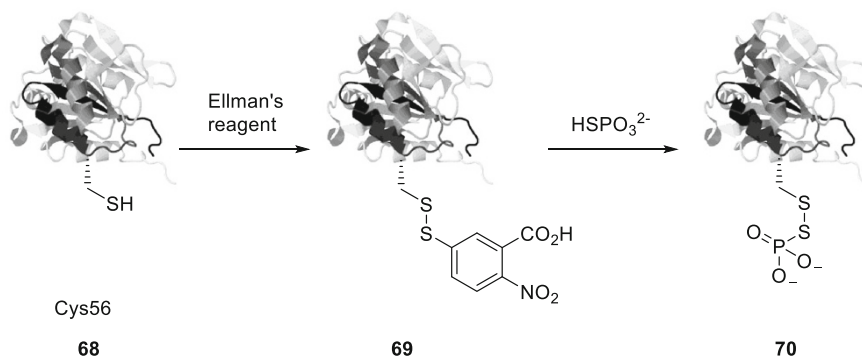
6.2 Synthesis

pAsp (**8**) was first synthesized from a protected aspartic acid (**62**) (Scheme 16) [106, 107]. Reaction of **62** with silver phosphate gave the acyl phosphate **63** in about 80% yield. Deprotection by hydrogenolysis in cold aqueous potassium bicarbonate provided the free pAsp **8** in solution with 28% yield. The solution of pAsp at approximately pH 6.5 could be stored at $-20\text{ }^\circ\text{C}$, but showed signs of slow degradation [107].

Another approach that was utilized for pAsp and pGlu, involves the initial reaction of the silver carboxylate of the protected derivatives **64** and **65** with dibenzyl chlorophosphate, providing fully protected phosphoamino acids **66** and **67** in excellent yields (Scheme 17). Protonolysis with anhydrous hydrogen bromide produced the free amino acids **8** and **9** in high purity [108].

6.3 Stable Analogs of pAsp

pAsp is involved in amino acid biosynthesis in prokaryotes and fungi, via its conversion into aspartate semialdehyde by the essential aspartate-semialdehyde dehydrogenase. The absence of this enzyme in mammals prompted the investigation of pAsp analogs as potential antibiotics [8].



Scheme 18 Synthesis of a stable analog (**70**) of pAsp in CheB, applying site-directed mutagenesis and chemical modification

Notably, the incorporation of a pAsp mimic (**70**) into a protein was achieved [109] for the bacterial response regulator methyl esterase CheB. CheB is activated by phosphorylation of Asp56. Substitution of a cysteine residue at position 56 (**68**) allowed this site to be chemically modified with Ellman's reagent, followed by a disulfide exchange of **69** with thiophosphate to give the phosphoprotein **70** (Scheme 18). The analog had a half-life of 28 days and displayed a comparable activity to the native pAsp in CheB, which rendered it valuable for the characterization of the active form of CheB, which has an extremely short half-life of just 2 s at 25 °C [110].

6.4 Stable Analogs of pGlu

Several non-hydrolysable analogs of pGlu were synthesized and evaluated as inhibitors of glutamine synthetase (GS) [111] which is one of the most central enzymes in nitrogen metabolism. GS catalyzes the conversion of glutamate and ammonia to glutamine in the presence of ATP. GS inhibitors have potential applications as anti-cancer drugs, as it was proven that glutamine supply is essential for the growth of some cancer lines [112]. However, integration of pGlu analogs into peptides and proteins has not been described to date, presumably due to the scarcity of known pGlu containing proteins and the uncertainty about relevance of this modification.

7 Pyrophosphoserine

7.1 Background

Protein serine pyrophosphorylation (**73**, ppSer) is a recently discovered PTM mediated by the highly phosphorylated intracellular small molecule messenger 5-diphosphoinositol pentakisphosphate (5PP-InsP₅, **72**). A study conducted in 2004, in which the β-phosphoryl group of 5PP-InsP₅ was enzymatically labeled with ³²P,

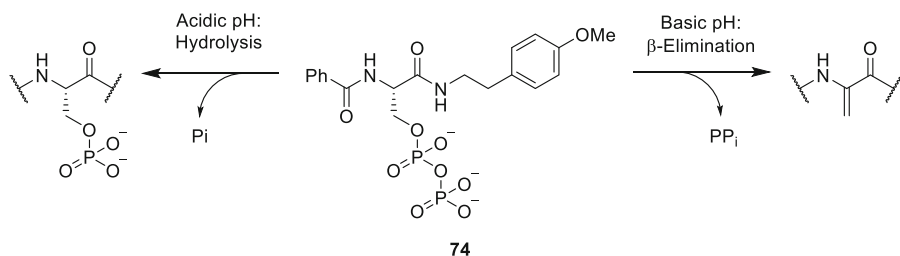
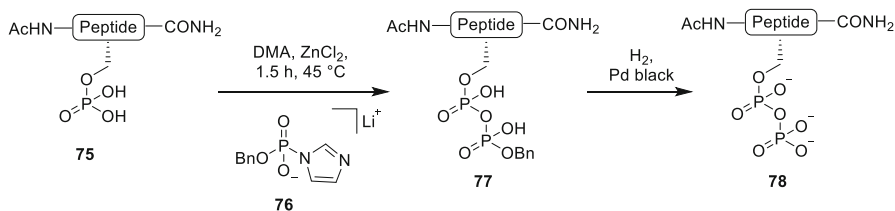


Fig. 6 Pyrophosphoserine analog **65** for stability studies. Degradation was monitored by analytical HPLC

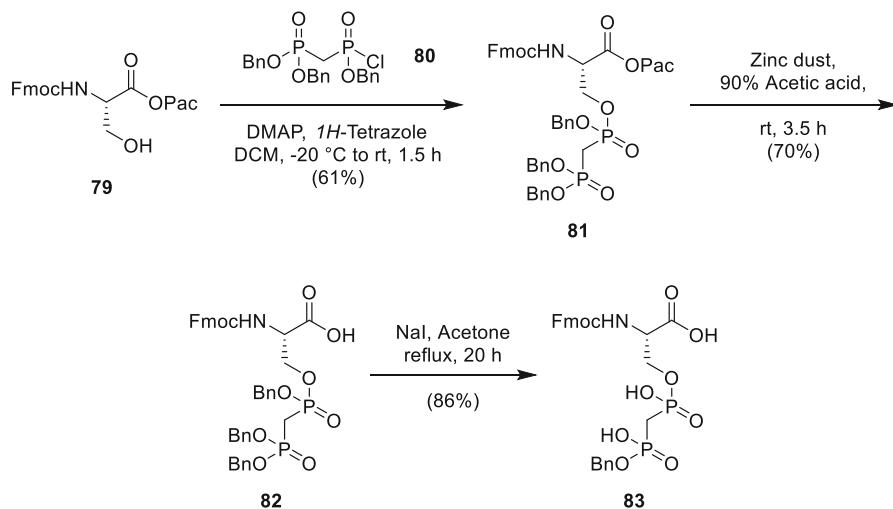


Scheme 19 Chemical pyrophosphorylation of peptides

7.2 Synthesis

In light of stability studies showing that di-esters of the ppSer analog **74** were unstable under the conditions of Fmoc-mode SPPS, Fiedler and co-workers [119] explored a global approach for the pyrophosphorylation of pre-phosphorylated synthetic peptides. Phosphorimidazolide reagents such as **76** were found to functionalize pre-existing phosphoryl groups selectively over other nucleophilic amino acid side chains, and provided pyrophosphorylated peptides in high yields after reverse-phase HPLC purification (Scheme 19). The free pyrophosphopeptides (**78**) were obtained via hydrogenolysis to remove the benzyl protecting groups. The phosphorimidazolide reagents also function in methanol and water, expanding their utility and allowing compatibility with more polar phosphorylated substrates, including a peptide fragment of NOPP140, an *in vitro* identified target of protein pyrophosphorylation. In addition to probing the stability and enzymatic reversibility of this modification [120], these pyrophosphopeptides were used to develop tools for *in vivo* identification of pyrophosphorylated proteins, including an affinity reagent [121], a gel stain [122], and a mass spectrometry-based identification method capable of differentiating bis-phosphorylated peptides from pyrophosphorylated peptides (Penkert et al. unpublished data).

With the goal of raising antibodies against pyrophosphorylated peptide sequences, non-hydrolyzable methylene-bisphosphonate (PCP) peptide analogs were developed [123]. Starting from Fmoc and phenacyl (Pac) protected *L*-serine (**79**), the bisphosphonate group was installed using ((bis(benzyloxy)phosphoryl)methyl) phosphonochloridate (**80**) (Scheme 20). Removal of the Pac group gave



Scheme 20 Synthesis of Fmoc-protected PCP-serine building block **83**

the tribenzyl protected Fmoc-PCP-L-serine **82**. One benzyl group could be removed from each phosphoryl unit on **82** using NaI to yield **83**.

Building block **83** was successfully incorporated into several peptide sequences from *in vitro* identified pyrophosphorylation substrates using standard SPPS conditions and reagents. Furthermore, these peptides were stable to common methods used to append peptide antigens to carrier proteins such as glutaraldehyde ligation, maleimide conjugation, and native chemical ligation [123]. The generation of antibodies with carrier proteins made using the above techniques is currently in progress.

8 Lysine Polyphosphorylation

Most of the diversity of protein phosphorylation discussed so far is generated by changing the amino acid to which the phosphoryl group is tethered. Only more recently, with the discovery of protein pyrophosphorylation, have scientists begun to appreciate the ways in which protein phosphorylation can be diversified by changing the character of the phosphoryl group itself. It now appears that nature has taken this strategy still further: just last year, evidence that chains of inorganic polyphosphate (poly-P) can covalently modify protein substrates was disclosed [5]. Poly-P is a polymer of repeating phosphoryl groups connected through phosphoanhydride bonds. It is present in all forms of life and functions primarily as a means of phosphate and energy storage [124] and a chaperone-like factor preventing protein misfolding and aggregation under stress [125]. In higher mammals, it has recently been shown to have roles in neuronal and glial cell signaling [126], as well as α -synuclein amyloid fibril formation and stability [127].

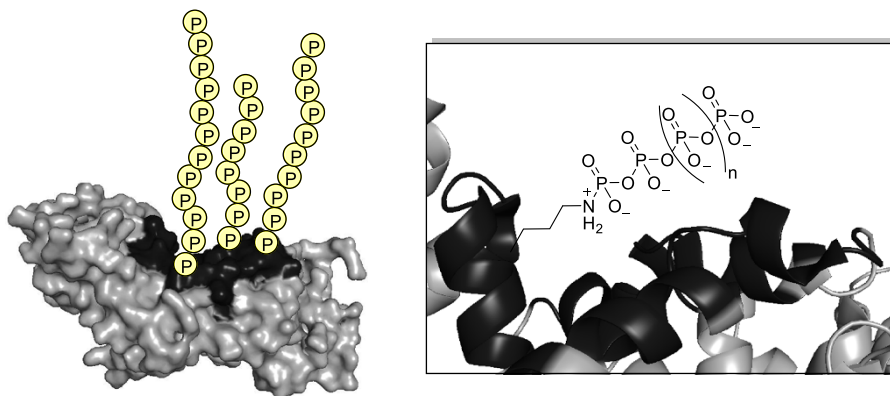


Fig. 7 $(P)_n$ Lys occurs in poly-acidic serine and lysine (PASK) sequences and has been found to affect the localization and activity of Top1 and Nsr1

Protein lysine polyphosphorylation (**11**, $(P)_n$ Lys) was discovered after the observation that the *S. cerevisiae* protein Nuclear Signal Recognition 1 (Nsr1) ran at a much higher molecular weight on a gel than expected. Knockout of VTC4—the subunit of the Vacuolar Transporter Chaperone (VTC) complex responsible for catalyzing poly-P synthesis—resulted in a dramatic decrease in Nsr1 mass, more consistent with the protein's expected molecular weight. A subsequent investigation found that poly-P was able to non-enzymatically modify purified Nsr1 and its binding partner Topoisomerase 1 (Top1). Hydrolysis studies with perchloric acid and mutagenesis data indicate that the poly-P polymers are attached to multiple lysine residues in poly-acidic serine and lysine (PASK) sequences present in both of these proteins (Fig. 7). Polyphosphorylation appears to regulate the localization of Nsr1 and Top1 within the nucleus, and was found to inhibit Top1 activity in a supercoiled DNA relaxation assay. Interestingly, $(P)_n$ Lys is indirectly regulated by the levels of inositol pyrophosphates, which modulate poly-P synthesis [5, 128]. No techniques exist for the controlled synthesis of polyphosphorylated proteins, let alone inorganic polyphosphate of uniform length. The ability to synthesize site-selectively polyphosphorylated proteins could allow the elucidation of the specific changes that polyphosphate has on protein structure and function. Likewise, synthesis of poly-P with carefully controlled chain lengths or specific tags could facilitate the high-throughput identification of other polyphosphorylated proteins, giving a more comprehensive picture of this modification's role in cell physiology.

9 Outlook

Reversible protein phosphorylation is one of the most frequently encountered PTMs and confers great diversity on the proteome. This diversity is generated by the spatially and temporally controlled activity of specific protein kinases and phosphatases, the expanding set of amino acids capable of undergoing

phosphorylation, and potentially the opportunity to add more than one phosphoryl group to a single amino acid side chain.

In eukaryotes in particular, most of our attention has focused on just three types of modifications; pSer, pThr, and pTyr. To decipher their role within cellular signaling networks, a large set of specific tools first had to be developed. These tools included synthetic access to site-specifically modified phosphopeptides to elucidate the stability of these modifications, as well as their fragmentation patterns during mass spectrometric analysis. Furthermore, these phosphopeptides enabled the development of specific enrichment strategies of phosphopeptides and proteins, for subsequent proteomic analysis, or antibody-based detection. Only with these tools in hand has it been possible to now routinely interrogate specific signaling pathways, and to annotate the dynamic behavior of these signaling pathways within complex cellular settings.

In contrast, the presence of phosphorylated basic amino-acid residues in proteins is often overlooked because of the short lifetime and acid lability of the phosphoramidate bond. Likewise, phosphorylated acidic amino-acids and pyro- and polyphosphorylated side chains pose a significant analytical challenge. For these more labile modifications, dedicated tools, such as clean peptide standards, or specific antibodies are sparse, and these modifications typically remain poorly characterized. This neglect has been recognized over the past few years though, and new chemical and biochemical approaches have surfaced that are deemed to make the study of labile-amino acid phosphorylation more accessible. In this chapter, we have summarized some of these approaches, such as new chemical methodology to obtain site-specifically modified peptides, and the design of stabilized analogs with the goal to generate pan-specific antibodies. Application of these tools now offers the possibility to annotate the more exotic phosphorylation events, and to determine how widely used labile phosphorylation is within cellular signaling pathways.

References

1. Hunter T (2012) Why nature chose phosphate to modify proteins. *Philos Trans R Soc B Biol Sci* 367:2513–2516
2. Zhang J, Yang PL, Gray NS (2009) Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* 9:28–39
3. Freschi L, Osseni M, Landry CR (2014) Functional divergence and evolutionary turnover in mammalian phosphoproteomes. *PLoS Genet* 10:e1004062
4. Bhandari R, Saiardi A, Ahmadibeni Y, Snowman AM, Resnick AC, Kristiansen TZ, Molina H, Pandey A, Werner JK, Juluri KR, Xu Y, Prestwich GD, Parang K, Snyder SH (2007) Protein pyrophosphorylation by inositol pyrophosphates is a posttranslational event. *Proc Natl Acad Sci USA* 104:15305–15310
5. Azevedo C, Livermore T, Saiardi A (2015) Protein polyphosphorylation of lysine residues by inorganic polyphosphate. *Mol Cell* 58:71–82
6. Attwood PV, Piggott MJ, Zu XL, Besant PG (2007) Focus on phosphohistidine. *Amino Acids* 32:145–156
7. Besant PG, Attwood PV, Piggott MJ (2009) Focus on phosphorarginine and phospholysine. *Curr Protein Pept Sci* 10:536–550

8. Attwood PV, Besant PG, Piggott MJ (2011) Focus on phosphoaspartate and phosphoglutamate. *Amino Acids* 40:1035–1051
9. Boyer PD, DeLuca M, Ebner KE, Hultquist DE, Peter JB (1962) Identification of phosphohistidine digests from a probable intermediate of oxidative phosphorylation. *J Biol Chem* 237:PC3306–PC3308
10. Perry J, Koteva K, Wright G (2011) Receptor domains of two-component signal transduction systems. *Mol BioSyst* 7:1388–1398
11. Khorchid A, Ikura M (2006) Bacterial histidine kinase as signal sensor and transducer. *Int J Biochem Cell Biol* 38:307–312
12. Neiditch MB, Federle MJ, Miller ST, Bassler BL, Hughson FM (2005) Regulation of LuxPQ receptor activity by the quorum-sensing signal autoinducer-2. *Mol Cell* 18:507–518
13. Mayville P, Ji G, Beavis R, Yang H, Goger M, Novick RP, Muir TW (1999) Structure-activity analysis of synthetic autoinducing thiolactone peptides from *Staphylococcus aureus* responsible for virulence. *Proc Natl Acad Sci USA* 96:1218–1223
14. Koteva K, Hong H-J, Wang XD, Nazi I, Hughes D, Naldrett MJ, Buttner MJ, Wright GD (2010) A vancomycin photoprobe identifies the histidine kinase VanSsc as a vancomycin receptor. *Nat Chem Biol* 6:327–329
15. Chen CC, Smith DL, Bruegger BB, Halpern RM, Smith RA (1974) Occurrence and distribution of acid-labile histone phosphates in regenerating rat liver. *Biochemistry* 13:3785–3789
16. Kee JM, Villani B, Carpenter LR, Muir TW (2010) Development of stable phosphohistidine analogues. *J Am Chem Soc* 132:14327–14329
17. Kee JM, Muir TW (2012) Chasing phosphohistidine, an elusive sibling in the phosphoamino acid family. *ACS Chem Biol* 7:44–51
18. Morera S, Chiadmi M, LeBras G, Lascu I, Janin J (1995) Mechanism of phosphate transfer by nucleoside diphosphate kinase: X-ray structures of the phosphohistidine intermediate of the enzymes from *Drosophila* and *Dictyostelium*. *Biochemistry* 34:11062–11070
19. Bond CS, White MF, Hunter WN (2001) High resolution structure of the phosphohistidine-activated form of *Escherichia coli* cofactor-dependent phosphoglycerate mutase. *J Biol Chem* 276:3247–3253
20. Hultquist DE (1968) The preparation and characterization of phosphorylated derivatives of histidine. *Biochim Biophys Acta Bioenerg* 153:329–340
21. Duclos B, Marcandier S, Cozzone AJ (1991) Chemical properties and separation of phosphoamino acids by thin-layer chromatography and/or electrophoresis. *Methods Enzymol* 201:10–21
22. Lloyd GJ, Cooperman BS (1971) Nucleophilic attack by zinc (II)-pyridine-2-carbaldoxime anion on phosphorylimidazole. A model for enzymatic phosphate transfer. *J Am Chem Soc* 93:4883–4889
23. Hohenester UM, Ludwig K, König S (2013) Chemical phosphorylation of histidine residues in proteins using potassium phosphoramidate—a tool for the analysis of acid-labile phosphorylation. *Curr Drug Deliv* 10:58–63
24. Medzihradzsky KF, Philipps NJ, Senderowicz L, Wang P, Turck CW (1997) Synthesis and characterization of histidine-phosphorylated peptides. *Protein Sci* 6:1405–1411
25. Attwood PV, Ludwig K, Bergander K, Besant PG, Adina-Zada A, Kriegelstein J, Klumpp S (2010) Chemical phosphorylation of histidine-containing peptides based on the sequence of histone H4 and their dephosphorylation by protein histidine phosphatase. *Biochim Biophys Acta Proteins Proteomics* 1804:199–205
26. Gustafson C, Wagner-Jauregg T (1954) Phosphorimidazole and phosphohistidine. *Fed Proc* 13:222
27. Rosenberg TH (1964) A simple preparation method for diphosphoimidazole. *Arch Biochem Biophys* 105:315–318
28. Muller T, Rathlev T, Rosenberg T (1956) Special cases of non-enzymic transphosphorylation. *Biochim Biophys Acta* 19:563–564
29. Wagner-Jauregg T, Hackley BE (1953) Model reactions of phosphorus-containing enzyme inactivators. III. Interaction of imidazole, pyridine, and some of their derivatives with dialkyl halogeno-phosphates. *J Am Chem Soc* 75:2125–2130
30. Hultquist DE, Moyer RW, Boyer PD (1966) The preparation and characterization of 1-phosphohistidine and 3-phosphohistidine. *Biochemistry* 5:322–331
31. Wei Y-F, Matthews HR (1991) Identification of phosphohistidine in proteins and purification of protein-histidine kinases. *Methods Enzymol* 200:388–414
32. Kowalewska K, Stefanowicz P, Ruman T, Frączyk T, Rode W, Szewczuk Z (2010) Electron capture dissociation mass spectrometric analysis of lysine-phosphorylated peptides. *Biosci Rep* 30:433–443

33. Kee J-M, Oslund RC, Perlman DH, Muir TW (2013) A pan-specific antibody for direct detection of protein histidine phosphorylation. *Nat Chem Biol* 9:416–421
34. Oslund RC, Kee JM, Couvillon AD, Bhatia VN, Perlman DH, Muir TW (2014) A phosphohistidine proteomics strategy based on elucidation of a unique gas-phase phosphopeptide fragmentation mechanism. *J Am Chem Soc* 136:12899–12911
35. Schenkels C, Erni B, Reymond J-L (1999) Phosphofurylalanine, a stable analog of phosphohistidine. *Bioorganic Med Chem Lett* 9:1443–1446
36. McAllister TE, Webb ME (2012) Triazole phosphohistidine analogues compatible with the Fmoc-strategy. *Org Biomol Chem* 10:4043–4049
37. Kee JM, Oslund RC, Couvillon AD, Muir TW (2015) A second-generation phosphohistidine analog for production of phosphohistidine antibodies. *Org Lett* 17:187–189
38. Lilley MB, Mambwe B, Thompson MJ, Jackson RFW, Muimo R (2015) 4-Phosphopyrazol-2-yl alanine: a non-hydrolysable analogue of phosphohistidine. *Chem Commun* 51:7305–7308
39. Fuhs SR, Meisenhelder J, Aslanian A, Ma L, Zagorska A, Stankova M, Binnie A, Al-Obeidi F, Mauger J, Lemke G, Yates JR, Hunter T (2015) Monoclonal 1- and 3-phosphohistidine antibodies: new tools to study histidine phosphorylation. *Cell* 162:198–210
40. Ennor AH, Morrison JF (1958) Biochemistry of the phosphagens and related guanidines. *Physiol Rev* 36:631–674
41. Cieřla J, Fraczyk T, Rode W (2011) Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta Biochim Pol* 58:137–148
42. Fuhrmann J, Clancy KW, Thompson PR (2015) Chemical biology of protein arginine modifications in epigenetic regulation. *Chem Rev* 115:5413–5461
43. Fuhrmann J, Schmidt A, Spiess S, Lehner A, Turgay K, Mechtler K, Charpentier E, Clausen T (2009) McsB is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR. *Science* 324:1323–1327
44. Elsholz AKW, Turgay K, Michalik S, Hessling B, Gronau K, Oertel D, Mader U, Bernhardt J, Becher D, Hecker M, Gerth U (2012) Global impact of protein arginine phosphorylation on the physiology of *Bacillus subtilis*. *Proc Natl Acad Sci USA* 109:7451–7456
45. Schmidt A, Trentini DB, Spiess S, Fuhrmann J, Ammerer G, Mechtler K, Clausen T (2014) Quantitative phosphoproteomics reveals the role of protein arginine phosphorylation in the bacterial stress response. *Mol Cell Proteomics* 13:537–550
46. Mijakovic I, Grangeasse C, Turgay K (2016) Exploring the diversity of protein modifications: special bacterial phosphorylation systems. *FEMS Microbiol Rev* 40:398–417
47. Wozniak DJ, Tiwari KB, Soufan R, Jayaswal RK (2012) The mcsB gene of the clpC operon is required for stress tolerance and virulence in *Staphylococcus aureus*. *Microbiology* 158:2568–2576
48. Trentini DB, Suskiewicz MJ, Heuck A, Kurzbauer R, Deszcz L, Mechtler K, Clausen T (2016) Arginine phosphorylation marks proteins for degradation by a Clp protease. *Nature* 539:48–53
49. Fujitaki JM, Smith RA (1984) Techniques in the detection and characterization of phosphoramidate-containing proteins. *Methods Enzymol* 107:23–36
50. Ellington WR (2001) Evolution and physiological roles of phosphagen systems. *Annu Rev Physiol* 63:289–325
51. Trentini DB, Fuhrmann J, Mechtler K, Clausen T (2014) Chasing phosphoarginine proteins: development of a selective enrichment method using a phosphatase trap. *Mol Cell Proteomics* mcp.O113.035790
52. Schmidt A, Ammerer G, Mechtler K (2013) Studying the fragmentation behavior of peptides with arginine phosphorylation and its influence on phospho-site localization. *Proteomics* 13:945–954
53. Marcus F, Morrison JF (1964) The preparation of phosphoarginine: a comparative study. *Biochem J* 92:429–435
54. Cramer F, Scheiffele E, Vollmar A (1962) Die synthese der argininphosphorsäure und die reaktion von isoureidophosphonaten mit aminen. *Chem Ber* 95:1670–1682
55. Kumon A, Yokoi F, Hiraishi H (1996) *N*-phosphoarginine phosphatase (17 kDa) and alkaline phosphatase as protein arginine phosphatases. *J Biochem* 119:719–724
56. Hofmann FT, Lindemann C, Salia H, Adamitzki P, Karanicolas J, Seebeck FP (2011) A phosphoarginine containing peptide as an artificial SH2 ligand. *Chem Commun (Camb)* 47:10335–10337
57. Fuhrmann J, Mierzwa B, Trentini DB, Spiess S, Lehner A, Charpentier E, Clausen T (2013) Structural basis for recognizing phosphoarginine and evolving residue-specific protein phosphatases in Gram-positive bacteria. *Cell Rep* 3:1832–1839


58. Fuhrmann J, Subramanian V, Thompson PR (2015) Synthesis and use of a phosphonate amidine to generate an anti-phosphoarginine-specific antibody. *Angew Chemie Int Ed* 54:14715–14718
59. Ouyang H, Fu C, Fu S, Ji Z, Sun Y, Deng P, Zhao Y (2016) Development of a stable phosphoarginine analog for producing phosphoarginine antibodies. *Org Biomol Chem* 14:1925–1929
60. Zetterqvist Ö, Engström L (1967) Isolation of *N*-ε-[32P]phosphoryl-lysine from rat-liver cell sap after incubation with [32P] adenosine triphosphate. *Biochem Biophys Acta* 141:523–532
61. Wälinder O (1968) Identification of a phosphate-incorporating protein from bovine liver as nucleoside identification of a phosphate-incorporating protein from bovine liver as nucleoside diphosphate kinase and isolation of and *N*-ε-32-*P*-phospholysine from erythrocytic nucleol. *J Biol Chem* 243:3947–3952
62. Chen CC, Bruegger BB, Kern CW, Lin YC, Halpern RM, Smith RA (1977) Phosphorylation of nuclear proteins in rat regenerating liver. *Biochemistry* 16:4852–4855
63. Smith DL, Bruegger BB, Halpern RM, Smith RA (1973) New histone kinases in nuclei of rat tissues. *Nature* 246:103–104
64. Postel EH, Abramczyk BM, Levit MN, Kyin S (2000) Catalysis of DNA cleavage and nucleoside triphosphate synthesis by NM23-H2/NDP kinase share an active site that implies a DNA repair function. *Proc Natl Acad Sci USA* 97:14194–14199
65. Ohmori H, Kuba M, Kumon A (1994) 3-Phosphohistidine/6-phospholysine phosphatase from rat brain as. *J Biochem* 116:380–385
66. Benkovic SJ, Sampson EJ (1971) Structure-reactivity correlation for the hydrolysis of phosphoramidate monoanions. *J Am Chem Soc* 93:4009–4016
67. Modro TA (1981) Phosphoric and carboxylic amides. *ACS Symp Ser* 171:619–622
68. Denehy E, White JM, Williams SJ (2007) Electronic structure of the sulfonyl and phosphoryl groups: a computational and crystallographic study. *Inorg Chem* 46:8871–8886
69. Bertran-Vicente J, Serwa RA, Schümann M, Schmieder P, Krause E, Hackenberger CPR (2014) Site-specifically phosphorylated lysine peptides. *J Am Chem Soc* 136:13622–13628
70. Fujitaki JM, Steiner AW, Nichols SE, Helander ER, Lin YC, Smith RA (1980) A simple preparation of *N*-phosphorylated lysine and arginine. *Prep Biochem* 10:205–213
71. Bertran-Vicente J, Schumann M, Schmieder P, Krause E, Hackenberger CPR (2015) Direct access to site-specifically phosphorylated-lysine peptides from a solid-support. *Org Biomol Chem* 13:6839–6843
72. Pas HH, Robillard GT (1988) S-Phosphocysteine and phosphohistidine are intermediates in the phosphoenolpyruvate-dependent mannitol transport catalyzed by *Escherichia coli* EIIMtl. *Biochemistry* 27:5835–5839
73. Cho H, Krishnaraj R, Kitas E, Bannwarth W, Walsh CT, Anderson KS (1992) Isolation and structural elucidation of a novel phosphocysteine intermediate in the LAR protein tyrosine phosphatase enzymic pathway. *J Am Chem Soc* 114:7296–7298
74. Brandão TAS, Hengge AC, Johnson SJ (2010) Insights into the reaction of protein-tyrosine phosphatase 1B: crystal structures for transition state analogs of both catalytic steps. *J Biol Chem* 285:15874–15883
75. Asthagiri D, Liu T, Noodleman L, Van Etten RL, Bashford D (2004) On the role of the conserved aspartate in the hydrolysis of the phosphocysteine intermediate of the low molecular weight tyrosine phosphatase. *J Am Chem Soc* 126:12677–12684. doi:10.1021/JA048638O
76. Pas HH, Meyer GH, Kruizinga WH, Tamminga KS, van Weeghel RP, Robillard GT (1991) 31phospho-NMR demonstration of phosphocysteine as a catalytic intermediate on the *Escherichia coli* phosphotransferase system EIIMtl. *J Biol Chem* 266:6690–6692
77. Meins M, Jenö P, Müller D, Richter WJ, Rosenbusch JP, Erni B (1993) Cysteine phosphorylation of the glucose transporter of *Escherichia coli*. *J Biol Chem* 268:11604–11609
78. Sun F, Ding Y, Ji Q, Liang Z, Deng X, Wong CCL, Yi C, Zhang L, Xie S, Alvarez S, Hicks LM, Luo C, Jiang H, Lan L, He C (2012) Protein cysteine phosphorylation of SarA/MgrA family transcriptional regulators mediates bacterial virulence and antibiotic resistance. *Proc Natl Acad Sci USA* 109:15461–15466
79. Buchowiecka AK (2014) Puzzling over protein cysteine phosphorylation—assessment of proteomic tools for S-phosphorylation profiling. *Analyst* 139:4118–4123
80. Chalker JM, Lercher L, Rose NR, Schofield CJ, Davis BG (2012) Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. *Angew Chemie Int Ed* 51:1835–1839

81. Chooi KP, Galan SRG, Raj R, McCullagh JSO, Mohammed S, Jones LH, Davis BG (2014) Synthetic phosphorylation of p38# recapitulates protein kinase activity. *J Am Chem Soc* 136:1698–1701
82. Rowan FC, Richards M, Bibby RA, Thompson A, Bayliss R, Blagg J (2013) Insights into aurora-a kinase activation using unnatural amino acids incorporated by chemical modification. *ACS Chem Biol* 8:2184–2191
83. Åkerfeldt S, Willman N-E, Berggren B, Thomelius H, Westin G (1960) Cysteamine S-phosphoric acid. *Acta Chem Scand* 14:1980–1984
84. Åkerfeldt S, Hasselquist H, Prange I, Dam H, Sjöberg B, Toft J (1961) Further studies on S-substituted phosphorothioic acids. Mixed lithiumsodium salts of S-(1-carboxyethyl) phosphorothioic acid and S-(2-carboxyethyl) phosphorothioic acid. *Acta Chem Scand* 15:575–582
85. Åkerfeldt S, Weidler A-M, Mandell L, Kvande PC, Meisinger E (1963) Further studies on S-substituted phosphorothioic acids. III. Rates of hydrolysis and dissociation constants. *Acta Chem Scand* 17:319–328
86. Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JEP, Coon JJ (2007) Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* 79:477–485
87. Bertran-Vicente J, Penkert M, Nieto-Garcia O, Jeckelmann J-M, Schmieder P, Krause E, Hackenberger CPR (2016) Chemoselective synthesis and analysis of naturally occurring phosphorylated cysteine peptides. *Nat Commun* 7:12703
88. Ruman T, Długopolska K, Jurkiewicz A, Rut D, Fraczyk T, Cieśla J, Leś A, Szewczuk Z, Rode W (2010) Thiophosphorylation of free amino acids and enzyme protein by thiophosphoramidate ions. *Bioorg Chem* 38:74–80
89. Bernardes GJL, Chalker JM, Errey JC, Davis BG (2008) Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: versatile and switchable access to functionalized proteins. *J Am Chem Soc* 130:5052–5053
90. Chalker JM, Gunnoo SB, Boutureira O, Gerstberger SC, Fernández-González M, Bernardes GJL, Griffin L, Hailu H, Schofield CJ, Davis BG (2011) Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem Sci* 2:1666
91. Chen Z, Cole PA (2015) Synthetic approaches to protein phosphorylation. *Curr Opin Chem Biol* 28:115–122
92. Ridder IS, Dijkstra BW (1999) Identification of the Mg^{2+} -binding site in the P-type ATPase and phosphatase members of the HAD (haloacid dehalogenase) superfamily by structural similarity to the response regulator protein CheY. *Biochem J* 339:223–226
93. Allen KN, Dunaway-Mariano D (2004) Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem Sci* 29:495–503
94. Lee SY, Cho HS, Pelton JG, Yan D, Henderson RK, King DS, Huang L, Kustu S, Berry EA, Wemmer DE (2001) Crystal structure of an activated response regulator bound to its target. *Nat Struct Biol* 8:52–56
95. Lee SY, Cho HS, Pelton JG, Yan D, Berry EA, Wemmer DE (2001) Crystal structure of activated CheY. Comparison with other activated receiver domains. *J Biol Chem* 276:16425–16431
96. Zhao R, Collins EJ, Bourret RB, Silversmith RE (2002) Structure and catalytic mechanism of the *E. coli* chemotaxis phosphatase CheZ. *Nat Struct Biol* 9:570–575
97. Ibrahim IM, Puthiyaveetil S, Allen JF (2016) A two-component regulatory system in transcriptional control of photosystem stoichiometry: redox-dependent and sodium ion-dependent phosphoryl transfer from cyanobacterial histidine kinase Hik2 to response regulators Rre1 and RppA. *Front Plant Sci* 7:137
98. Trumbore MW, Wang RH, Enkemann SA, Berger SL (1997) Prothymosin alpha in vivo contains phosphorylated glutamic acid residues. *J Biol Chem* 272:26394–26404
99. Wang RH, Tao L, Trumbore MW, Berger SL (1997) Turnover of the acyl phosphates of human and murine prothymosin alpha in vivo. *J Biol Chem* 272:26405–26412
100. Tao L, Wang RH, Enkemann SA, Trumbore MW, Berger SL (1999) Metabolic regulation of protein-bound glutamyl phosphates: insights into the function of prothymosin alpha. *J Cell Physiol* 178:154–163
101. Koshland DE (1952) Effect of catalysts on the hydrolysis of acetyl phosphate. nucleophilic displacement mechanisms in enzymatic reactions. *J Am Chem Soc* 74:2286–2292
102. Purich DL (2002) Use of sodium borohydride to detect acyl-phosphate linkages in enzyme reactions. *Methods Enzymol* 354:168–177

103. Andersson J, Barth A (2006) FTIR studies on the bond properties of the aspartyl phosphate moiety of the Ca^{2+} -ATPase. *Biopolymers* 82:353–357
104. Schlemmer H, Sontheimer GM, Kalbitzer HR (1988) ^{31}P nuclear magnetic resonance spectroscopy of the phosphorylated tetrapeptide Gly–Gly–Asp–Ala. *Magn Reson Chem* 26:260–263
105. Platzer G, Okon M, McIntosh LP (2014) pH-dependent random coil ^1H , ^{13}C , and ^{15}N chemical shifts of the ionizable amino acids: a guide for protein pKa measurements. *J Biomol NMR* 60:109–129
106. Black S, Wright NG (1953) Enzymatic reduction of β -aspartylphosphate to homoserine. *J Am Chem Soc* 75:5766
107. Wright G, Black S, Wright NG (1955) β -Aspartokinase and β -aspartyl phosphate. *J Biol Chem* 213:27–38
108. Katchalsky A, Paecht M (1954) Phosphate anhydrides of amino acids. *J Am Chem Soc* 76:6042–6044
109. Saxl RL, Anand GS, Stock AM (2001) Synthesis and biochemical characterization of a phosphorylated analogue of the response regulator CheB. *Biochemistry* 40:12896–12903
110. Stewart RC (1993) Activating and inhibitory mutations in the regulatory domain of CheB, the methylesterase in bacterial chemotaxis. *J Biol Chem* 268:1921–1930
111. Berlicki Ł (2008) Inhibitors of glutamine synthetase and their potential application in medicine. *Mini Rev Med Chem* 8:869–878
112. Colquhoun A, Newsholme E (1997) Aspects of glutamine metabolism in human tumour cells. *IUBMB Life* 41:583–596
113. Saiardi A, Bhandari R, Resnick AC, Snowman AM, Snyder SH (2004) Phosphorylation of proteins by inositol pyrophosphates. *Science* 306:2101–2105
114. Chakraborty A, Koldobskiy MA, Bello NT, Maxwell M, Potter JJ, Juluri KR, Maag D, Kim S, Huang AS, Dailey MJ, Saleh M, Snowman AM, Moran TH, Mezey E, Snyder SH (2010) Inositol pyrophosphates inhibit akt signaling, thereby regulating insulin sensitivity and weight gain. *Cell* 143:897–910
115. Thota SG, Unnikannan CP, Thampatty SR, Manorama R, Bhandari R (2015) Inositol pyrophosphates regulate RNA polymerase I-mediated rRNA transcription in *Saccharomyces cerevisiae*. *Biochem J* 466:105–114
116. Voglmaier SM, Bembenek ME, Kaplin AI, Dormán G, Olszewski JD, Prestwich GD, Snyder SH (1996) Purified inositol hexakisphosphate kinase is an ATP synthase: diphosphoinositol pentakisphosphate as a high-energy phosphate donor. *Proc Natl Acad Sci USA* 93:4305–4310
117. Brown NW, Marmelstein AM, Fiedler D (2016) Chemical tools for interrogating inositol pyrophosphate structure and function. *Chem Soc Rev* 45:6311–6326
118. Shears SB (2015) Inositol pyrophosphates: why so many phosphates? *Adv Biol Regul* 57:203–216
119. Marmelstein AM, Yates LM, Conway JH, Fiedler D (2014) Chemical pyrophosphorylation of functionally diverse peptides. *J Am Chem Soc* 136:108–111
120. Yates LM, Fiedler D (2015) Establishing the stability and reversibility of protein pyrophosphorylation with synthetic peptides. *ChemBioChem* 16:415–423
121. Conway JH, Fiedler D (2015) An affinity reagent for the recognition of pyrophosphorylated peptides. *Angew Chemie Int Ed* 54:3941–3945
122. Williams FJ, Fiedler D (2015) A fluorescent sensor and gel stain for detection of pyrophosphorylated proteins. *ACS Chem Biol* 10:1958–1963
123. Yates LM, Fiedler D (2016) A stable pyrophosphoserine analog for incorporation into peptides and proteins. *ACS Chem Biol* 11:1066–1073
124. Kornberg A, Rao NN, Ault-riché D (1999) Inorganic polyphosphate: a molecule of many functions. *Annu Rev Biochem* 68:89–125
125. Gray MJ, Wholey WY, Wagner NO, Cremers CM, Mueller-Schickert A, Hock NT, Krieger AG, Smith EM, Bender RA, Bardwell JCA, Jakob U (2014) Polyphosphate is a primordial chaperone. *Mol Cell* 53:689–699
126. Holmström KM, Marina N, Baev AY, Wood NW, Gourine AV, Abramov AY (2013) Signalling properties of inorganic polyphosphate in the mammalian brain. *Nat Commun* 4:1362
127. Cremers CM, Knoefler D, Gates S, Galvan V, Southworth DR, Jakob U, Cremers CM, Knoefler D, Gates S, Martin N, Dahl J, Lempart J, Xie L, Chapman MR, Galvan V, Southworth DR, Jakob U (2016) Polyphosphate: a conserved modifier of amyloidogenic processes. *Mol Cell* 63:1–13

128. Wild R, Gerasimaite R, Jung J-Y, Truffault V, Pavlovic I, Schmidt A, Saiardi A, Jessen HJ, Poirier Y, Hothorn M, Mayer A (2016) Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains. *Science* 352:986–990

Applications of Phosphate Modification and Labeling to Study (m)RNA Caps

Marcin Warminski¹ · Pawel J. Sikorski² ·
Joanna Kowalska¹ · Jacek Jemielity² 

Received: 17 November 2016 / Accepted: 10 January 2017 / Published online: 23 January 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract The cap is a natural modification present at the 5' ends of eukaryotic messenger RNA (mRNA), which because of its unique structural features, mediates essential biological functions during the process of gene expression. The core structural feature of the mRNA cap is an N7-methylguanosine moiety linked by a 5'–5' triphosphate chain to the first transcribed nucleotide. Interestingly, other RNA 5' end modifications structurally and functionally resembling the m⁷G cap have been discovered in different RNA types and in different organisms. All these structures contain the 'inverted' 5'–5' oligophosphate bridge, which is necessary for interaction with specific proteins and also serves as a cleavage site for phosphohydrolases regulating RNA turnover. Therefore, cap analogs containing oligophosphate chain modifications or carrying spectroscopic labels attached to phosphate moieties serve as attractive molecular tools for studies on RNA metabolism and modification of natural RNA properties. Here, we review chemical, enzymatic, and chemoenzymatic approaches that enable preparation of modified cap structures and RNAs carrying such structures, with emphasis on phosphate-modified mRNA cap analogs and their potential applications.

M. Warminski and P. J. Sikorski have contributed equally to this work.

This article is part of the Topical Collection "Phosphate Labeling in Chemical Biology"; edited by Henning Jessen.

✉ Joanna Kowalska
asia@biogeo.uw.edu.pl

✉ Jacek Jemielity
jacekj@biogeo.uw.edu.pl

¹ Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Zwirki i Wigury 93, 02-089 Warsaw, Poland

² Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland

Keywords RNA labeling · Capping · Molecular probe · Nucleotide · Cap analog · 7-methylguanosine

1 RNA Cap Structures and Their Functions

1.1 7-Methylguanosine mRNA Cap

The existence of methylated guanine joined by a 5′–5′-triphosphate linker to the first transcribed nucleotide of eukaryotic and viral messenger RNA (mRNA) was first reported in the mid-1970s independently by Wei [1] and Furuichi [2] (Fig. 1). Since then, cap has emerged as a master regulator of several processes contributing to gene expression in eukaryotes. Besides its two main cytoplasmic roles, i.e., protection of the mRNA 5′ end from premature degradation by exonucleases and participation in the initiation of protein synthesis (translation), cap is also required for proper splicing of precursor mRNA (pre-mRNA) in the nucleus, engaged in cleavage and polyadenylation of pre-mRNA, and mRNA export (Fig. 2) [3–5].

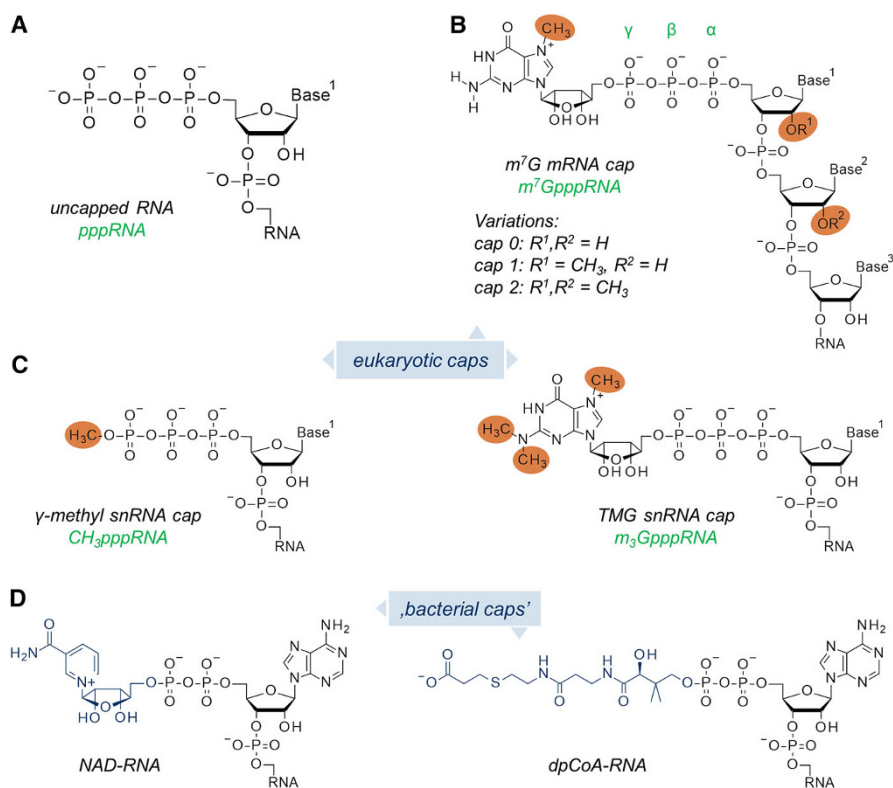


Fig. 1 Structures of different RNA 5′ ends: **a** uncapped (5′-triphosphate) RNA; **b** monomethylguanosine (m^7G) mRNA cap; **c** trimethylguanosine (m_3G) cap and γ -*O*-methyl cap found in snRNAs; **d** examples of ‘cap-like’ structures found on bacterial RNAs

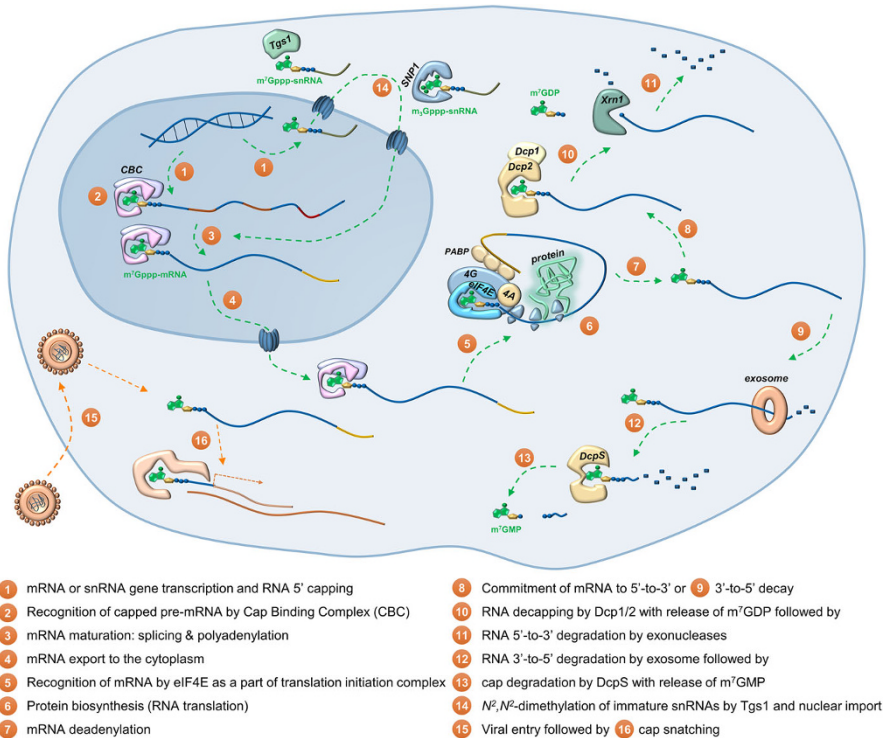


Fig. 2 An overview on RNA cap functions in gene expression and main proteins involved in cap recognition and metabolism

Therefore, it is not surprising that cap formation is the first co-transcriptional event of a nascent RNA synthesized by RNA polymerase II (RNAP II), which occurs after synthesis of 20–40 nucleotide long transcripts [6]. Many viruses that infect eukaryotes have evolved mechanisms to either synthesize their own cap structures or salvage cap moieties from host mRNA through so-called cap snatching to maximize the chances for their own mRNA survival and expression [7]. The protective function of mRNA cap relies on the fact that cytoplasmic 5'–3' exoribonucleases, such as Xrn1, do not recognize capped mRNAs as substrates. Therefore, mRNA degradation initiated from the 5' end must be preceded by cap removal by specialized decapping enzymes. Another class of decapping enzymes is involved in degradation of free cap structures generated after 3'–5' mRNA decay (Fig. 2).

Synthetic cap analogs are useful research tools that facilitate investigation of cap-related processes both at a molecular level and in biochemical and biological experiments, which may eventually lead even to development of novel therapeutic approaches [8]. In this paper, we review approaches employed to develop cap-based molecular tools that could benefit biological studies, with the focus on phosphate-labeled and phosphate-modified cap analogs of m⁷G cap and other structurally related RNA cap structures.

1.2 m⁷G Cap 0, 1, 2 and Other Cap Structures

Although the term cap usually refers to the m⁷G moiety linked to the 5' end of a transcript, different versions of the m⁷G cap have been recognized along with other types of cap structures found on different RNA types and in various organisms. Depending on the methylation pattern of the first few 5'-nucleotides in an m⁷G-capped RNA body, cap variants referred to as cap 0, cap 1, cap 2, and so on, can be distinguished (Fig. 1b). In addition, other types of methylated cap structures have been identified in eukaryotes, trimethylguanosine and γ -methyl cap being the prime examples (Fig. 1c). Finally, it was recently discovered that prokaryotes also synthesize conjugates of RNA with small molecules that structurally resemble eukaryotic 5' caps (Fig. 1d).

Cap 0 is m⁷G cap without additional methylations in the RNA body. However, cap 0 can undergo further methylations at the 2'-O position within the ribose of the first or within the first and second nucleotide to produce cap 1 or cap 2, respectively (Fig. 1b) [9]. Generally, cap 0 structures are more common in lower eukaryotes, whereas cap 1 and cap 2 structures are found in higher organisms, including mammals [9]. Additional methylations at the first and second nucleobase of mRNAs can also take place in cap 1 and 2. For a long time, it was unclear why the 5' end of mRNAs would undergo such extensive methylation. Recent studies have revealed that 2'-O methylation of cellular RNA plays a central role in discrimination of self from non-self RNA, e.g., distinction of viral from host RNA [10, 11], while reversible interconversion of N⁶, 2'-O-dimethyladenosine and 2'-O-methyladenosine in the cap controls mRNA stability [12].

Not only nucleotides downstream from m⁷G undergo methylation. In some small nuclear RNAs (snRNAs) involved in splicing of pre-mRNA, the m⁷G moiety undergoes hypermethylation, i.e., addition of two methyl groups at the N² position (Fig. 1c). This structure, called a trimethylguanosine cap (TMG cap or m₃G), is characteristic of a portion of snRNAs transcribed by RNAP II, namely U1, U2, U4, and U5 snRNAs, and small nucleolar RNAs (snoRNAs) engaged in post-transcriptional modification of precursor ribosomal RNAs (pre-rRNAs) [13]. These snRNAs are hypermethylated by trimethylguanosine synthase 1 (TGS1) after being exported to the cytoplasm in association with Sm proteins (Fig. 4) [13, 14]. The presence of a TMG cap allows binding to the transport protein snurportin 1 (SNP1) and the import of fully matured snRNAs back to the nucleus, where they participate in pre-mRNA splicing [14]. Moreover, the TMG cap has been found at the 5' end of a certain pool of mRNAs in nematodes (e.g., *Caenorhabditis elegans*) [15], and most recently at the 5' end of mRNAs encoding selenoproteins in mammals [16].

Interestingly, other snRNAs, such U6 and 7SK, which are synthesized by RNA polymerase III (RNAP III), possess at their 5' end another modification, a methyl group added directly to triphosphate bridge to generate a γ -methyl phosphate cap (Figs. 1c, 4) [17]. However, the presence of a methyl group linked directly to a triphosphate bridge is not a general feature of RNAP III-transcribed RNAs (e.g., tRNAs), as usually the triphosphorylated 5' ends of nascent transcripts are trimmed by nucleases to RNAs without cap-like structures upon maturation.

Although it has been hypothesized for a long time that the presence of the 5' cap is one of the key structural features distinguishing eukaryotic from prokaryotic RNAs, recent mass spectrometry-aided studies on bacterial transcriptomes revealed

that in prokaryotes, the 5' ends of some portion of RNA are modified by moieties, which could be considered as cap-like structures. These include a nicotinamide adenine dinucleotide (NAD^+) [18], 3'-dephospho-coenzyme A (dpCoA) (Fig. 1d) [18], and other moieties attached to the 5' end of RNA by an oligophosphate bridge. The structure, biosynthesis, function, and degradation of these so-called bacterial caps have currently come under intense investigation [19–22].

1.3 Recognition of Cap Structures by Proteins

The complex network of biological processes regulating mRNA expression and turnover is in large extent orchestrated by numerous cap-binding proteins that recognize the 7-methylguanosine triphosphate moiety as a hallmark of the mRNA 5' end. Similarly, the localization, transport, and function of immature and mature sn/snoRNA largely rely on binding by proteins that recognize their m^7G and m_3G caps, respectively. Although these proteins are usually unrelated in terms of function and sequence, they often share similarities in cap recognition modes. Based on the crystal structures and thermodynamic data of numerous cap-binding proteins and enzymes in complex with m^7G or m_3G derivatives, the selectivity of cap recognition could be attributed to its two distinctive features: a positively charged nucleobase and a negatively charged 5'–5'-triphosphate bridge. Interaction patterns that are common in all proteins include (i) cation- π stacking interactions involving 7-methylguanine or 2,2,7-trimethylguanidine and (ii) hydrogen bonds between the oligophosphate bridge and positively charged side chains of basic amino acids assisted by electrostatic attraction of unlike charges (Fig. 3). In the majority of complexes, 7-methylguanine is stacked with the indole moiety of tryptophan residue (eIF4E, SPN1, DcpS, cNIIIB, TGS1) and often forms sandwich complexes employing another side chain of an aromatic amino acid such as Trp, Tyr, Phe (eIF4E, cNIIIB, CBC, VP39) or even the base of the second nucleoside (SPN1) [23–30]. Interestingly, in most of these structures, the N^7 -methyl group is not in direct contact with protein, yet its removal reduces the affinity of m^7GTP to eIF4E by several orders of magnitude, suggesting that the specificity of m^7G recognition is mainly determined by the presence of a positive charge within the nucleobase [24]. Another structural feature influencing the affinity to cap-binding proteins is the oligophosphate chain. Usually, at least two arginine or lysine residues are engaged in hydrogen bonds with the triphosphate bridge, but the geometry of those interactions differs from one protein to another.

2 Chemically and Enzymatically Labeled RNA Caps

2.1 Utility of Labeled Capped RNAs and Cap Analogs

Nowadays, capped RNAs can be easily obtained on a small scale by standard molecular biology techniques and used for the purpose of RNA-driven gene expression of proteins of interest in living cells or cell lysates [31]. Various modified cap analogs are used as small molecule binding and activity probes for cap-binding proteins and cap-processing enzymes. Finally, capped RNAs and small

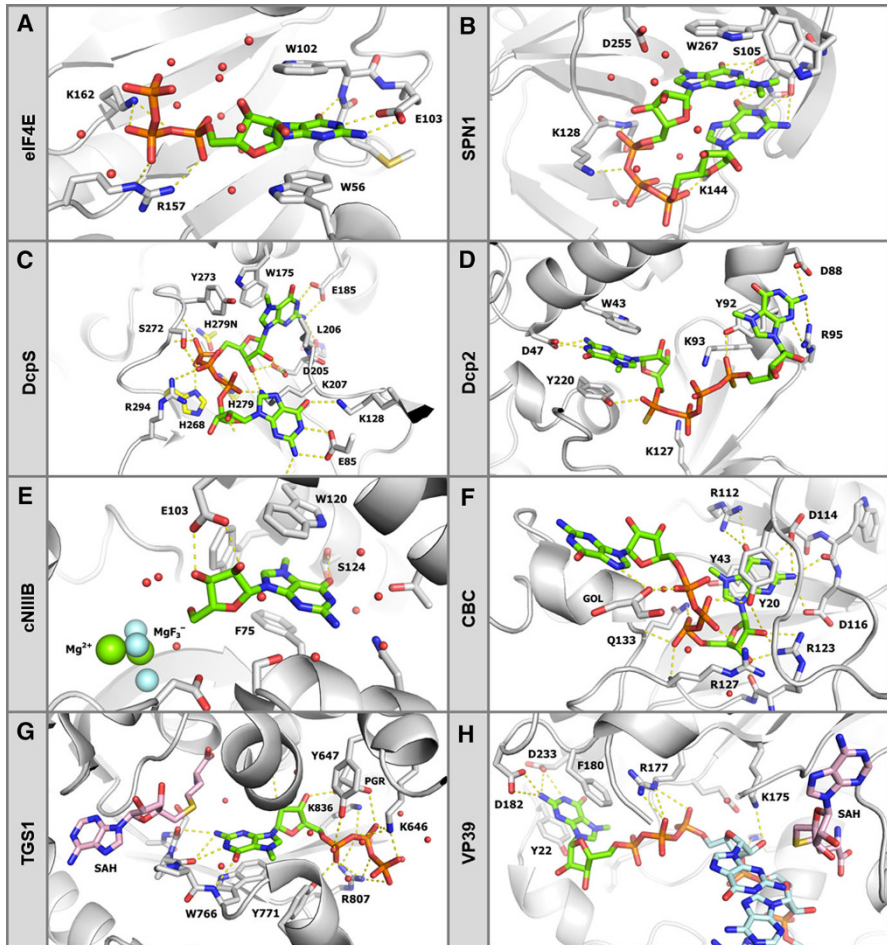


Fig. 3 Recognition of mRNA cap structure by different proteins relies on cation- π stacking of 7-methylguanosine and electrostatic interactions with an oligophosphate chain. **a** Murine translation initiation factor 4E (eIF4E) in complex with m^7 GpppG (pdb entry 1L8B) [24]; **b** cap-binding domain of human snurportin 1 (SPN1) in complex with m_3 GpppG (pdb entry 1XK5) [25]; **c** human decapping scavenger mutant (hDcpS H277N) in complex with substrate, m^7 GpppG (pdb entry 1ST0) [26]; **d** *Saccharomyces cerevisiae* decapping protein 2 (Dcp2) in complex with cap analog, m^7 Gpsppps m^7 G (pdb entry 5KQ4) [27]; **e** *Drosophila melanogaster* cytosolic nucleotidase IIIIB (Ds cNIIIIB) in complex with product, m^7 G, and MgF_3^- as a transition state phosphate mimic (pdb entry 4NV0) [28]; **f** human nuclear cap-binding complex subunit (CBC20) in complex with m^7 GpppG (pdb entry 1N52) [29]; **g** human trimethylguanosine cap synthase 1 (hTGS1) in complex with substrate analog, m^7 GTP and product S-adenosylhomocysteine (SAH) (pdb entry 3GDH) [30]; **h** *Vaccinia* virus 2'-O methylase VP39 in complex with a capped RNA fragment and SAH (pdb entry 1AV6) [23]

molecule cap analogs are utilized to study RNA turnover in various biological systems. For biochemical studies, capped RNAs, cap analogs, and their degradation products isolated from biological mixtures are often resolved by thin layer chromatography (small nucleotides) or electrophoresis (oligos and nucleic acids) and visualized by different methods. Because of low concentrations of RNA and cap

metabolites in typical samples, standard visualization techniques, such as UV shadowing, do not provide sufficient sensitivity nor selectivity in detection, which impairs analysis of complex nucleotide and nucleic acid mixtures. Although fluorescent staining reagents are used to increase sensitivity of nucleic acid detection, they are not useful for detection of short oligos and small nucleotides. Therefore, methods for generation of specifically labeled RNAs carrying radioactive, fluorescent, and other spectroscopic labels are constantly being developed to provide highly sensitive and selective tools for detection of the molecule of interest in a complex mixture or to selectively monitor a particular enzymatic reaction or binding event. In the following sections, we will review enzymatic, chemoenzymatic, and chemical approaches enabling RNA cap labeling, with the main emphasis on modifications of the triphosphate chain and their utility in studying RNA-related processes. We start from the well-established enzymatic approaches, which rely on the use of radioactively labeled Nucleoside triphosphates (NTPs), and RNA-processing enzymes. We then discuss chemical approaches to the synthesis of phosphate-modified or -labeled cap analogs, followed by examples of their utility as molecular probes for studying cap-related processes and reagents for modification of RNA oligonucleotides and long transcripts.

2.2 Cap Biosynthesis

In vivo m^7G capping of a nascent transcript occurs in three sequential steps catalyzed by specific enzymes, namely, RNA triphosphatase (TPase), RNA guanylyltransferase (GTase), and guanine- N^7 methyltransferase (guanine- N^7 MTase) (Fig. 4) [32]. First, RNA TPase removes the 5' γ -phosphate from ppp-RNA to generate RNA 5'-diphosphate. Subsequently, RNA GTase transfers guanine monophosphate (GMP) from GTP to the β -phosphate of RNA to form Gppp-RNA in a reversible two-step ping-pong reaction with lysine-GMP intermediate [33]. The last step of m^7G cap 0 formation is transfer of a methyl group from S-adenosyl-L-methionine (SAM) to the N^7 position of the terminal guanine catalyzed by guanine N^7 MTase [33]. Further optional methylations leading to cap 1 and cap 2 are performed by other enzymes called 2'-*O* methyltransferases (2'-*O* MTases) (Fig. 4). Interestingly, formation of cap 0 and cap 1 takes place in the nucleus, while methylation of cap 1 to produce cap 2 is a cytoplasmic event [34]. The m^7G -capping mechanism dependent on three distinct enzymatic activities is shared from fungi to higher mammals and utilized by some eukaryote-associated viruses. One such virus is *vaccinia*, which encodes its own heterodimeric 127 kDa capping enzyme (*vaccinia* capping enzyme, VCE) that combines all three capping activities necessary for cap 0 biosynthesis. The enzyme was first characterized in 1980 [35], and since then has been extensively studied, and is now a commercially available molecular biology tool widely used to produce capped transcripts, including radioactively labeled ones.

The biosynthesis of other eukaryotic cap structures usually relies on post-transcriptional RNA modification. TMG-capped RNA ($m_3GpppRNA$) is produced by methylation of $m^7GpppRNA$ with TGS1 methyltransferase, which acts in vivo on immature U1, U2, U4, and U5 snRNAs, whereas the γ -methyl cap

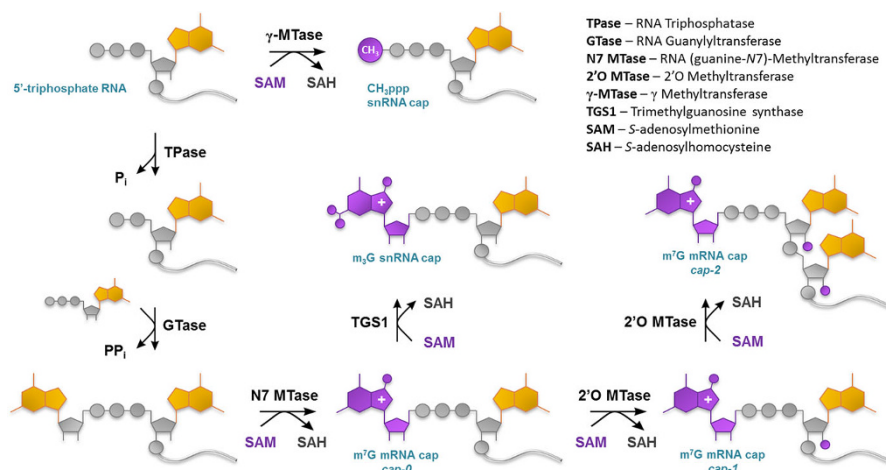


Fig. 4 Enzymatic pathways to generate various eukaryotic cap structures on RNA 5' ends

(CH_3pppRNA) is generated by methylation of U6 5'-triphosphate snRNA with γ -MTase (Fig. 4). Similarly, it was first proposed that biosynthesis of NAD-capped RNAs in bacteria occurs through the post-synthetic addition of a nicotinamide nucleotide to the 5' end of RNA [21, 32], but recent studies indicate that this kind of 'capping' is rather carried out by bacterial RNA polymerases, which use NAD as a transcription initiator [34].

2.3 Possible Sites for Cap Modification

From the bioorganic chemist's point of view, the structural studies provide insights not only into the recognition mode of natural cap structures but also hints at the design of chemically modified cap analog-based molecular tools tailored for specific applications. In general, modifications of the cap should serve at least one of three purposes: (i) to perturb or even completely destroy biological activity (e.g., inhibition of degradation by decapping enzymes); (ii) to increase affinity to particular cap-binding proteins, thereby augmenting associated biological activity; (iii) to confer a new property to the cap structure without any interference with its biological activity (biorthogonal modifications).

Sites for biorthogonal modifications of the cap could be envisaged based on their crystal structure in complex with a targeted cap-binding protein (Fig. 3). For example, in the case of eIF4E complex, both m^7G hydroxyl groups and the second nucleoside are exposed to the solvent, and, in fact, these groups were successfully functionalized with various substituents without significant decrease in affinity constant. Consequently, the translational activity of mRNAs carrying such cap structures was at least retained [36–39]. In contrast, ribose of m^7G in complex with DcpS is engaged in numerous contacts and any modification within 2'-O and 3'-O positions results in a considerable decrease of the rate of hydrolysis catalyzed by DcpS [40, 41]. Importantly, the triphosphate bridge appears to be an attractive site

for modification, especially in the context of conferring resistance to decapping enzymes. Because the decapping enzymes involved in 5′–3′ and 3′–5′ decay cleave the triphosphate chain at different positions (α , β - and β , γ -, respectively) and employ different mechanisms to perform catalysis, it is possible to develop modifications that affect decapping at only one selected position or at both positions.

2.4 ^{32}P Labeling of RNA Caps by Enzymatic Approaches

Capped RNAs and cap analogs radiolabeled with ^{32}P were one of the first tools that enabled biochemical studies of the structural details, functions, and metabolism of mRNA and snRNA caps. One of the most important applications of such tools is to investigate the biochemical activities of cap-specific enzymes involved in RNA turnover. Synthesis of ^{32}P -labeled capped mRNA is usually achieved by transcription of a DNA template *in vitro*, followed by post-transcriptional enzymatic capping using radiolabeled NTP in at least one of the synthetic steps. Knowing the mechanistic details of phosphate transfer reactions taking place during transcription and capping reactions allows the design of methods to produce transcripts that are site-specifically radiolabeled within or in the vicinity of the cap attached to either a uniformly radiolabeled or unlabeled RNA body.

2.4.1 Incorporation of ^{32}P into RNA Caps

Usually, RNAs used for biochemical studies are generated in transcription reactions, in which DNA-dependent RNA polymerase uses four unlabeled (‘cold’) NTPs to transcribe the DNA sequence encoding a specific promoter region recognized by the polymerase and sequence the region of interest into RNA (Fig. 5). To obtain RNAs radiolabeled at the α position of the cap’s triphosphate bridge ($\text{m}^7\text{Gppp}^*\text{RNA}$ or $\text{CH}_3\text{ppp}^*\text{RNA}$, where p^* denotes ^{32}P -labeled phosphate), a ‘hot’, α -phosphate-labeled version of the first transcribed NTP ($[\alpha\text{-}^{32}\text{P}]\text{NTP}$) is added to the transcription reaction. This initially yields transcripts radiolabeled at the α position of the 5′-triphosphate moiety (ppp^*RNA), which can be subsequently subjected to capping by VCE or γ -MTase under standard conditions to yield the desired product (Fig. 5). To obtain capped RNAs radiolabeled within the first phosphodiester bond (e.g., $\text{m}^7\text{GpppNp}^*\text{RNA}$) the $[\alpha\text{-}^{32}\text{P}]\text{NTP}$ version of the second transcribed nucleotide is added to the transcription reaction, followed by enzymatic capping (Fig. 5). Finally, ^{32}P can be introduced at the γ position of a cap’s triphosphate chain (to produce $\text{m}^7\text{Gp}^*\text{ppRNA}$) by adding $[\alpha\text{-}^{32}\text{P}]\text{GTP}$ into the VCE-mediated capping reaction performed on ‘cold’ RNA (Fig. 5). Notably, this approach is different from the two previous ones as it leaves the rest of the RNA body completely unlabeled. It is also possible to obtain RNA selectively radiolabeled at the β -phosphate of the cap by adding an appropriate $[\beta\text{-}^{32}\text{P}]\text{NTP}$ to the transcription reaction [42]. However, to our knowledge, this approach is rarely used, likely owing to a lack of commercial availability of $[\beta\text{-}^{32}\text{P}]\text{NTPs}$.

^{32}P labeling can be performed *in vitro* to create RNA of a pre-determined sequence, but can also be applied to the synthesis of total RNA in cell extracts followed by biochemical analysis. The latter approach has been employed a number

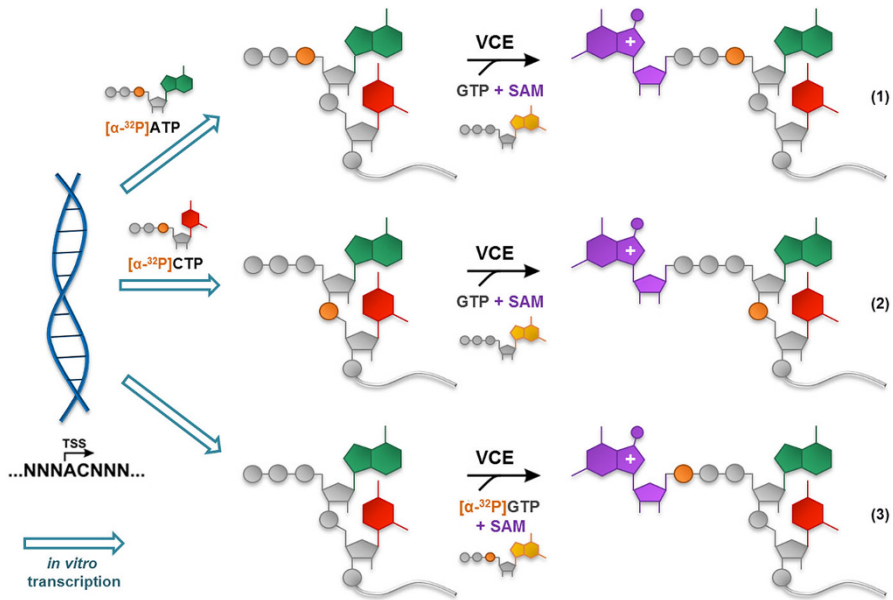


Fig. 5 Generation of differently ^{32}P -labeled RNAs using *in vitro* transcription followed by enzymatic capping. Depending on the template sequence and ‘hot’ NTP used in the transcription reaction, RNA can be labeled at the α position within the cap triphosphate bridge (1) or at the first phosphodiester bond of transcribed RNA (2). Alternatively, the γ -phosphate of the cap structure can be specifically labeled if ‘hot’ GTP is used in VCE-catalyzed capping (3)

of times in different organisms and on various RNA types to analyze the RNA 5' end heterogeneity, with the focus on both the distal cap modification and the nucleosides at the positions +1 and +2. To this end, total RNA is transcribed in cell extract in the presence of one or more ‘hot’ NTPs. The transcripts are resolved by electrophoresis, fractions of interest excised from the gel, and subjected to degradation by 3'–5' nuclease such as RNase P1 or a cocktail of ribonucleases, optionally followed by alkaline phosphatase, which releases nucleosides and inorganic phosphate from the RNA body but leaves undigested RNA 5' ends of $\text{Xppp}(\text{Np})_n$ (where $n = 1\text{--}3$). Released radioactive products can be analyzed using 2D TLC, usually on cellulose plates, or high-performance liquid chromatography (HPLC) on reversed-phase or ion exchange columns. Alternatively, RNA cleavage can be performed with a more specific endonuclease. For instance, RNase T1 cleaves RNA after 3'-guanylic residues leaving a 3'-phosphate and a 5'-OH in the products. This RNase is particularly useful for the analysis of transcripts initiated with G and labeled within the first transcribed phosphodiester bond ($\text{XpppGp}^*\text{N-RNA}$), because it cleaves them to release radiolabeled cap structures of XpppGp^* type (along with other RNA fragments). The identity of the cap structures can be confirmed by isolation and further digestion by a pyrophosphatase with broad specificity, such as tobacco acid pyrophosphatase (TAP), which cleaves both triphosphate bonds. This approach has been used to identify 5' ends of eukaryotic and viral mRNAs, snRNAs, and snoRNAs [42–45], as well as to verify the

efficiency of incorporation and orientation of synthetic dinucleotide cap structures introduced into RNA co-transcriptionally [36, 46, 47].

2.4.2 ^{32}P -Labeled RNAs and Cap Analogs in the Study of RNA Turnover

Another important use of radiolabeled capped RNAs is for gaining new insights into mRNA and cap turnover. Cap turnover starts when a transcript is subjected to cellular RNA degradation machinery. There are two main mRNA degradation pathways 5′–3′ and 3′–5′, both initiated by poly(A) tail shortening [48]. Although decapping is not the first event in the 5′–3′ mRNA degradation pathway, as it is preceded by deadenylation, it is considered as the first highly irreversible step. In eukaryotes, Dcp1/2 serves as the main decapping enzyme [49]. Dcp2 is a catalytic subunit belonging to the Nudix family of hydrolases [50, 51] whereas Dcp1 is a regulatory subunit, which can interact with additional decapping enhancers such as Edc1, 2, and 3, or PNRC2 [52].

Biochemical analysis using $m^7\text{Gp}^*\text{pp}$ - and $m_3\text{Gp}^*\text{pp}$ -capped mRNAs and snRNAs revealed that human and yeast Dcp2 hydrolyze the cap exclusively between α and β phosphates to yield $[\alpha\text{-}^{32}\text{P}]m^7\text{GDP}$ or $[\alpha\text{-}^{32}\text{P}]m_3\text{GDP}$ and RNA 5′-monophosphate [50, 51, 53] which can undergo further degradation by 5′-exonucleases (Fig. 6). Dcp2 requires the RNA body for activity as it does not hydrolyze $m^7\text{Gp}^*\text{ppG}$ in vitro [53]. Recently, it has been found that, at least in vitro, several other enzymes, such as Nudt16, Dxo1, and Rai1, can catalyze RNA decapping at the α - β position [54–56], but their exact roles in RNA degradation in vivo are yet to be established.

In contrast, the second main enzyme engaged in cap turnover—DcpS, which is responsible for depletion of products of 3′–5′ RNA decay—prefers cap dinucleotides as substrates [57]. $m^7\text{Gp}^*\text{ppN}$ dinucleotides and short $m^7\text{Gp}^*\text{pp}$ -RNAs are hydrolyzed by recombinant DcpS between β and γ phosphates to yield $m^7\text{Gp}^*$ and a downstream nucleoside or RNA 5′-diphosphate [57–59].

Both decapping pathways produce either $m^7\text{GMP}$ or $m^7\text{GDP}$ mononucleotides as reaction products. For a long time, it was hypothesized that cells must have developed mechanisms that protect them from accumulation of $m^7\text{G}$ nucleotides, which either could inhibit cap-dependent proteins or, after conversion into $m^7\text{GTP}$, be erroneously salvaged by RNA polymerases. A study on degradation of enzymatically generated $m^7\text{Gp}^*\text{ppG}$ -RNA and $m^7\text{Gp}^*\text{p}$ in mammalian cell extracts using TLC and autoradiography suggested that an additional function of DcpS in mRNA decay is to cleave $m^7\text{GDP}$ to $m^7\text{GMP}$ [60]. However, it was later shown that $m^7\text{GDP}$, in contrast to $m^7\text{GpppG}$ and $m^7\text{GTP}$, is not a substrate for DcpS, and at least in vitro acts as a DcpS inhibitor [61]. A new study performed in yeast and mammalian cell extracts on $m^7\text{Gp}^*\text{ppG}$, $m^7\text{Gp}^*\text{p}$, and $m^7\text{Gp}^*$ revealed that $m^7\text{GDP}$ metabolism may proceed through its phosphorylation to $m^7\text{GTP}$ carried out by nucleoside diphosphate kinase (NDPK) and subsequent degradation by DcpS to $m^7\text{GMP}$ [62]. Interestingly, $m^7\text{GMP}$ is not the end product of cap metabolism in yeast, as degradation of $m^7\text{Gp}^*\text{ppG}$ led to only small amounts of $m^7\text{Gp}^*$ and high amounts of radioactive Pi and some unidentified metabolite X with TLC mobility similar to Pi [62]. This study has also shown that in the absence of DcpS activity

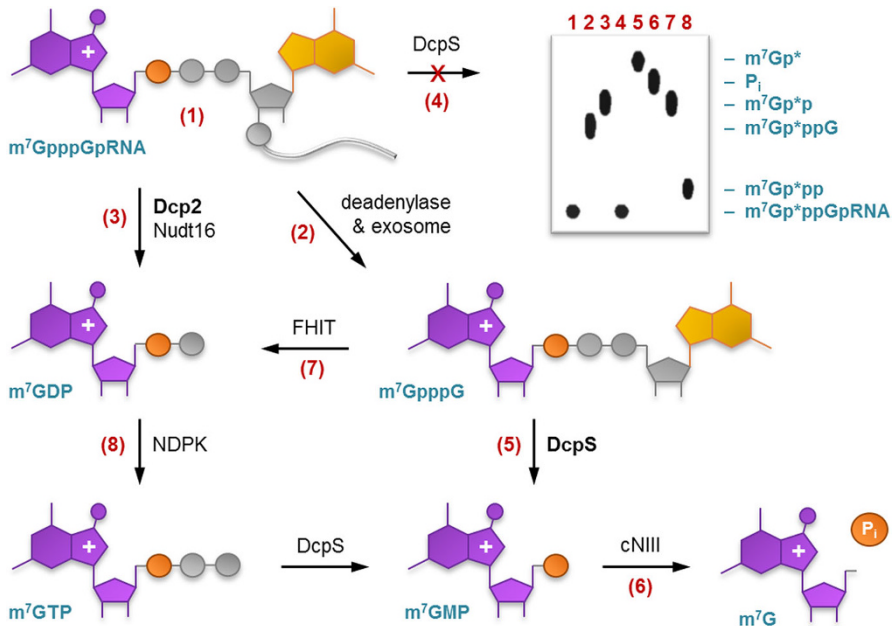


Fig. 6 Investigation of RNA degradation and cap turnover pathways using ‘hot’ γ -phosphate-labeled capped RNA. The main enzymes involved in these processes *in vivo* are marked in **bold**. Schematic representation of a typical TLC analysis of cap turnover in the presence of different enzymes

(gene knockout), the *Fhit* enzyme can take part in m^7Gp^*ppG turnover by cleaving it between α and β phosphates to release m^7Gp^*p [62]. Even less is known about m^7GMP metabolism, although a cytosolic nucleotidase III-like enzyme (*cNIII*) has been described that utilizes m^7GMP as a substrate (Fig. 6) [63].

^{32}P -labeled capped RNA oligonucleotides were used to evaluate the susceptibility of various synthetic cap structures modified in the triphosphate bridge (discussed in more detail below) to *Dcp2* [64, 65] and to screen potential inhibitors of *Dcp2*-catalyzed decapping [66]. If short capped RNA oligonucleotides (up to 50 nt) were used as *Dcp2* substrates, the electrophoretic resolution of capped and decapped RNAs could be performed directly on sequencing gels.

2.4.3 Enzymatic Labeling Beyond Radioactive Phosphates

As described in the previous section, by taking advantage of enzymes engaged in cap biosynthesis and turnover, it is possible to generate ^{32}P -labeled versions of variously capped RNAs, different cap structures, and cap metabolites, which provide tools instrumental in numerous biological studies (Fig. 6). However, labeling with ^{32}P has some limitations such as the production of unpredicted radioactive metabolites or incompatibility with continuous (on-line) enzymatic activity monitoring, *in vivo* studies, and *in cellulo* visualization. Therefore, labeling methods directed towards introduction of fluorescent labels or biological tags into

capped RNA are constantly being developed. Interestingly, some of those methods rely on the same biochemical approaches as those developed for radioactive labeling. For example, it has been recently reported that VCE apart from GTP accepts certain GTP analogs as substrates, including those functionalized at the ribose moiety [67–69]. Taking advantage of this, VCE has been independently used to transfer a biotinylated [69] or anthraniloylated [68] GMP moiety to the 5' end of RNA from a GTP analog appropriately functionalized at the ribose moiety. Notably, the label structure is of great importance for success in this approach, since a similar reaction using GTP carrying a manthraniloyl dye failed to produce capped transcripts [68]. Another interesting chemo-enzymatic approach to enzymatic cap labeling proposed by Rentmeister et al. takes advantage of methyltransferase activity of trimethylguanosine synthase 2 from *Giardia lamblia* (GlaTgs2-Var1). GlaTgs2-Var1 was first mutated (V34A) to accommodate bulkier, chemically modified analogs of *S*-adenosyl-L-methionine (SAM) as co-substrates [70–72], thereby allowing for transfer of various functional groups from synthetic SAM analogs onto the N^2 position of an m^7G moiety within the cap. The introduced functional groups were reactive in so-called click reactions such as copper(I) (CuAAC) or strain promoted azide-alkyne cycloaddition (SPAAC), tetrazole photoclick (PC), thiol-ene conjugation (TEC) and inverse electron-demand Diels–Alder cycloaddition (IEDDA) and, as such, were used to label a dinucleotide 5' cap analog m^7GpppA with properly functionalized fluorescent dyes. Although such modifications of N^2 position are expected to disturb mRNA activity in translation, they have great potential for applications related to labeling and quantification of endogenous or exogenously delivered RNAs in vitro and in vivo. Recently, another methyltransferase—Ecm1 from the microsporidian parasite *Encephalitozoon cuniculi*—was shown to transfer some bulky residues from SAM analogs to the N^7 atom of GpppG capped transcripts [72]. Such functionalized mRNAs were then delivered into HeLa cells by transfection where they reacted with dibenzocyclooctyne functionalized sulforhodamine B in a SPAAC reaction.

2.5 Chemical and Chemoenzymatic Approaches Towards Phosphate-Modified Cap Analogs

2.5.1 Possible Sites for Chemical Modifications of the RNA Caps

The repertoire of available enzymatic methods to modify cap structures is restricted by substrate specificity of particular enzymes. Much greater versatility of cap analogs can be generated by means of chemical synthesis. The influence of chemical modifications within the cap on its biological properties has been studied almost since the very moment of cap discovery. To date, almost every position of m^7G and TMG cap structures has been chemically modified, and the influence of those modifications on interaction with cap-binding proteins has been studied, often followed by the determination of the biological consequences of such modifications on RNA stability, transport, and translation. Interestingly, triphosphate chain modifications were explored much later than modifications of the 7-methylguanine or ribose moieties [36, 73–76], but turned out to have great potential to modulate the

interactions of the cap-structure with cap-binding proteins and its susceptibility to different decapping enzymes.

In the next paragraphs, we describe the variety of phosphate modifications that can be introduced into mRNA cap analogs by means of chemical synthesis followed by their selected biochemical, biological and even medicinal applications, as small molecules, as a part of RNA oligonucleotides, and as full-length transcripts.

2.5.2 Synthesis of RNA Caps: P-Imidazolides and MCl_2 -Mediated Coupling

The main challenge in chemical synthesis of cap analogs is the formation of the pyrophosphate bonds from two different phosphate derivatives necessary to attain an asymmetrically substituted oligophosphate bridge. The most commonly used method relies on conversion of one of the phosphate substrates into P-imidazolidine, which could be then coupled with an appropriate nucleophile in a MCl_2 -mediated reaction (where M stands for metal, usually Zn or Mg) [77, 78]. Such an approach provides a relatively simple way to obtain a wide scope of chemically modified oligophosphate mono- and diesters (Fig. 7), including phosphorothioates [79–81], seleno- [82], borano- [83], and fluorophosphates [84], as well as phosphoramidates [85], C-phosphonates [86], and bisphosphonates [87]. P-imidazolides were shown to react readily with other inorganic nucleophiles such as fluorides [84] and sulfates [88] to yield fluorophosphates and phosphosulfates, respectively.

An important limitation of this method lies in the synthesis of P-imidazolides, which is very efficient for non-modified mono-, di-, and even triphosphates but fails to produce a satisfactory yield of the desired product from compounds modified within the terminal phosphate. A solution to this problem was proposed by introducing electrophilic phosphorylating reagents, such as cyanoethyl phosphate or thiophosphate P-imidazolides, which are capable of reacting with non-activated nucleotides (Fig. 7) [89]. Such reagents were shown to react with a series of phosphates, phosphorothioates, and seleno- and borano-phosphates, which was followed by one-step removal of cyanoethyl-protecting groups, provided straightforward access to α -modified nucleoside diphosphates and β -modified triphosphates.

The P-imidazolidine-based approach was employed for the synthesis of many phosphate-modified cap analogs used as research tools for probing interaction with cap-binding proteins and enzymes. One example is a therapeutically important cap analog called β -S-ARCA [80, 90], bearing a 2'-O-methyl group within the ribose of 7-methylguanosine and a phosphorothioate modification at the β position of the triphosphate bridge (Fig. 8a). The synthetic pathway included two consecutive $ZnCl_2$ -mediated coupling reactions: first P-imidazolidine of 2'-O-methyl-7-methylguanosine monophosphate was reacted with triethylammonium thiophosphate and the resulting β -thiodiphosphate was coupled with P-imidazolidine of guanosine monophosphate. Notably, the final product was obtained as a roughly equimolar mixture of R_P and S_P diastereoisomers resulting from the β -phosphate modification. Nonetheless, diastereoisomeric cap analogs could be separated by semi-preparative reversed-phase (RP) HPLC even at a multi-milligram scale.

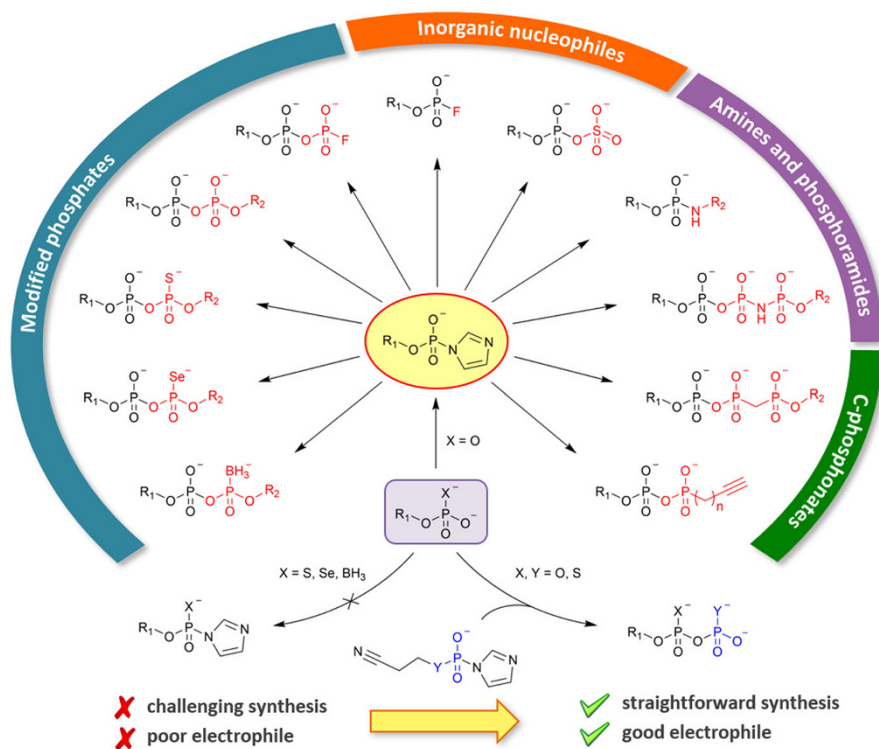


Fig. 7 Reactivity scope and limitations of P-imidazolides as substrates for the synthesis of chemically modified phosphate derivatives

Another example of the usefulness of MCl_2 -mediated coupling reactions involving P-imidazolides is the synthesis of a symmetrical two-headed tetraphosphate cap analog [91], which was found to be a potent inhibitor of Dcp2 decapping enzyme acting as an mRNA 5' cap mimic [27, 66]. Owing to the symmetry of the final compound, it was possible to perform two coupling reactions in one synthetic step, starting with P^1, P^2 -diimidazolyl-pyrophosphate and two equivalents of 7-methylguanosine 5'-phosphorothioate (Fig. 8b).

MCl_2 -mediated coupling reactions usually reach very high conversion of reactants in a reasonable length of time (1–48 h depending on the sterical hindrance); however, they require an excess of divalent metal chloride (4–16 molar equivalents) [77, 78]. $ZnCl_2$, $MgCl_2$, or $MnCl_2$ are the most commonly used for that purpose (Fig. 8). The role of metal ions is probably complex and has never been studied in detail, but the catalytic effect could be mainly attributed to three aspects related to the formation of phosphate complexes (Fig. 8d): (i) eliminating the electrostatic repulsion between negatively charged moieties and templating the reactants, (ii) increasing the solubility of reactants in an organic solvent (usually DMF), and (iii) increasing the electrophilicity of phosphorus atoms by complexation of imidazole.

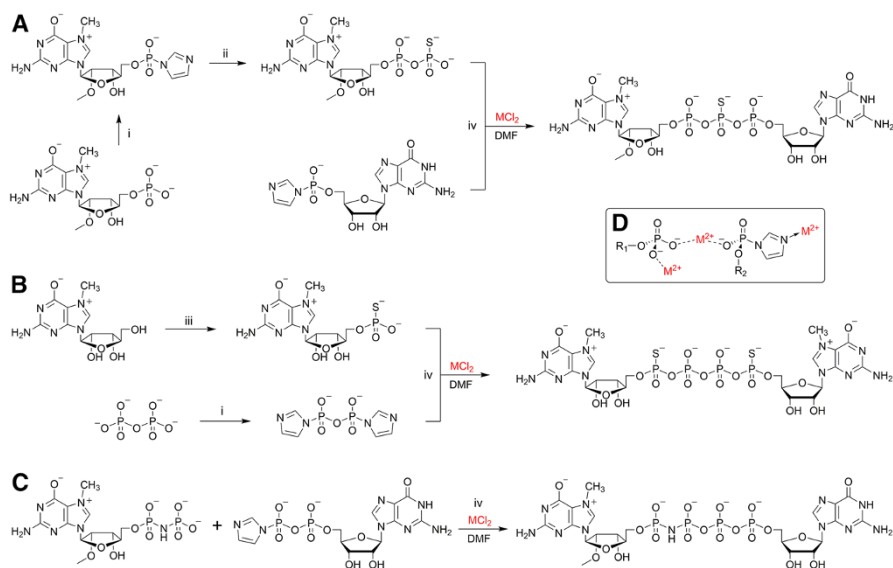


Fig. 8 Synthetic pathways leading to important chemically modified cap analogs. **a** The key steps in the synthesis of β -S-ARCA analog; **b** synthesis of a two-headed tetraphosphate cap analog—inhibitor of Dcp2 enzyme; **c** synthesis of a cap analog containing a γ - δ imidodiphosphate moiety in the tetraphosphate bridge; **d** putative role of divalent metal cations in coupling reactions. Reagents: (i) imidazole, 2,2'-dithiodipyridine (DTPD), triphenylphosphine (Ph_3P), triethylamine (TEA), DMF; (ii) triethylammonium thiophosphate, zinc chloride, DMF; (iii) thiophosphoryl chloride, 2,6-dimethylpyridine, trimethyl phosphite; (iv) zinc chloride, DMF

2.5.3 Phosphate-Modified Cap Analogs as Molecular Probes for Cap-Dependent Processes

A significant number of dinucleotide phosphate-modified cap analogs carrying single or multiple O-to-X substitutions in the tri- or tetraphosphate chain (where X is a single atom or a group of atoms) have been systematically studied over the last decade to probe the structural requirements of eIF4E protein for cap recognition [81, 83, 86, 92, 93]. These studies revealed that while some modifications disturb recognition by eIF4E, some have a rather negligible effect, whereas others stabilize the cap-eIF4E complex (Fig. 9). The affinity of eIF4E influences the ability of the cap to promote mRNA translation, thereby modulating properties of mRNA molecules *in vitro* and *in vivo*, as described further in chapter 2.4.4. Similarly, phosphate-modified m_3G cap analogs were employed to study interaction of m_3G cap with SNP1, degradation by hNudT16, and nuclear import of m_3G cap bioconjugates [94–96].

Phosphate moieties of small molecule cap analogs such as 7-methylguanine mononucleotides or dinucleotides can be easily modified to carry a spectroscopic label sensitive to changes in the local environment caused by protein binding or enzymatic cleavage. Synthesis of such probes can be achieved by chemical modification of the terminal phosphate moiety in a mononucleotide. The terminal phosphate can be conveniently functionalized using a variety of chemistries to

enabled both online DcpS activity monitoring as well as inhibitor screening assay development (Fig. 11c) [99]. The vivid difference between the influence of guanine and 7-methylguanine on AcPy fluorescence opens avenues for using GTP γ SAcPy to study biologically relevant guanine N^7 methylation processes.

2.5.4 mRNA Modification with Synthetic Cap Analogs (Transcription In Vitro)

Beside their use as small molecular ligands, phosphate-modified dinucleotide cap analogs can be used as reagents for mRNA 5' end modification, as previously reviewed by us in detail [92].

The incorporation of dinucleotide cap analogs into RNA can be achieved co-transcriptionally by in vitro transcription reactions. To this end, a DNA template encoding G as the first transcribed nucleotide is transcribed by a bacteriophage RNA polymerase in the presence of all four NTPs, and a dinucleotide m⁷GpppG. The transcription is initiated by the nucleophilic attack of the 3'-OH group of guanosine from m⁷GpppG or GTP on the α phosphate of the second transcribed nucleotide to produce, eventually, a mixture of capped and uncapped transcripts (Fig. 12). To increase the fraction of capped RNAs in the transcription product (capping efficiency), the concentration of GTP in the reaction mixture is usually decreased, whereas m⁷GpppG concentration is elevated (up to tenfold over GTP) [100]. Pasquinelli et al. have shown that T7 RNA polymerase can initiate the transcription by the attack of the 3'-OH group of 7-methylguanosine in m⁷GpppG, thereby incorporating the dinucleotide in the so-called reverse orientation (Fig. 12) [46]. Such event results in 30–50% of capped mRNAs having non-functional cap structures, which are not translated efficiently by cap-dependent mechanisms. A solution to that problem was proposed in the form of anti-reverse cap analogs (ARCAs). These analogs are modified (or blocked) at the 3'- or 2'-OH hydroxyl of 7-methylguanosine to prevent recognition of this moiety by RNA polymerase

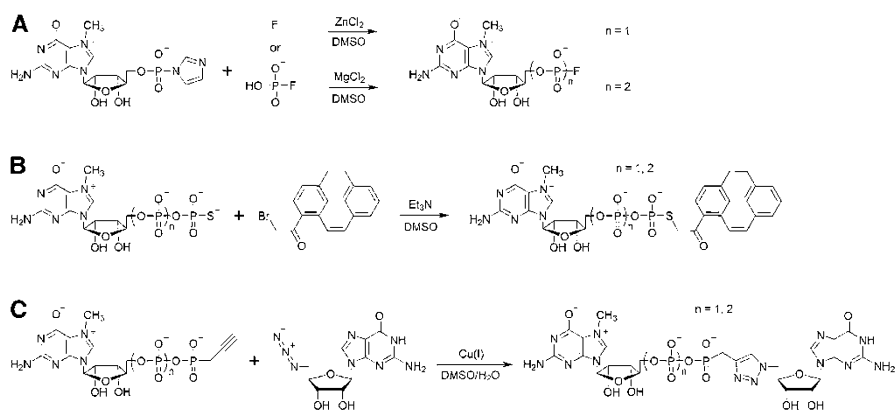


Fig. 10 Various approaches to cap analog functionalization via terminal phosphate. **a** Synthesis of 7-methylguanine mononucleotides labeled with fluorine atom via phosphorimidazole chemistry; **b** synthesis of cap analogs carrying an acetylpyrene label via S-alkylation of the terminal phosphorothioate group; **c** synthesis of triazole-containing dinucleotide analogs via CuAAC

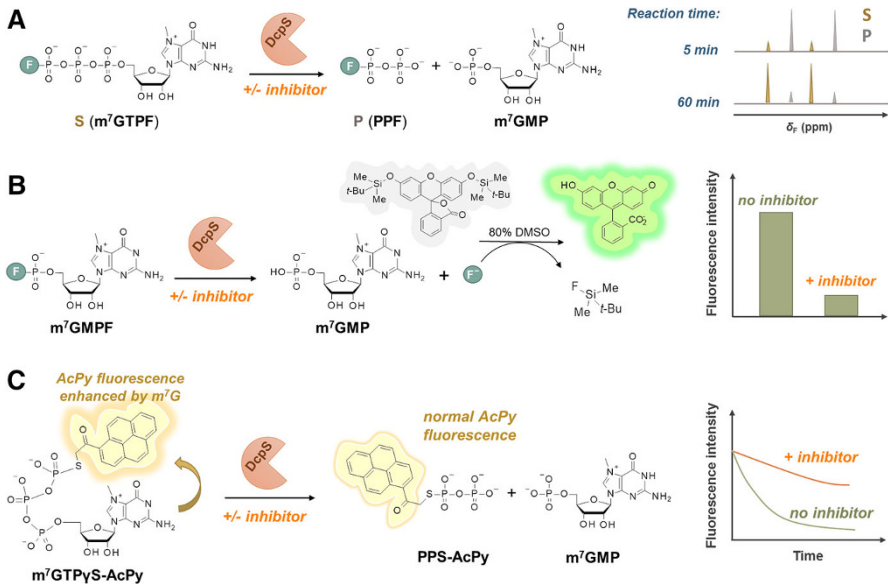


Fig. 11 Applications of phosphate-modified cap analogs for monitoring or assaying activity of the human DcpS enzyme. **a** DcpS activity monitoring using ^{19}F NMR and m^7GTPF as a substrate; **b** the idea of high-throughput inhibitor screening assay for DcpS using m^7GMPF as a substrate and a fluoride-sensitive chemosensor as a fluorogenic probe; **c** assaying DcpS activity by monitoring hydrolysis of an acetylpyrene-labeled fluorescent probe

[36, 101, 102]. The simplest and most commonly used modification is methylation of one of those groups, but ARCAs containing hydrogen (O–H substitution) or fluorine (O–F substitution) or even bulkier substituents have also been reported [37, 38, 103–106]. Importantly, chemical modifications at these positions do not affect interactions with eIF4E or translation efficiency of such capped mRNAs, although, in the case of some bulkier substituents, the translation efficiency can be slightly decreased in comparison to ARCAs carrying simply a methyl group [37, 38]. Adenine dinucleotides such as NAD, FAD, and dpCoA can be introduced into RNA 5' ends by transcription in vitro of a DNA template encoding A as the first transcribed nucleotide [107].

Importantly, the co-transcriptional capping approach opens the possibility for straightforward introduction of chemical modifications into the cap. First, the ‘anti-reverse’ type modifications of the m^7G cap can be additionally functionalized to carry a label or a biological tag such as biotin or fluorescent dye. Second, a number of phosphate-modified nucleotides were shown to be accepted as transcription initiators for T7 and SP6 RNA polymerases, giving access to phosphate-modified capped mRNAs as tools to study the specificity of cleavage by RNA-dependent capping enzymes, with Dcp1/2 being the prime example [64, 65, 108]. Two methylene-bisphosphonate ARCAs differing in the site of O to CH_2 substitution (either α - β or β - γ , Fig. 9) allowed for differential inhibition of decapping pathways [64]. First, in vitro experiments were performed on short radiolabeled RNAs to

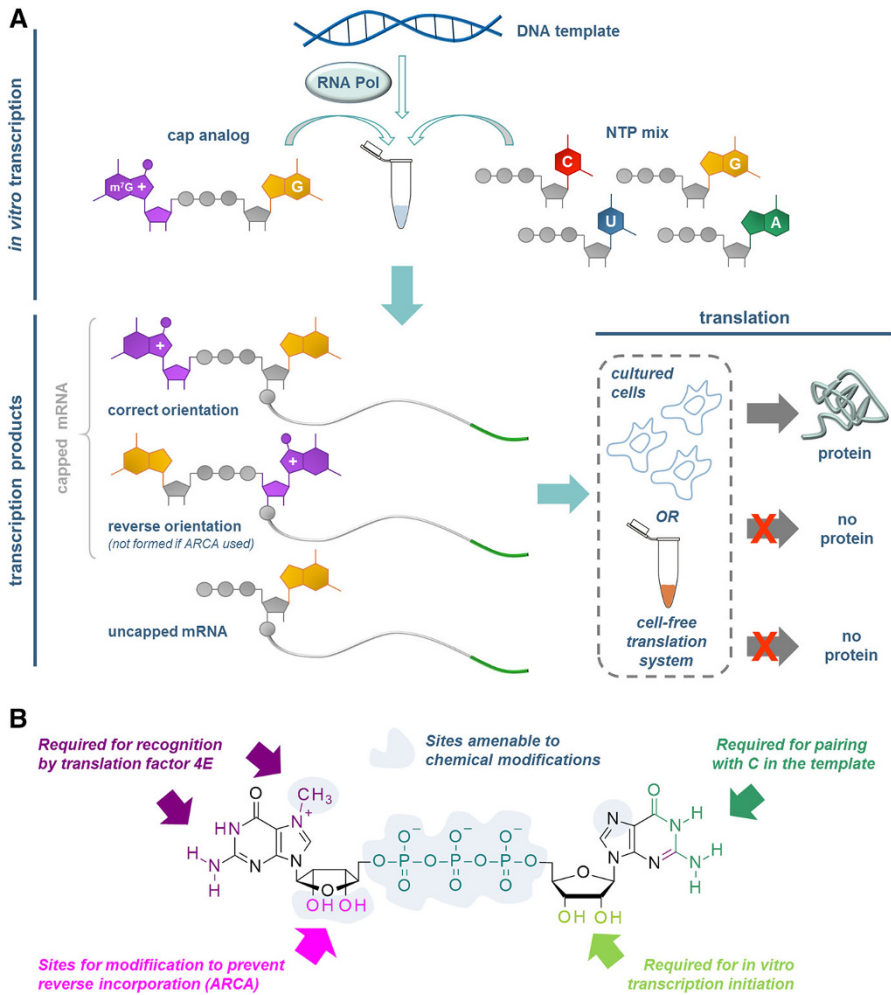


Fig. 12 Synthesis of capped RNAs by so-called co-transcriptional capping during transcription *in vitro*. **a** General scheme of the process; **b** structural requirements for a cap analog to be incorporated into mRNA and to produce anti-reverse cap analogs (ARCAs)

determine the susceptibility of the modified caps to recombinant hDcp2. It was found that the RNAs with the cap modified at the α - β position were resistant to cleavage by Dcp2, while the RNAs with the cap modified at the β - γ position were similarly susceptible to the unmodified cap. A study on degradation of full-length mRNAs in mammalian cells revealed that preventing the cleavage at the α - β position of the triphosphate bridge (performed by Dcp1/2) increases the half-life of mRNA. This observation combined with earlier studies which showed that modification at the β - γ position protects the cap from cleavage only by DcpS leads to the conclusion that cap degradation is a limiting step for the 5'-3' degradation pathway but not for the 3'-5' pathway (Fig. 9).

Despite the extended half-life of mRNAs capped with α - β modified analogs, such transcripts had poor translational properties, which inspired the quest for novel cap structures that could confer to mRNA both resistance to Dcp2 and high translational efficiency. The first analogs, which combined these two features, were compounds carrying a non-bridging O-S substitution at the beta position of the triphosphate chain, referred to as β -S-ARCA [108]. Later, it was found that similar properties could be conferred to mRNA by an O-to-BH₃ substitution [65, 83]. Recent studies have shown that applying two O-S substitutions at two adjacent phosphate moieties in a tri- or tetraphosphate bridge may provide biological properties superior to β -S-ARCA [81]. A series of phosphate-modified analogs carrying O-to-S, O-to-NH, O-to-BH₃, and O-to-Se substitutions at different positions have been used to study mRNA degradation in HeLa cells [65].

Typically, even small bridging modifications of the oligophosphate chain decrease the translation potential of mRNA, but the relation between size of the substituent and biological properties is not so obvious. Recently, a series of 34 cap analogs carrying a 1,2,3-triazole moiety within the oligophosphate chain were synthesized using copper-catalyzed azide-alkyne cycloaddition [86]. Biochemical evaluation of RNA co-transcriptionally capped with these analogs led to identification of two analogs, which despite this bulky triazole modification had translational properties similar to cap structures carrying an unmodified triphosphate chain. This unexpected finding highlights possibilities of novel approaches towards the synthesis of small molecule cap analogs as well as capped RNAs.

2.5.5 Phosphate Modifications for Chemical Capping Approaches

Over the 40 years since the discovery of the cap structure, chemical synthesis has provided access to a still-growing library of small molecule cap analogs, while the molecular biology techniques enabled incorporation of these structures into long RNAs. Somewhat surprisingly, exploring the 'middle ground'—which is the efficient synthesis of short capped oligonucleotides in high purity—still poses a challenge both for chemistry and biology. Short m⁷GpppRNAs, m₃GpppRNAs, and NAD-RNAs have been synthesized by chemical reaction of RNA-5' phosphates with appropriate imidazole-activated nucleotides under aqueous conditions [109–112], but these reactions are only moderately efficient and require time-consuming chromatographic and/or enzymatic work-ups to separate capped and uncapped RNAs. An improvement in solution synthesis of capped oligomers has been proposed by using a 4,4'-dimethoxytrityl (DMT) group as a lipophilic purification handle, which facilitates separation of capped RNAs from uncapped RNAs by RP HPLC [113]. Solid-phase synthesis of capped oligonucleotides has been attempted as well, but found to be challenging due to incompatibility of the m⁷G moiety with standard oligonucleotide de-immobilization and de-protection protocols [114–118]. Combination of RNA capping via these approaches with phosphate modifications appears even more challenging and remains to be demonstrated.

As such, a universal non-enzymatic method for the generation of bulk amounts of short capped RNAs would be of great interest to both chemical and biological

communities. To this end, alternative approaches to chemical RNA capping based on click chemistry have recently come into focus. One interesting alternative to the P-imidazolide approaches that has been recently proposed is based on the selective reaction of benzyl diazomethane derivatives with the terminal phosphate of unmodified RNA 5'-(mono/di/tri)phosphates to give appropriate benzyl esters [119]. Such a reaction was applied for the functionalization of the RNA 5' end with an orthosteric inhibitor of translation targeting the cap-binding site of eIF4E (Fig. 13a). Another example is chemical functionalization of short RNAs containing a modified 5'-(oligo)phosphate carrying an alkyne handle with azide-containing m₃G cap analogs or 5'-azido-5'-deoxy-7-methylguanosine (Fig. 13 b, c) [86, 96]. Interestingly, some RNAs capped with m₃G analogs obtained by CuAAC were still found to be recognized by decapping enzyme hNudT16, suggesting that these types of conjugates may be useful for biochemical and structural studies [96]. Conditions for high-yielding RNA click modifications that enables retaining desired biological activities of capped RNAs is yet to be determined, but if achieved it could finally provide access to high-yielding synthesis of short capped RNA mimics.

3 Summary and Future Prospects

Cap is a natural tag attached to the 5' end of messenger RNA and small nuclear RNA, which due to its unique structural features fulfills essential biological functions during the process of gene expression. The 'inverted' 5'-to-5' triphosphate bridge, alongside with a 7-methylguanosine moiety, is one of the two crucial elements of the mRNA cap structure. The triphosphate bridge is a site of cleavage for RNA decapping enzymes and is involved in the electrostatic interactions with other cap-related proteins. Therefore, synthetic cap analogs modified within the triphosphate bridge or carrying spectroscopic labels attached to phosphate moieties turned out to be valuable tools for studies on mRNA metabolism, modification of natural mRNA properties, or molecular probes for screening inhibitors of cap-dependent processes/proteins. Biorthogonal methods for modification of biologically relevant molecules give exceptional opportunities to follow or even to manipulate these processes inside the cells. mRNA has a great potential for use as a therapeutic agent in gene therapy, emerging in several ongoing clinical trials, which is a driving motivation for the development of new molecular tools for mRNA cap modification and for the construction of new cap-based molecular probes. Much progress has been made in microscopic techniques and single-molecule experiments allowing investigation into the properties of individual molecules, and providing another motivation to develop new methods for mRNA cap-labeling. The authors of this chapter are convinced that the coming years will bring intensification of studies aimed at developing new applications of labeled cap analogs in a cellular context. The continuous improvements in understanding of how the cap interacts with its cellular partners, protein factors, and decapping enzymes lead us to believe that in the future, modifications introduced into the mRNA cap structure will minimize the interruption of mRNA biological functions. Progress in the design and synthesis of molecular probes will enable investigation of processes that are poorly understood,

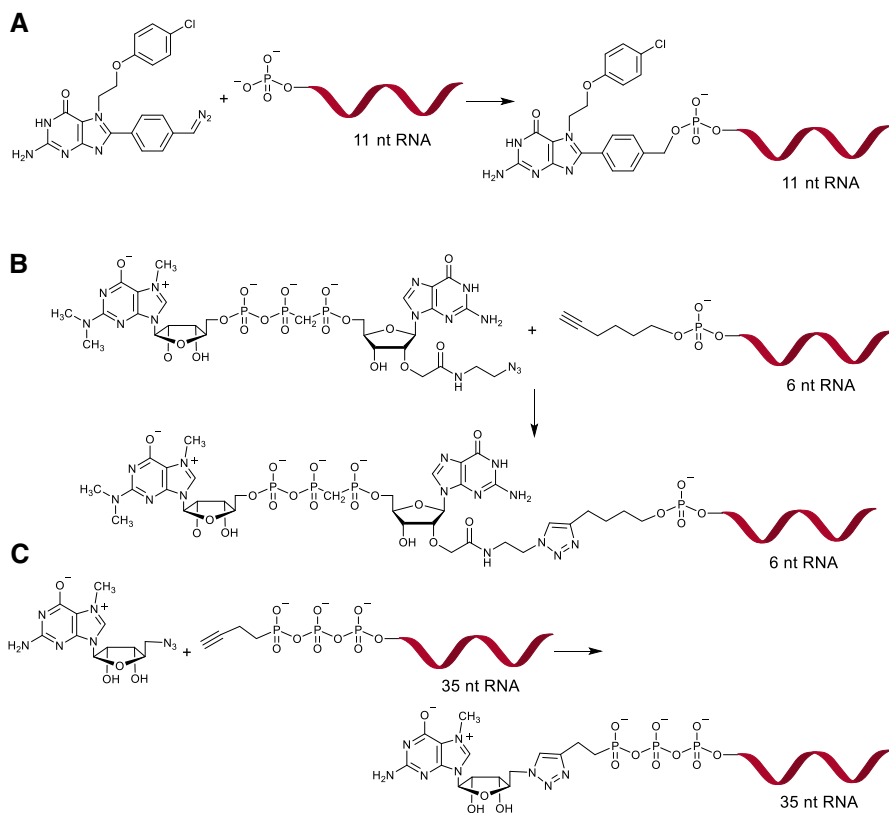


Fig. 13 Modification of the RNA 5' end by direct attachment of cap analogs or mimics (chemical capping). **a** Chemical esterification of phosphorylated RNA using a diazobenzyl derivative of translation inhibitor-targeting eIF4E protein; **b** application of CuAAC for conjugation of a dinucleotide m3G cap analog and chemically synthesized RNA; **c** chemical capping of enzymatically synthesized RNA with 5'-azido-5'-deoxy-7-methylguanosine under CuAAC conditions

such as cellular correlation between initiation of translation and mRNA degradation, post-transcriptional methylation of mRNA 5' end, and antiviral immune response. The discovery of structures similar to the mRNA cap, such as the NAD cap in bacteria, may also open a new chapter in this field. Various approaches applied for the investigation of mRNA and snRNA caps can be adapted to study biological functions of their bacterial relatives. Finally, advancements in screening methodologies will allow researchers to find more selective inhibitors of cap-dependent processes with high therapeutic potency. The immense progress in nucleotide and nucleic acid delivery methodologies we are currently witnessing will likely bring phosphate-modified cap analogs to in vivo applications.

Acknowledgements This work was financially supported by the National Science Centre (Poland, grants UMO-2013/09/B/ST5/01341 to JJ and UMO-2011/01/D/ST5/05869 to JK) and by the Ministry of Science and Higher Education (Poland, DI2012 024842 to MW).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Wei CM, Gershowitz A, Moss B (1975) Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. *Cell* 4:379–386
2. Furuichi Y, Muthukrishnan S, Shatkin AJ (1975) 5'-Terminal m⁷G(5')ppp(5')GMP in vivo - identification in reovirus genome RNA. *Proc Natl Acad Sci USA* 72:742–745
3. Ramanathan A, Robb GB, Chan SH (2016) mRNA capping: biological functions and applications. *Nucl Acid Res* 44:7511–7526
4. Topisirovic I, Svitkin YV, Sonenberg N, Shatkin AJ (2011) Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip Rev Rna* 2:277–298
5. Furuichi Y, Shatkin AJ (2000) *Advances in virus research*, vol 55. Academic Press Inc, San Diego, pp 135–184
6. Saldittgeorgieff M, Harpold M, Chenkiang S, Darnell JE (1980) Addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell* 19:69–78
7. Decroly E, Ferron F, Lescar J, Canard B (2012) Conventional and unconventional mechanisms for capping viral mRNA. *Nat Rev Microbiol* 10:51–65
8. Ziemiak M, Strenkowska M, Kowalska J, Jemielity J (2013) Potential therapeutic applications of RNA cap analogs. *Fut Med Chem* 5:1141–1172
9. Furuichi Y (2015) Discovery of m(7)G-cap in eukaryotic mRNAs. *Proc Jpn Acad Ser B Phys Biol Sci* 91:394–409
10. Daffis S, Szretter KJ, Schriewer J, Li JQ, Youn S, Errett J, Lin TY, Schneller S, Zust R, Dong HP et al (2010) 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* 468:452–456
11. Devarkar SC, Wang C, Miller MT, Ramanathan A, Jiang FG, Khan AG, Patel SS, Marcotrigiano J (2016) Structural basis for m⁷G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proc Natl Acad Sci USA* 113:596–601
12. Mauer J, Luo X, Blanjoie A, Jiao X, Grozhik AV, Patil DP, Linder B, Pickering BF, Vasseur J-J, Chen Q et al (2016) Reversible methylation of m⁶A_m in the 5' cap controls mRNA stability. *Nature*. doi:10.1038/nature21022
13. Mattaj IW (1986) Cap trimethylation of U-snRNA is cytoplasmic and dependent on U-snRNP protein-binding. *Cell* 46:905–911
14. Huber J, Dickmanns A, Luhrmann R (2002) The importin-beta binding domain of snurportin1 is responsible for the Ran- and energy-independent nuclear import of spliceosomal U snRNPs in vitro. *J Cell Biol* 156:467–479
15. Liou RF, Blumenthal T (1990) Trans-spliced *Caenorhabditis elegans* messenger RNAs retain trimethylguanosine caps. *Mol Cell Biol* 10:1764–1768
16. Wurth L, Gribling-Burrer AS, Verheggen C, Leichter M, Takeuchi A, Baudrey S, Martin F, Krol A, Bertrand E, Allmann C (2014) Hypermethylated-capped selenoprotein mRNAs in mammals. *Nucl Acid Res* 42:8663–8677
17. Gupta S, Busch RK, Singh R, Reddy R (1990) Characterization of U6 small nuclear-RNA cap-specific antibodies: identification of gamma-monomethyl-GTP cap structure in 7SK and several other human small RNAs. *J Biol Chem* 265:19137–19142
18. Chen YG, Kowtoniuk WE, Agarwal I, Shen YH, Liu DR (2009) LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat Chem Biol* 5:879–881
19. Hofer K, Li SS, Abele F, Frindert J, Schlotthauer J, Grawenhoff J, Du JM, Patel DJ, Jaschke A (2016) Structure and function of the bacterial decapping enzyme NudC. *Nat Chem Biol* 12:730–734

20. Bird JG, Zhang Y, Tian Y, Panova N, Barvik I, Greene L, Liu M, Buckley B, Krasny L, Lee JK et al (2016) The mechanism of RNA 5' capping with NAD(+), NADH and desphospho-CoA. *Nature* 535:444–447
21. Luciano DJ, Belasco JG (2015) NAD in RNA: unconventional headgear. *Trends Biochem Sci* 40:245–247
22. Jaschke A, Hofer K, Nubel G, Frindert J (2016) Cap-like structures in bacterial RNA and epitranscriptomic modification. *Curr Opin Microbiol* 30:44–49
23. Hodel AE, Gershon PD, Quijcho FA (1998) Structural basis for sequence-nonspecific recognition of 5'-capped mRNA by a cap-modifying enzyme. *Mol Cell* 1:443–447
24. Niedzwiecka A, Marcotrigiano J, Stepinski J, Jankowska-Anyszka M, Wyslouch-Cieszyńska A, Dadlez M, Gingras AC, Mak P, Darzynkiewicz E, Sonenberg N et al (2002) Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J Mol Biol* 319:615–635
25. Strasser A, Dickmanns A, Luhrmann R, Ficner R (2005) Structural basis for m(3)G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *EMBO J* 24:2235–2243
26. Gu MG, Fabrega C, Liu SW, Liu HD, Kiledjian M, Lima CD (2004) Insights into the structure, mechanism, and regulation of scavenger mRNA decapping activity. *Mol Cell* 14:67–80
27. Mugridge JS, Ziemniak M, Jemielity J, Gross JD (2016) Structural basis of mRNA-cap recognition by Dcp1-Dcp2. *Nat Struct Mol Biol* 23:987–994
28. Monecke T, Buschmann J, Neumann P, Wahle E, Ficner R (2014) Crystal structures of the novel cytosolic 5'-nucleotidase IIIb explain its preference for m(7)GMP. *PLoS One* 9(13):e90915
29. Calero G, Wilson KF, Ly T, Rios-Steiner JL, Clardy JC, Cerione RA (2002) Structural basis of m(7)GpppG binding to the nuclear cap-binding protein complex. *Nat Struct Biol* 9:912–917
30. Monecke T, Dickmanns A, Ficner R (2009) Structural basis for m(7)G-cap hypermethylation of small nuclear, small nucleolar and telomerase RNA by the dimethyltransferase TGS1. *Nucl Acid Res* 37:3865–3877
31. Quabius ES, Krupp G (2015) Synthetic mRNAs for manipulating cellular phenotypes: an overview. *New Biotechnol* 32:229–235
32. Shuman S (2015) RNA capping: progress and prospects. *RNA* 21:735–737
33. Fabrega C, Hausmann S, Shen V, Shuman S, Lima CD (2004) Structure and mechanism of mRNA cap (guanine-N7) methyltransferase. *Mol Cell* 13:77–89
34. Langberg SR, Moss B (1981) Post-transcriptional modifications of messenger-RNA - purification and characterization of cap-I and cap-II RNA nucleoside-2'-O-methyltransferases from HeLa-cells. *J Biol Chem* 256:54–60
35. Shuman S, Surks M, Furneaux H, Hurwitz J (1980) Purification and characterization of a GTP-pyrophosphate exchange activity from vaccinia virions: association of the GTP-pyrophosphate exchange activity with vaccinia messenger-RNA guanylyltransferase. RNA (guanine-7)-methyltransferase complex (capping enzyme). *J Biol Chem* 255:1588–1598
36. Stepinski J, Waddell C, Stolarski R, Darzynkiewicz E, Rhoads RE (2001) Synthesis and properties of mRNAs containing the novel "anti-reverse" cap analogs 7-methyl(3'-O-methyl)GpppG and 7-methyl(3'-deoxy)GpppG. *RNA-a Publ RNA Soc* 7:1486–1495
37. Warminski M, Kowalska J, Buck J, Zuberek J, Lukaszewicz M, Nicola C, Kuhn AN, Sahin U, Darzynkiewicz E, Jemielity J (2013) The synthesis of isopropylidene mRNA cap analogs modified with phosphorothioate moiety and their evaluation as promoters of mRNA translation. *Bioorg Med Chem Lett* 23:3753–3758
38. Jemielity J, Lukaszewicz M, Kowalska J, Czarnecki J, Zuberek J, Darzynkiewicz E (2012) Synthesis of biotin labelled cap analogue: incorporable into mRNA transcripts and promoting cap-dependent translation. *Org Biomol Chem* 10:8570–8574
39. Ziemniak M, Szabelski M, Lukaszewicz M, Nowicka A, Darzynkiewicz E, Rhoads RE, Wieczorek Z, Jemielity J (2013) Synthesis and evaluation of fluorescent cap analogues for mRNA labelling. *RSC Adv* 3:20943–20958
40. Rydzik A, Zuberek J, Kowalska J, Darzynkiewicz E, Jemielity J (2008) Bisphosphonate modification in tetraphosphate 5' mRNA cap analogs: synthesis and biochemical properties. *Chem Nucl Acid Comp* 10:444–448
41. Darzynkiewicz ZM, Bojarska E, Kowalska J, Lewdorowicz M, Jemielity J, Kalek M, Stepinski J, Davis RE, Darzynkiewicz E (2007) Interaction of human decapping scavenger with 5' mRNA cap analogues: structural requirements for catalytic activity. *J Phys Condens Matter* 19:285217

42. Contreras R, Fiers W (1981) Initiation of transcription by RNA polymerase-II in permeable, SV40-infected or noninfected, CV1 cells: evidence for multiple promoters of SV40 late transcription. *Nucl Acid Res* 9:215–236
43. Flavell AJ, Cowie A, Legon S, Kamen R (1979) Multiple 5' terminal cap structures in late polyomavirus RNA. *Cell* 16:357–371
44. Gidoni D, Kahana C, Canaani D, Groner Y (1981) Specific in vitro initiation of transcription of Simian Virus-40 early and late genes occurs at the various cap nucleotides including cytidine. *Proc Natl Acad Sci USA Biol Sci* 78:2174–2178
45. Shimba S, Buckley B, Reddy R, Kiss T, Filipowicz W (1992) Cap structure of U3 small nucleolar RNA in animal and plant-cells is different: gamma-monomethyl phosphate cap structure in plant RNA. *J Biol Chem* 267:13772–13777
46. Pasquinelli AE, Dahlberg JE, Lund E (1995) Reverse 5' caps in RNAs made in vitro by phage RNA polymerases. *RNA-a Publ RNA Soc* 1:957–967
47. Grudzien E, Stepinski J, Jankowska-Anyszka M, Stolarski R, Darzynkiewicz E, Rhoads RE (2004) Novel cap analogs for in vitro synthesis of mRNAs with high translational efficiency. *RNA-a Publ RNA Soc* 10:1479–1487
48. Labno A, Tomecki R, Dziembowski A (2016) Cytoplasmic RNA decay pathways: enzymes and mechanisms. *Biochim Biophys Acta* 1863:3125–3147
49. Coller J, Parker R (2004) Eukaryotic mRNA decapping. *Annu Rev Biochem* 73:861–890
50. Piccirillo C, Khanna R, Kiledjian M (2003) Functional characterization of the mammalian mRNA decapping enzyme hDcp2. *RNA-a Publ RNA Soc* 9:1138–1147
51. van Dijk E, Cougot N, Meyer S, Babajko S, Wahle E, Seraphin B (2002) Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J* 21:6915–6924
52. Arribas-Layton M, Wu DH, Lykke-Andersen J, Song HW (2013) Structural and functional control of the eukaryotic mRNA decapping machinery. *Biochim Biophys Acta Gene Regul Mech* 1829:580–589
53. Cohen LS, Mikhli C, Jiao XF, Kiledjian M, Kunkel G, Davis RE (2005) Dcp2 decaps m(2,2,7)GpppN-capped RNAs, and its activity is sequence and context dependent. *Mol Cell Biol* 25:8779–8791
54. Song M-G, Li Y, Kiledjian M (2010) Multiple mRNA decapping enzymes in Mammalian cells. *Mol Cell* 40:423–432
55. Song MG, Bail S, Kiledjian M (2013) Multiple Nudix family proteins possess mRNA decapping activity. *RNA-a Publ RNA Soc* 19:390–399
56. Jiao XF, Xiang S, Oh C, Martin CE, Tong LA, Kiledjian M (2010) Identification of a quality-control mechanism for mRNA 5'-end capping. *Nature* 467:608–U137
57. Liu HD, Rodgers ND, Jiao X, Kiledjian M (2002) The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. *EMBO J* 21:4699–4708
58. Liu SW, Jiao XF, Liu HD, Gu MG, Lima CD, Kiledjian M (2004) Functional analysis of mRNA scavenger decapping enzymes. *RNA-a Publ RNA Soc* 10:1412–1422
59. Wang ZR, Kiledjian M (2001) Functional link between the mammalian exosome and mRNA decapping. *Cell* 107:751–762
60. van Dijk E, Le Hir H, Seraphin B (2003) DcpS can act in the 5'-3' mRNA decay pathway in addition to the 3'-5' pathway. *Proc Natl Acad Sci USA* 100:12081–12086
61. Wypijewska A, Bojarska E, Lukaszewicz M, Stepinski J, Jemielity J, Davis RE, Darzynkiewicz E (2012) 7-Methylguanosine diphosphate (m(7)GDP) is not hydrolyzed but strongly bound by decapping scavenger (DcpS) enzymes and potently inhibits their activity. *Biochemistry* 51:8003–8013
62. Taverniti V, Seraphin B (2015) Elimination of cap structures generated by mRNA decay involves the new scavenger mRNA decapping enzyme Aph1/FHIT together with DcpS. *Nucl Acid Res* 43:482–492
63. Buschmann J, Moritz B, Jeske M, Lilie H, Schierhorn A, Wahle E (2013) Identification of drosophila and human 7-Methyl GMP-specific nucleotidases. *J Biol Chem* 288:2441–2451
64. Grudzien E, Kalek M, Jemielity J, Darzynkiewicz E, Rhoads RE (2006) Differential inhibition of mRNA degradation pathways by novel cap analogs. *J Biol Chem* 281:1857–1867
65. Su W, Slepnev S, Grudzien-Nogalska E, Kowalska J, Kulis M, Zuberek J, Lukaszewicz M, Darzynkiewicz E, Jemielity J, Rhoads RE (2011) Translation, stability, and resistance to decapping

- of mRNAs containing caps substituted in the triphosphate chain with BH₃, Se, and NH. RNA-a Publ RNA Soc 17:978–988
66. Ziemniak M, Mugridge JS, Kowalska J, Rhoads RE, Gross JD, Jemielity J (2016) Two-headed tetraphosphate cap analogs are inhibitors of the Dcp1/2 RNA decapping complex. RNA-a Publ RNA Soc 22:518–529
 67. Issur M, Bougie I, Despins S, Bisaillon M (2013) Enzymatic synthesis of RNAs capped with nucleotide analogues reveals the molecular basis for substrate selectivity of RNA capping enzyme: impacts on RNA metabolism. PLoS One 8:12
 68. Gunawardana D, Domashevskiy AV, Gayler KR, Goss DJ (2016) Efficient preparation and properties of mRNAs containing a fluorescent cap analog: anthraniloyl-m⁷G pppG (vol 3, e988538, 2014). Translation 4:1
 69. Ettwiller L, Buswell J, Yigit E, Schildkraut I (2016) A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. BMC Genom 17, 14, No 199
 70. Schulz D, Holstein JM, Rentmeister A (2013) A chemo-enzymatic approach for site-specific modification of the RNA cap. Angew Chem Int Ed 52:7874–7878
 71. Holstein JM, Schulz D, Rentmeister A (2014) Bioorthogonal site-specific labeling of the 5'-cap structure in eukaryotic mRNAs. Chem Commun 50:4478–4481
 72. Holstein JM, Anhauser L, Rentmeister A (2016) Modifying the 5'-cap for click reactions of eukaryotic mRNA and to tune translation efficiency in living cells. Angew Chem Int Ed 55:10899–10903
 73. Darzynkiewicz E, Stepinski J, Ekiel I, Goyer C, Sonenberg N, Temeriusz A, Jin YX, Sijuwade T, Haber D, Tahara SM (1989) Inhibition of eukaryotic translation by nucleoside 5'-monophosphate analogs of messenger RNA 5'-cap: changes in N⁷ substituent affect analog activity. Biochemistry 28:4771–4778
 74. Brown CJ, McNae I, Fischer PM, Walkinshaw MD (2007) Crystallographic and mass spectrometric characterisation of eIF4E with N-7-alkylated cap derivatives. J Mol Biol 372:7–15
 75. Chen X, Kopecky DJ, Mihalic J, Jeffries S, Min X, Heath J, Deignan J, Lai S, Fu Z, Guimaraes C et al (2012) Structure-guided design, synthesis, and evaluation of guanine-derived inhibitors of the eIF4E mRNA-Cap interaction. J Med Chem 55:3837–3851
 76. Ogasawara S, Maeda M (2011) Photoresponsive 5'-cap for the reversible photoregulation of gene expression. Bioorg Med Chem Lett 21:5457–5459
 77. Sawai H, Wakai H, Shimazu M (1991) Facile synthesis of cap portion of messenger-RNA by Mn(II)-catalyzed pyrophosphate formation in aqueous-solution. Tetrahedron Lett 32:6905–6906
 78. Kadokura M, Wada T, Urashima C, Sekine M (1997) Efficient synthesis of gamma-methyl-capped guanosine 5'-triphosphate as a 5'-terminal unique structure of U6 RNA via a new triphosphate bond formation involving activation of methyl phosphorimidazolide using ZnCl₂ as a catalyst in DMF under anhydrous conditions. Tetrahedron Lett 38:8359–8362
 79. Kowalska J, Lewdorowicz M, Darzynkiewicz E, Jemielity J (2007) A simple and rapid synthesis of nucleotide analogues containing a phosphorothioate moiety at the terminal position of the phosphate chain. Tetrahedron Lett 48:5475–5479
 80. Kowalska J, Lewdorowicz M, Zuberek J, Grudzien-Nogalska E, Bojarska E, Stepinski J, Rhoads RE, Darzynkiewicz E, Davis RE, Jemielity J (2008) Synthesis and characterization of mRNA cap analogs containing phosphorothioate substitutions that bind tightly to eIF4E and are resistant to the decapping pyrophosphatase DcpS. RNA-a Publ RNA Soc 14:1119–1131
 81. Strenkowska M, Grzela R, Majewski M, Wnek K, Kowalska J, Lukaszewicz J, Zuberek J, Darzynkiewicz E, Kuhn AN, Sahin U et al (2016) Cap analogs modified with 1,2-dithiodiphosphate moiety protect mRNA from decapping and enhance its translational potential. Nucl Acid Res 44:9578–9590
 82. Kowalska J, Lukaszewicz M, Zuberek J, Darzynkiewicz E, Jemielity J (2009) Phosphoroselenoate dinucleotides for modification of mRNA 5' end. ChemBioChem 10:2469–2473
 83. Kowalska J, Wypijewska del Nogal A, Darzynkiewicz ZM, Buck J, Nicola C, Kuhn AN, Lukaszewicz M, Zuberek J, Strenkowska M, Ziemniak M et al (2014) Synthesis, properties, and biological activity of boranophosphate analogs of the mRNA cap: versatile tools for manipulation of therapeutically relevant cap-dependent processes. Nucl Acid Res 42:10245–10264
 84. Baranowski MR, Nowicka A, Rydzik AM, Warminski M, Kasprzyk R, Wojtczak BA, Wojcik J, Claridge TDW, Kowalska J, Jemielity J (2015) Synthesis of fluorophosphate nucleotide analogues and their characterization as tools for ¹⁹F NMR studies. J Org Chem 80:3982–3997

85. Rydzik AM, Kulis M, Lukaszewicz M, Kowalska J, Zuberek J, Darzynkiewicz ZM, Darzynkiewicz E, Jemielity J (2012) Synthesis and properties of mRNA cap analogs containing imidodiphosphate moiety—fairly mimicking natural cap structure, yet resistant to enzymatic hydrolysis. *Bioorg Med Chem* 20:1699–1710
86. Walczak S, Nowicka A, Kubacka D, Fac K, Wanat P, Mroczek S, Kowalska J, Jemielity J (2017) A novel route for preparing 5' cap mimics and capped RNAs: phosphate-modified cap analogues obtained via click chemistry. *Chem Sci* 8:260–267
87. Kalek M, Jemielity J, Darzynkiewicz ZM, Bojarska E, Stepinski J, Stolarski R, Davis RE, Darzynkiewicz E (2006) Enzymatically stable 5' mRNA cap analogs: synthesis and binding studies with human DcpS decapping enzyme. *Bioorg Med Chem* 14:3223–3230
88. Kowalska J, Osowniak A, Zuberek J, Jemielity J (2012) Synthesis of nucleoside phosphosulfates. *Bioorg Med Chem Lett* 22:3661–3664
89. Strenkowska M, Wanat P, Ziemniak M, Jemielity J, Kowalska J (2012) Preparation of synthetically challenging nucleotides using cyanoethyl P-imidazolides and microwaves. *Org Lett* 14:4782–4785
90. Kuhn AN, Diken M, Kreiter S, Selmi A, Kowalska J, Jemielity J, Darzynkiewicz E, Huber C, Tureci O, Sahin U (2010) Phosphorothioate cap analogs increase stability and translational efficiency of RNA vaccines in immature dendritic cells and induce superior immune responses in vivo. *Gene Ther* 17:961–971
91. Ziemniak M, Kowalska J, Lukaszewicz M, Zuberek J, Wnek K, Darzynkiewicz E, Jemielity J (2015) Phosphate-modified analogues of m(7)GTP and m(7)Gppppm(7)G-Synthesis and biochemical properties. *Bioorg Med Chem* 23:5369–5381
92. Jemielity J, Kowalska J, Rydzik AM, Darzynkiewicz E (2010) Synthetic mRNA cap analogs with a modified triphosphate bridge: synthesis, applications and prospects. *New J Chem* 34:829–844
93. Rydzik AM, Lukaszewicz M, Zuberek J, Kowalska J, Darzynkiewicz ZM, Darzynkiewicz E, Jemielity J (2009) Synthetic dinucleotide mRNA cap analogs with tetraphosphate 5',5' bridge containing methylenebis(phosphonate) modification. *Org Biomol Chem* 7:4763–4776
94. Zytek M, Kowalska J, Lukaszewicz M, Wojtczak BA, Zuberek J, Ferenc-Mrozek A, Darzynkiewicz E, Niedzwiecka A, Jemielity J (2014) Towards novel efficient and stable nuclear import signals: synthesis and properties of trimethylguanosine cap analogs modified within the 5',5'-triphosphate bridge. *Org Biomol Chem* 12:9184–9199
95. Honcharenko M, Bestas B, Jezowska M, Wojtczak BA, Moreno PMD, Romanowska J, Bachle SM, Darzynkiewicz E, Jemielity J, Smith CIE et al (2016) Synthetic m(3)G-CAP attachment necessitates a minimum trinucleotide constituent to be recognised as a nuclear import signal. *RSC Adv* 6:51367–51373
96. Wojtczak BA, Warminski M, Kowalska J, Lukaszewicz M, Honcharenko M, Smith CIE, Stromberg R, Darzynkiewicz E, Jemielity J (2016) Clickable trimethylguanosine cap analogs modified within the triphosphate bridge: synthesis, conjugation to RNA and susceptibility to degradation. *RSC Adv* 6:8326–8337
97. Baranowski MR, Nowicka A, Jemielity J, Kowalska J (2016) A fluorescent HTS assay for phosphohydrolases based on nucleoside 5'-fluorophosphates: its application in screening for inhibitors of mRNA decapping scavenger and PDE-I. *Org Biomol Chem* 14:4595–4604
98. McEwen DP, Gee KR, Kang HC, Neubig RR (2001) Fluorescent BODIPY-GTP analogs: real-time measurement of nucleotide binding to G proteins. *Anal Biochem* 291:109–117
99. Kasprzyk R, Kowalska J, Wieczorek Z, Szabelski M, Stolarski R, Jemielity J (2016) Acetylpyrene-labelled 7-methylguanine nucleotides: unusual fluorescence properties and application to decapping scavenger activity monitoring. *Org Biomol Chem* 14:3863–3868
100. Grudzien-Nogalska E, Stepinski J, Jemielity J, Zuberek J, Stolarski R, Rhoads RE, Darzynkiewicz E (2007) Synthesis of anti-reverse cap analogs (ARCA) and their applications in mRNA translation and stability. *Transl Initiati: Cell Biol High-Through Method Cheml Based Approach* 431:203–227
101. Peng ZH, Sharma V, Singleton SF, Gershon PD (2002) Synthesis and application of a chain-terminating dinucleotide mRNA cap analog. *Org Lett* 4:161–164
102. Jemielity J, Fowler T, Zuberek J, Stepinski J, Lewdorowicz M, Niedzwiecka A, Stolarski R, Darzynkiewicz E, Rhoads RE (2003) Novel “anti-reverse” cap analogs with superior translational properties. *RNA-a Publ RNA Soc* 9:1108–1122
103. Kore AR, Shanmugasundaram M, Charles I, Vlassov AV, Barta TJ (2009) Locked nucleic acid (LNA)-modified dinucleotide mRNA cap analogue: synthesis, enzymatic incorporation, and utilization. *J Am Chem Soc* 131:6364–6365

104. Kore AR, Charles I (2010) Synthesis and evaluation of 2'-O-allyl substituted dinucleotide cap analog for mRNA translation. *Bioorg Med Chem* 18:8061–8065
105. Kore AR, Charles I, Shanmugasundaram M (2009) Synthesis and application of 2'-ara-fluoroguanosine-substituted cap analog. *Chem Lett* 38:652–653
106. Shanmugasundaram M, Charles I, Kore AR (2016) Design, synthesis and biological evaluation of dinucleotide mRNA cap analog containing propargyl moiety. *Bioorg Med Chem* 24:1204–1208
107. Huang FQ (2003) Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucl Acid Res* 31:8
108. Grudzien-Nogalska E, Jemielity J, Kowalska J, Darzynkiewicz E, Rhoads RE (2007) Phosphorothioate cap analogs stabilize mRNA and increase translational efficiency in mammalian cells. *RNA-a Publ RNA Soc* 13:1745–1755
109. Piecyk K, Davis RE, Jankowska-Anyszka M (2012) 5'-Terminal chemical capping of spliced leader RNAs. *Tetrahedron Lett* 53:4843–4847
110. Zuberek J, Wyslouch-Cieszynska A, Niedzwiecka A, Dadlez M, Stepinski J, Augustyniak W, Gingras AC, Zhang ZB, Burley SK, Sonenberg N et al (2003) Phosphorylation of eIF4E attenuates its interaction with mRNA 5' cap analogs by electrostatic repulsion: intein-mediated protein ligation strategy to obtain phosphorylated protein. *RNA-a Publ RNASoc* 9:52–61
111. Hofer K, Abele F, Schlotthauer J, Jaschke A (2016) Synthesis of 5'-NAD-Capped RNA. *Biochem Chem* 27:874–877
112. Sawai H, Wakai H, Nakamura-Ozaki A (1999) Synthesis and reactions of nucleoside 5'-diphosphate imidazolide. A nonenzymatic capping agent for 5'-monophosphorylated oligoribonucleotides in aqueous solution. *J Org Chem* 64:5836–5840
113. Veliath E, Gaffney BL, Jones RA (2014) Synthesis of capped RNA using a DMT group as a purification handle. *Nucleosides Nucleotides Nucl Acid* 33:40–52
114. Thillier Y, Decroly E, Morvan F, Canard B, Vasseur JJ, Debart F (2012) Synthesis of 5' cap-0 and cap-1 RNAs using solid-phase chemistry coupled with enzymatic methylation by human (guanine-N-7)-methyl transferase. *RNA-a Publ RNA Soc* 18:856–868
115. Kadokura M, Wada T, Seio K, Moriguchi T, Huber J, Luhrmann R, Sekine M (2001) Solid-phase synthesis of a 5'-terminal TMG-capped trinucleotide block of U1 snRNA. *Tetrahedron Lett* 42:8853–8856
116. Jemielity J, Heinonen P, Lonnberg H, Darzynkiewicz E (2005) A novel approach to solid phase chemical synthesis of oligonucleotide mRNA cap analogs. *Nucleosides Nucleotides Nucl Acids* 24:601–605
117. Ohkubo A, Sasaki K, Noma Y, Tsunoda H, Seio K, Sekine M (2009) Efficient solid-phase synthesis of oligodeoxynucleotides having a 5'-terminal 2,2,7-trimethylguanosine pyrophosphate linkage. *Bioorg Med Chem* 17:4819–4824
118. Ohkubo A, Kondo Y, Suzuki M, Kobayashi H, Kanamori T, Masaki Y, Seio K, Nagai K, Sekine M (2013) Chemical synthesis of U1 snRNA derivatives. *Org Lett* 15:4386–4389
119. Gampe CM, Hollis-Symynkywicz M, Zecri F (2016) Covalent chemical 5'-functionalization of RNA with Diazo Reagents. *Angew Chem Int Ed* 55:10283–10286