IUPAC Recommendations

G. Michael Blackburn*, Jacqueline Cherfils, Gerard P. Moss*, Nigel G. J. Richards, Jonathan P. Waltho, Nicholas H. Williams and Alfred Wittinghofer

How to name atoms in phosphates, polyphosphates, their derivatives and mimics, and transition state analogues for enzyme-catalysed phosphoryl transfer reactions (IUPAC Recommendations 2016)

DOI 10.1515/pac-2016-0202
Received February 4, 2016; accepted January 12, 2017

Abstract: Procedures are proposed for the naming of individual atoms, P, O, F, N, and S in phosphate esters, amidates, thiophosphates, polyphosphates, their mimics, and analogues of transition states for enzyme-catalyzed phosphoryl transfer reactions. Their purpose is to enable scientists in very different fields, *e.g.* biochemistry, biophysics, chemistry, computational chemistry, crystallography, and molecular biology, to share standard protocols for the labelling of individual atoms in complex molecules. This will facilitate clear and unambiguous descriptions of structural results, as well as scientific intercommunication concerning them. At the present time, perusal of the Protein Data Bank (PDB) and other sources shows that there is a limited degree of commonality in nomenclature, but a large measure of irregularity in more complex structures. The recommendations described here adhere to established practice as closely as possible, in particular to IUPAC and IUBMB recommendations and to "best practice" in the PDB, especially to its atom labelling of amino acids, and particularly to Cahn-Ingold-Prelog rules for stereochemical nomenclature. They are designed to work in complex enzyme sites for binding phosphates but also to have utility for non-enzymatic systems. Above all, the recommendations are designed to be easy to comprehend and user-friendly.

Keywords: atom names for transition states; naming phosphate transition states; P, O, F and N atom labels; phosphate analogues; phosphate atom labels; phosphate nomenclature; phosphate stereochemical naming; phosphoryl transfer; polyphosphates; recommendations.

CONTENTS

Article note: This manuscript (PAC-REC-16-02-02) was prepared in the framework of IUPAC project 2013-039-2-300.

***Corresponding authors: G. Michael Blackburn,** Department of Molecular Biology, Krebs Institute, University of Sheffield, S10 2TN, UK, e-mail: g.m.blackburn@sheffield.ac.uk; and **Gerard P. Moss,** Queen Mary University of London, School of Biological and Chemical Sciences, London E1 4NS, UK, e-mail: g.p.moss@qmul.ac.uk
Jacqueline Cherfils: Laboratoire de Biologie et Pharmacologie Appliquée, CNRS – École Normale Supérieure Paris-Saclay, Cachan, France. http://orcid.org/0000-0002-8966-3067
Nigel G. J. Richards: Department of Chemistry, Indiana University Purdue University Indianapolis, IL 46202, USA; and School of Chemistry, Cardiff University, Cardiff CF10 3AT, UK
Jonathan P. Waltho: Biosciences Institute, University of Manchester, M1 7DN, UK. http://orcid.org/0000-0002-7402-5492
Nicholas H. Williams: Chemistry Department, University of Sheffield, Sheffield S3 7HF, UK
Alfred Wittinghofer: Group for Structural Biology, Max-Planck-Institut für Molekulare Physiologie, 44227 Dortmund, Deutschland

1 Introduction

The advent of stereochemical studies on phosphate esters and diesters with particular reference to their enzyme catalysed reactions, initially through the work of Knowles [1] and of Lowe [2], placed new demands on the nomenclature of the oxygen atoms of the transferring phosphoryl group, PO_3^-.[1] In early work employing thiophosphates made chiral by the specific introduction of oxygen-18 paired with oxygen-16, the direct application of Cahn-Ingold-Prelog (CIP) Rules for prochirality [3, 4] resolved the problem by labelling the oxygen atoms (*R*p) and (*S*p), as appropriate [5–7]. The more advanced use of ^{16}O, ^{17}O, and ^{18}O bonded to the same phosphorus [8] led to the concept of *pro-pro-pro*-chirality at phosphorus, which was still capable of CIP identification [2, 8]. However, such isotopic labelling is experimentally demanding and not necessarily applicable to stereochemical problems now more readily amenable to analysis through advances in protein crystallography. The increasing frequency of binary and tertiary structures of proteins in complex with phosphate ester substrates and/or analogues has enabled a rapidly expanding number of enzyme catalysed reactions to be investigated by structural and computational methods [9–11]. Indeed, there are now over 1600 ligands in the PDB having a phosphoryl group component. These are associated with over 28 000 deposited structures. While many of these structures can be, and have been, labelled for their phosphorus and phosphoryl oxygen atoms through current practice, comparative studies of related structures easily identify multiple inconsistencies in labelling that arise from variable methods of naming N, O, and P atoms.

This situation has become increasingly complex as a result of the introduction and development of metal fluoride (MF_x) analogues of the PO_3^- group in studies on transition state analogues (TSA) for phosphoryl transfer enzymes. Trifluoroberyllate (BeF_3^-)[2] is a ground state analogue for phosphate, with characteristic tetrahedral geometry when ligated to anionic oxygen. Tetrafluoroaluminate (AlF_4^-)[3] is a mimic for concerted phosphoryl transfer in multiple enzymes, though it has octahedral geometry. Aluminium trifluoride (AlF_3)[4] forms trigonal bipyramidal (tbp) TSA complexes that have the correct 3-D geometry for concerted PO_3^- group transfer but lack its anionic charge. These two values converge in the relatively smaller number of trifluoromagnesate complexes (MgF_3^-),[5] which are both anionic and have tbp geometry. Indeed, some of the AlF_3 complexes have been shown in reality to be MgF_3^- complexes in solution [12–14]. The growth in use of these four types of MF_x complexes is illustrated in Fig. 1. In addition, there are many significant structures of phosphoryl transfer enzyme complexes that include vanadium(V) or tungsten(VI) complexes, either as tetrahedral phosphate mimics or as tbp mimics of transition states. The relative growth in use of these six species is presented in Fig. 1. The double change from four coordinate tetrahedral PO_4 to five coordinate tbp O–MF_3–O and six

1 The description of the trigonal planar monoanionic PO_3 species as a phosphoryl group is in line with long-established usage in biochemistry, biomolecular chemistry, and molecular biology. We note that the IUPAC recommendation is that the species 'phosphoryl group' relates to the neutral, diatomic trivalent species ≡PO.

2 BeF_3^-; PDB ligand name, Beryllium trifluoride ion; PDB code, **BEF**; IUPAC name, Trifluoridoberyllate.

3 AlF_4^-; PDB ligand code: **ALF**; IUPAC name Tetrafluoridoaluminate.

4 AlF_3; PDB and IUPAC, name Aluminum fluoride; PDB ligand code: **AF3**.

5 MgF_3^-; PDB ligand code, **MGF**; IUPAC name, Trifluoridomagnesate.

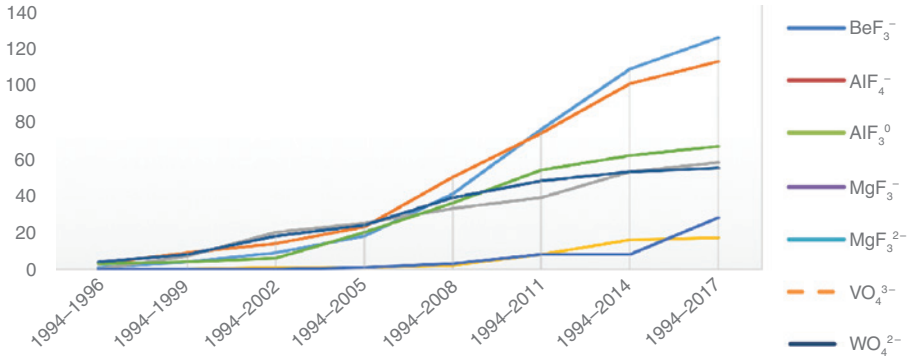

Fig. 1: Cumulative numbers of protein structures published in the PDB in successive triennia that contain ligands which are analogues of phosphoryl groups or their transition states (*data for mid-2016 extrapolated to 2017*).

coordinate octahedral O–MF_4–O complexes adds a new dimension to the problem of the atomic description of these complexes. The need to solve this general problem provided the principal motivation for the development of these standardized naming conventions.

As the development of our protocols progressed, it became apparent to us that a rational, logical set of labels for the 5- and 6-coordinate systems described above could only be established on the basis of a clear definition of the systematic labelling of phosphorus atoms in standard multiple phosphate molecules, that already extends to eight in the case of hexaphosphoinositol bisphosphates [15]. It needed to be followed by a comprehensive system for oxygen atom labelling to include both bridge and non-bridge atoms in linear chains of phosphates, as for the 13 oxygen atoms of 5′-adenosyl 5‴-guanosyl P^1, P^4-tetraphosphate [16] and the 3 non-isotopically identifiable oxygen atoms of the PO_3^- group of terminal phosphates. With those objectives accomplished, our recommendations could then be developed to incorporate both the fluorine ligands of MF_x systems and the oxygen atoms of vanadate and tungstate analogues of phosphates and their TSAs.

The basic strategy of the recommendations is built on the recognition that a phosphate monoester comprises an alkoxy group and a phosphoryl group ($ROH + PO_3^-$), a monoalkyl diphosphate comprises a phosphate monoester and a second phosphoryl group ($ROPO_3^- + PO_3^-$), a monoalkyl triphosphate comprises a monoalkyl diphosphate and a third phosphoryl group, and so on. For simplicity, we have ignored anionic charges on phosphoryl oxygen atoms and we have treated P=O "double bonds" as P–O single bonds because there is no π-bonding in the phosphoryl group. While we do not seek to claim that our coverage has been exhaustive, we believe that the principles for naming atoms set out here will prove generally applicable to all cognate molecular species which share a geometrical relationship to phosphates, *e.g.* sulfates, perchlorates, *etc.*

Lastly, we provide an Appendix as a simple guide to the application of Cahn-Ingold-Prelog Rules to label prochiral, non-bridge oxygen atoms in molecules under inspection.

2 Existing recommendations

Phosphorus nomenclature and related IUPAC Recommendations

a) The nomenclature of phosphorus-containing compounds of biochemical importance, Recommendations 1976, was published in 1977 [17]. It was concerned with the naming of compounds, but did not consider the identification of the individual atoms of the phosphate or polyphosphate groups, other than to label the phosphates of a nucleoside triphosphate α, β, and γ. It did cover the naming of polyphosphates where a bridging oxygen is replaced by a methylene or imino group. A variation on this was proposed in 1980 and revised in 1992 [18].

b) A document on the abbreviations and symbols for the description of conformation of polynucleotide chains, Recommendations 1982, was published in 1983 [19]. In a related paper, it was proposed that the

pro-S oxygen should be OP1 and the *pro-R* should be OP2 [20]. This is the *reverse* of the system proposed here and it is also contrary to CIP nomenclature, where Rule 5 gives priority to *R* over *S*. We have chosen to adhere to CIP priority Rule 5.

c) IUPAC Recommendations for preferred names of derivatives of phosphoric acid are pertinent [21]. They include the application of the CIP rules to chiral phosphates as well as CIP rules for a trigonal bipyramidal and octahedral systems. These are also described in IUPAC inorganic chemistry nomenclature systems for bipyramidal and octahedral structures [22].

3 Recommendations for labelling phosphorus atoms in phosphates

3.1 A. Labelling phosphorus atoms in polyphosphate species

3.1.1 A1. Species with one single polyphosphate chain

This requires a one-symbol code to describe the position of each phosphorus in a single chain of phosphates. Phosphorus descriptions use a capital letter that serves to discriminate sequential phosphorus atoms in the same chain (PDB usage).

a) Phosphorus atoms are named in progression from the RO- end as P^A, P^B, P^G, P^D, *etc.*[6] Hence, adenosine 5′-tetraphosphate (PDB ligand: **AQP**) has phosphorus atoms labelled as P^A, P^B, P^G, P^D starting from the ribose 5′-oxygen (Fig. A1a).

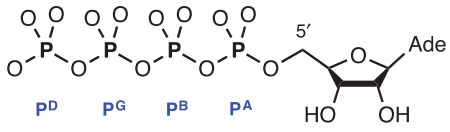

Fig. A1a

b) For P^1-(5′-adenosyl)-P^5-(5‴-guanosyl) pentaphosphate (PDB Ligand: **G5P**), phosphorus atoms should be named P^A, P^B, P^G, P^D, and P^E starting at the 5′-oxygen of the adenosine.

c) For the RO- group at the end of a phosphate chain, a nucleoside takes priority over a non-nucleoside. Thus, in uridine diphosphate glucose (PDB ligand: **UPG**), P^A is bonded to uridine-O5′ and P^B is bonded to O1″ of glucose (Fig. A1b).[7]

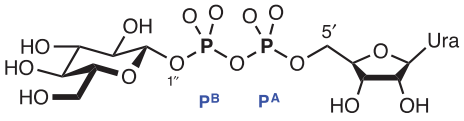

Fig. A1b

d) A nucleic acid base takes priority over non-nucleic acid base (*i.e.* adenosine > nicotinamide riboside). Thus, in NAD^+ (PDB ligand: **NAD**), P^A is bonded to O5′ of adenosine, with P^B bonded to O5‴ of the nicotinamide riboside (Fig. A1c).

6 PDB usage currently always replaces Pγ with P^G as it does not use a Greek/Symbol font.

7 Here, and throughout, negative charges on phosphates and P=O double bonds are omitted for simplicity.

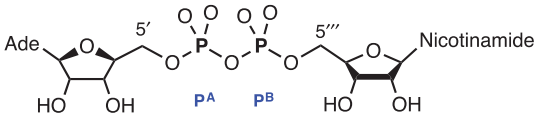

Fig. A1c

e) Nucleosides take priority in their alphabetical order (A > C > G > dT > U). Thus, in P^1-(5′-adenosyl) P^4-(5‴-deoxythymidyl) tetraphosphate (Ap_4dT) (PDB ligand: **4TA**), the phosphorus atoms should be named P^A, P^B, P^G, P^D, starting at the 5′-oxygen of the adenosine (Fig. A1d).

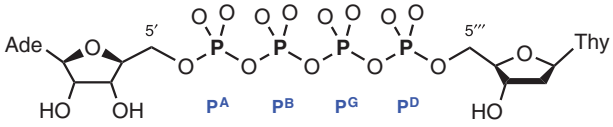

Fig. A1d

f) Pentoses use CIP priorities: D-ribose > L-ribose > 2-deoxy-D-ribose > 2-deoxy-L-ribose.[8] Thus, a transition state for dAMP kinase should label the four phosphorus atoms P^A, P^B, P^G, P^D, starting from the adenosine 5′-oxygen (Fig. A1e).

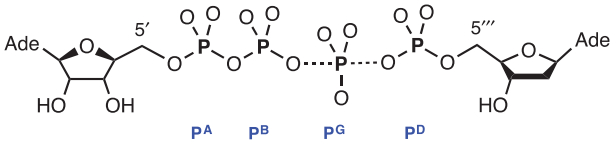

Fig. A1e

g) In phosphonate and phosphoramidate analogues of polyphosphates, phosphorus atoms will be labelled in the same manner as for the parent polyphosphate molecule. Hence, for β,γ-methylene-GTP (PDB ligand: **GCP**), phosphorus atoms should be named P^A, P^B and P^G from the 5′-oxygen (Fig. A1f1).

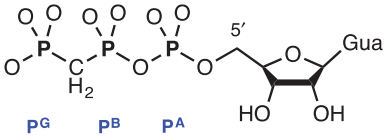

Fig. A1f1

Likewise, for 2′-deoxyuridine 5′-α,β-imidotriphosphate (PDB ligand: **DUP**), phosphorus atoms should be named P^A, P^B and P^G from the 5′-oxygen (Fig. A1f2).

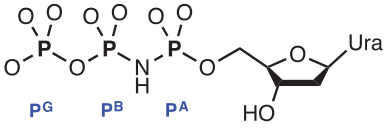

Fig. A1f2

8 This pentose order approximates to CIP Rule 5 priority (*R*) > (*S*). This rule will apply particularly to transition states for deoxynucleotide kinases, *e.g.* where ATP phosphorylates dAMP.

3.1.2 A2. Species with multiple single phosphate chains

This requires a one-symbol code to describe the relationship of each phosphate chain to the parent molecule.

a) Inositol polyphosphates require a phosphorus label derived from the identity of the oxygen to which each single phosphate is attached. Thus, for *myo*-inositol 1,3,4,5,6-pentakis phosphate (InsP5) (PDB ligand: **5MY**), the phosphorus atoms should be labelled P^1, P^3, P^4, P^5, and P^6 (Fig. A2a1).[9] For fructose 1,6-bisphosphate (PDB label: **FBP**), the phosphorus atoms should be labelled P^1 and P^6 (Fig. A2a2).

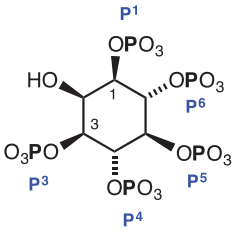

Fig. A2a1

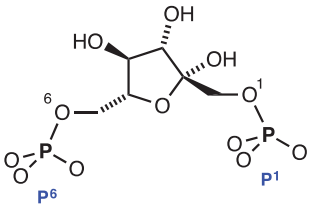

Fig. A2a2

3.1.3 A3. Species with multiple single phosphate and/or polyphosphate chains

This requires a two-symbol code to describe (i) the position of each phosphorus in a single chain of phosphates, and (ii) the relationship of that phosphate chain to the parent molecule.

a) Species with polyphosphates located on multiple oxygen atoms require a two-symbol code to designate their phosphorus atoms: a numerical code for the oxygen bridging to the parent molecule and an alphabetic code for the position of the phosphorus in the phosphate chain. Thus, in pppGpp (PDB ligand: **0O2**), the 5′-phosphorus atoms should be named P^{A5}, P^{B5}, and P^{G5}, and the 3′-phosphorus atoms named P^{A3} and P^{B3} (Fig. A3a).

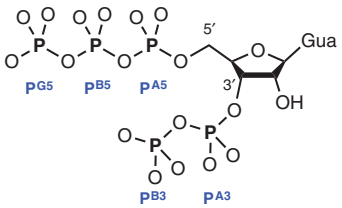

Fig. A3a

b) Inositol polyphosphates having polyphosphate moieties require a two-symbol code to designate their phosphorus atoms. A numerical symbol designates the oxygen to which each single phosphate is attached and an alphabetic code designates the position of the phosphorus in the phosphate chain. In the case of monophosphates, the labels P^1, P^2, *etc.*, should apply to single phosphorus entities, while P^{An} and P^{Bn} will apply to diphosphates, as in PP-InsP_5 (PDB ligand: **I7P**) (Fig. A3b).[4]

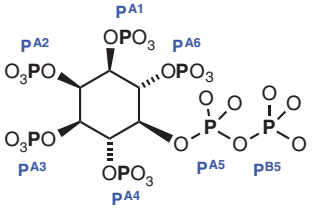

Fig. A3b

9 *cf.* R. F. Irvine & M. J. Schell, *Nature Rev. Molec. Cell Biol.* **2**, 327–338 (2001).

4 Recommendations for labelling oxygen atoms in phosphates

4.1 B1. Non-terminal phosphates in molecules with one single phosphate chain

This requires a two-symbol code to describe (i) the identity of the oxygen relative to its congeners, and (ii) the identity of the parent phosphorus atom. Oxygen codes use a number first, to discriminate oxygen atoms bonded to the same phosphorus, followed by a letter to indicate the parent phosphorus.

a) The oxygen linking P^A to the carbon moiety of the molecule will retain its regular label. Thus, in ATP, O5′ bonds P^A to the ribose (Fig. B1a1). In Ap_4G, O5′ bonds P^A to adenosine, while O5‴ bonds P^D to guanosine (Fig. B1a2) [16].

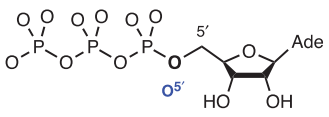

Fig. B1a1

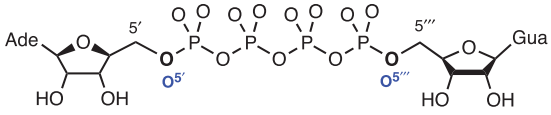

Fig. B1a2

b) In each non-terminal phosphoryl group (PO_3), the bridging oxygen bonding P^X to $P^{(X+1)}$ in the chain should be numbered O^{3X}. Hence, in ATP, O^{3A} joins P^A to P^B, and O^{3B} joins P^B to P^G (Fig. B1b).

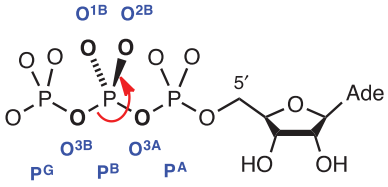

Fig. B1b

c) In each non-terminal phosphoryl group, the two non-bridging oxygen atoms will be labelled 1 and 2 according to their CIP *pro-R*- and *pro-S*-chiralities, respectively.[10] Hence, in ATP, P^A will have non-bridging oxygen atoms O^{1A} and O^{2A} for the *pro-R* and *pro-S* oxygen atoms, respectively (Fig. B1b).

d) Bridging atoms in polyphosphate chains should be given priorities $O^{n\prime} < O^{3A} < O^{3B} < O^{3G} < O^{3D} <$ *etc.* (Section B3). This provides a relatively direct CIP route to the assignment of paired non-bridging oxygens on non-terminal P^B, P^G, *etc.*, that can be illustrated here for O^{1B} and O^{2B} (Fig. B1d).

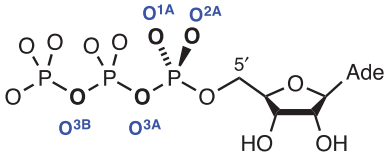

Fig. B1d

e) In chains containing a sulfur atom in a non-bridging, non-terminal position, the sulfur will take the name S^{1A} (for substituent on P^A), S^{1B} (for substituent on P^B), *etc.* The non-bridging oxygen is then named O^{2A},

10 This nomenclature is widely used in the PDB for oxygen atoms on P^A and P^G in nucleoside triphosphates, but is rather variably used for oxygen atoms on P^B.

O^{2B}, *etc.*, and the bridging oxygen is O^{3A}, O^{3B}, *etc.*, as above. This is shown for guanosine 5′-(*R*p)-α-thiotriphosphate (PDB ligand: **GAV**) (Fig. B1e).

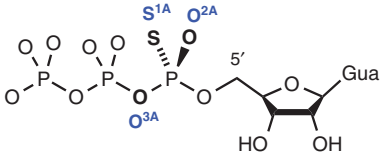

Fig. B1e

In modified polyphosphate chains having two-atom bridges replacing an O^{3N} (where N = A, B, *etc.*) the bridging atoms X and Y will be labelled X^{3A} and Y^{4A}, progressively. Thus in β,γ-oxymethylene-ATP (Ado*P*O*P*OCH_2*P*), the P^B, P^G-bridging atoms are O^{3B} and C^{4B}, respectively (Fig. B1f).

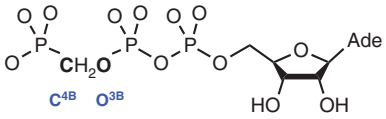

Fig. B1f

In polyphosphate chains with a bridging oxygen replaced by carbon or nitrogen, the prochirality[11] designations may change in consequence. Thus, in α,β-methylene adenosine 5′-triphosphate (PDB ligand: **APC**) (Fig. B1g), oxygen atoms O^{1A} and O^{2A} are reversed relative to their designation in ATP (Fig. B1b) because of the changed priority of C^{3A} relative to $O^{5'}$.

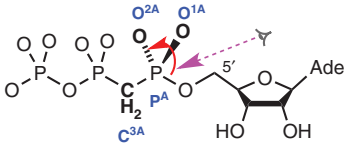

Fig. B1g

4.2 B2. Non-terminal phosphates in molecules with multiple phosphate chains

This requires a three-symbol code, using (i) one symbol to describe the identity of the oxygen relative to its congeners, and (ii) two symbols for the identity of the parent phosphorus atom (see above)

a) In each non-terminal phosphoryl group, the two non-bridging oxygen atoms will be labelled 1 and 2 according to their CIP *pro-R* and *pro-S* chiralities, respectively. Hence, in ppGpp (PDB ligand: G4P), P^{A5} will have non-bridging oxygen atoms O^{1A5} and O^{2A5} for the *pro-R* and *pro-S* oxygen atoms, respectively, and P^{A3} will have non-bridging oxygen atoms O^{1A3} and O^{2A3} for the *pro-R* and *pro-S* oxygen atoms, respectively[12] (Fig. B2a).

b) In NAD^+, the oxygen atoms on P^A will be labelled O^{1A} and O^{2A} for the *pro-R* and *pro-S* oxygen atoms, respectively, and the oxygen atoms on P^B will be labelled O^{1B} and O^{2B} for the *pro-R* and *pro-S* oxygen atoms, respectively (Fig. B2b). (**NB** Ade takes priority over Nicotinamide. Fig. A1d).

11 An Appendix has been added on a simple introduction to the use of CIP Rules on prochirality and the assignment of *pro-R* and *pro-S* descriptions.

12 For simplicity, the designation omits the prime symbol from *e.g.* $O^{2A3'}$.

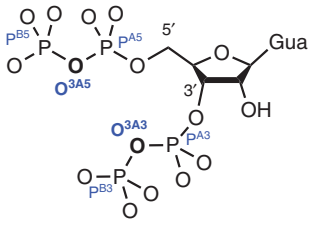

Fig. B2a

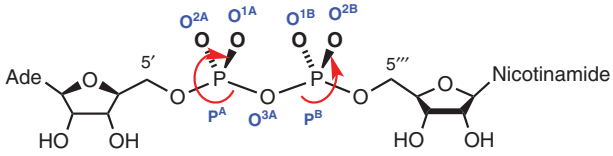

Fig. B2b

c) In ppIns5p, the oxygen atoms on P^{A5} will be labelled O^{1A5} and O^{2A5} for the *pro-R* and *pro-S* oxygen atoms, respectively (Fig. B2c).

d) In each non-terminal phosphoryl group (PO_3), the bridging oxygen bonding P^N to $P^{(N+1)}$ (where N = A, B, *etc.*) in the chain should be numbered O^{3Nx}, where x designates the parent oxygen of the polyphosphate chain. Hence, in ppGpp (PDB ligand: **P4G**), P^{A5} is joined to P^{B5} by O^{3A5}, and P^{A3} is joined to P^{B3} by O^{3A3} (Fig. B2d).

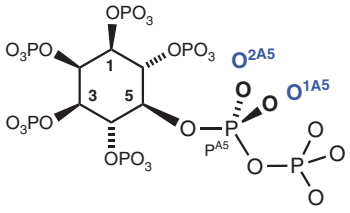

Fig. B2c

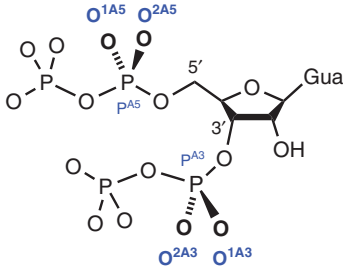

Fig. B2d

4.3 B3. Non-terminal phosphates in molecules with a doubly-capped, single phosphate chain

This requires a two-symbol code to describe (i) the identity of the oxygen relative to its congeners and (ii) one symbol for the identity of the parent phosphorus atom (v.s.).

There are several examples of natural and non-natural dinucleosidyl polyphosphates with phosphate chain lengths from 2 to 6 phosphates. P^1-(5′-adenosyl) P^4-(5‴-deoxythymidyl) tetraphosphate (Ap_4dT) (PDB

ligand: **4TA**) (Fig. A1d) and P^1-(5′-adenosyl)-P^4-(5‴-guanosyl) tetraphosphate (Ap_4G) (Fig. B1a2) have been discussed above. While the accurate, and sometimes tricky, application of CIP rules can deliver an appropriate descriptor for paired non-bridging oxygens, for symmetrical molecules, including Ap_3A and Ap_5A, the oxygens on the central phosphate are stereochemically identical (being related by C2 symmetry). They become a diastereotopic pair only in a chiral environment, as when bound to a protein, for which an additional Rule is needed. The problem can easily be resolved by the application of an additional priority Rule for sequential phosphates in polyphosphate chains.

"Bridging atoms in polyphosphate chains are given priorities that increase with increasing separation from the priority nucleoside oxygen: $O^{n\prime} < O^{3A} < O^{3B} < O^{3G} < O^{3D} < etc.$"
The application of this rule is illustrated for P^1-(5′-adenosyl)-P^3-(5‴-adenosyl) triphosphate, Ap_3A (Fig. B3a). By extension, this rule also provides a relatively direct assignment for the non-terminal phosphates of ATP, p_4A, *etc.* It is further illustrated in the Appendix (Section 9.5).

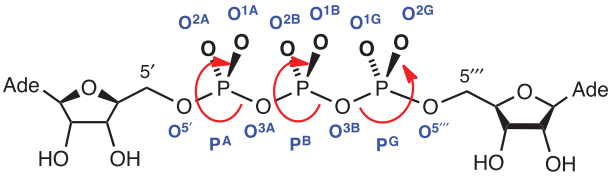

Fig. B3a

4.4 B4. Terminal phosphates in molecules with multiple phosphate chains

This requires a two-symbol code to describe (i) the identity of the oxygen relative to its congeners and (ii) the identity of the parent phosphorus atom (v.s.)*. The three oxygen atoms of a terminal phosphoryl group (PO_3) are* pro-pro-*chiral. They can thus be labelled according to CIP rules in those (rare) cases where they are identified by isotopes ^{16}O, ^{17}O, and ^{18}O.*

a) In cases of a terminal phosphoryl oxygen being replaced by *e.g.* sulfur, fluorine, or nitrogen, the remaining two terminal oxygen atoms are prochiral and can be appropriately identified by CIP chirality rules. Thus, in GTPγS (PDB ligand: **GSP**), the sulfur has priority to be labelled S1G and the oxygen atoms are labelled O^{2G} (*pro-R*) and O^{3G} (*pro-S*), respectively (Fig. B4a).

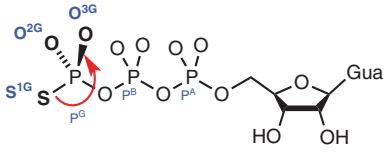

Fig. B4a

b) Prochirality identification can be applied if one of the three oxygen atoms is promoted relative to the other two. In the context of enzyme-bound nucleotides, such promotion can often be identified by coordination of the terminal phosphate to a protein-bound metal ion, typically magnesium. Thus, for ATP bound in many kinases, the γ-phosphate is coordinated from one of its three oxygen atoms to magnesium. This oxygen is thus designated O^{1G}. The remaining oxygen atoms are now prochiral and can be identified in the priority series: $O^{3B} > O^{1G} > O^{2G} > O^{3G}$. CIP rules then designate O^{2G} as the *pro-R* oxygen and O^{3G} as the *pro-S* oxygen, as illustrated for ATP bound in phosphoglycerate kinase (Fig. B4b; PDB entry: **1VJC**).

c) In the absence of metal ion coordination to the terminal phosphate, H-bond donation from amino acids in the protein provides a means of priority identification for O1N. Hydrogen bonds are considered only

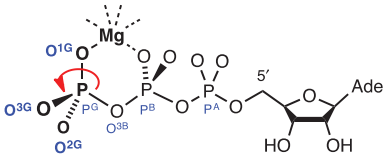

Fig. B4b

if they have a length ≤ 3.0 Å.[13] Priority will be given according to donor atom XH priority with CIP rules (S > O > N). Hydrogen bonding to the amino acid of lowest primary sequence number will identify O^{1G} in ATP, *etc*. If there is still ambiguity in the assignment, then backbone NH takes priority over sidechain NH.[14] This selection makes O^{2G} and O^{3G} prochiral, hence they can be assigned by the application of CIP rules.[15] Thus, in human bisphosphoglycerate mutase (PDB entry: **2A9J**), the 3-phosphoglycerate has phosphoryl oxygen coordination from Arg^{100} and Arg^{116} to O^{1A}, from Arg^{117} and Asn^{190} to O^{2A}, and from Arg^{117} to O^{3A} (Fig. B4c1). After O^{1A} is promoted by amino acid linkage priority, O^{2A} and O^{3A} are assigned by prochirality rules ($O^3 > O^{1A} > O^{2A} > O^{3A}$).

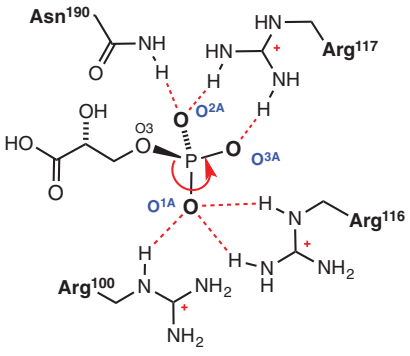

Fig. B4c1

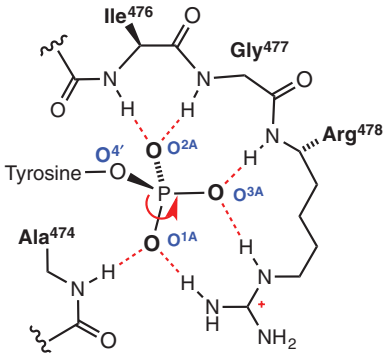

Fig. B4c2

In the case of human protein tyrosine phosphatase **ptpn5** (C472S mutant), the tyrosine phosphate moiety is coordinated to residues in the loop Ala^{474}-Arg^{478} (PDB entry: **2CJZ**). Consideration of hydrogen bonds ≤ 3.0 Å shows oxygen O^{1P} coordinated to Gly^{477} and Ile^{476} oxygen O^{2P} coordinated to Ala^{474} and Arg^{478}, and oxygen O^{3P} coordinated to Arg^{478}. Thus, we can now designate O^{1A} as being coordinated to the lowest numbered amino acid, Ala^{474} (it is labelled as O^{2P} in **2CJZ**).[10] The oxygen atom priority is $O^{4'} > O^{1A} > O^{2A} > O^{3A}$, in which O^{2A} and O^{3A} are designated by CIP rules for prochirality[10] as shown (O^{2A} being *pro-R* and O^{3A} is *pro-S*) (Fig. B4c2). (**NB** There are H-bonds from Ser^{472}(OH) to O^{2P} and O^{3P} but both are longer than 3.0 Å and thus are ignored; distance from heavy atom to heavy atom).

4.5 B5. Terminal phosphates in molecules with multiple phosphate chains

This requires a three-symbol code, (i) one symbol to describe the identity of the oxygen relative to its congeners, and (ii) two symbols for the identity of the parent phosphorus atom (v.s.).

13 The Ångstrom as the unit of distance in protein structures is standard throughout biochemistry, bioorganic chemistry, and molecular biology. The IUPAC use of **nm** as the IS unit of sub-micrometer distance is readily accommodated by the conversion factor 1 nm = 10 Å.

14 In determining priorities, coordination to an isolated water is ignored, because the presence or absence of a particular isolated water in a crystal structure can be a function of the structural resolution achieved, which makes water a variable object. However, waters coordinated to metal ions can be used.

15 For CIP Rules see the IUPAC Blue Book p1162 [21]. For the use of *pro-R* and *pro-S* see "Basic Terminology of Stereochemistry (IUPAC Recommendations 1996)" *Pure Appl. Chem.* **68**, 2193–2222 (1996).

a) The rules described above (section B3) for single phosphate chains will apply with the addition of a descriptor symbol designating the point of attachment of that chain to the parent molecule. Thus, for human aldolase reductase (PDB entry: **2J8T**), the bound $NADP^+$ (PDB ligand: **NAP**) has the oxygen atoms of P^{A2} coordinating no metal and hydrogen bonded to Lys^{262}, Ser^{263}, Val^{264}, Thr^{265}, and Arg^{268}. Thus, the oxygen coordinating Ser^{263} takes priority and is named O^{1A2}. The oxygen atom priorities for P^{A2} are thus $O^2 > O^{1A2} > O^{2A2} > O^{3A}$, as shown (Fig. B5a).

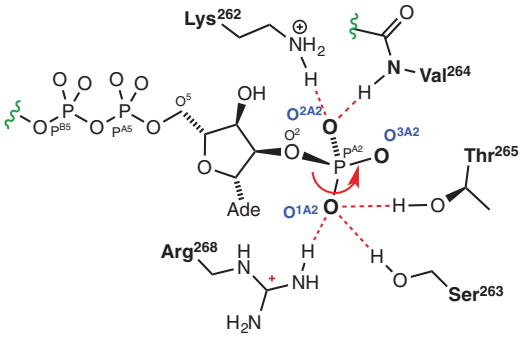

Fig. B5a

4.6 B6. Isolated single phosphates

This requires prioritisation of two oxygen atoms by their coordination features thus allowing the third and fourth oxygen atoms to be assigned their prochirality by CIP rules.

Isolated phosphate with no metal ions. In a structure of the small G protein Rab-5c with GDP and Pi ligands in the catalytic site (PDB entry: **1Z0D**), the isolated phosphate (PDB ligand: **PO4**) is not metal coordinated. Thus, the relative priorities of its 4 oxygen atoms are determined by H-bonds to amino acid residues. Ignoring H-bonds ≥ 3.0 Å, the PDB file assigns O^1 coordinated to Ser^{30}(OH), O^2 coordinated to Leu^{80}(NH), and O^3 coordinated to Lys^{34}(NH_3^+). O^4 is only coordinated to ligands at distances ≥ 3.0 Å (*oxygen atoms numbered as in* **1Z0D**) (Fig. B6a). (*Hence, the PDB priority order is* $O^1 > O^3 > O^2 > O^4$).

Assigning the top two oxygen priorities as O^{1P} and O^{2P}, respectively (Fig. B6b) converts the two remaining oxygen atoms into a prochiral pair. Promoting the 'front' oxygen to ^{18}O gives phosphorus *S* chirality, thus identifying it as *pro-S* (O^{4P}). By a similar analysis, the 'rear' oxygen (H-bonded to Leu^{80}) is *pro-R*. Hence, the rear oxygen is designated O^{3P} and the front oxygen is O^{4P} (Fig. B6b) (**NB** *The PDB file assigns* P^A *and* P^B *to the GDP ligand*).

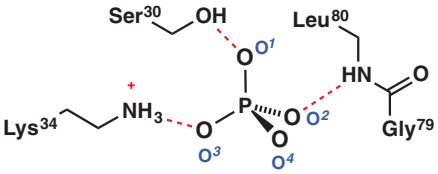

Fig. B6a

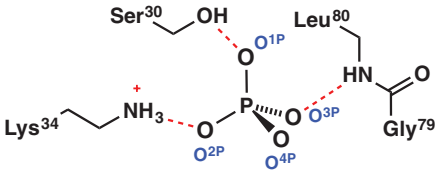

Fig. B6b

5 Recommendations for labelling fluorine and other atoms in phosphate transition state analogues

5.1 C1. Tetrahedral phosphate mimics – trifluoroberyllates

These use a two-symbol code that may be expanded to four when there are additional fluorine atoms in the species.

a) There are over 100 examples of trifluoroberyllates (BeF_3^-) in the PDB (PDB ligand: **BEF**). This phosphate mimic is invariably attached to a carboxylate or terminal phosphate oxyanion. Labelling the three fluorine atoms will follow the same rules as for the three oxygen atoms in a terminal tetrahedral phosphate. Prochirality[10] identification can be applied if one of the three fluorine atoms is promoted relative to the other two. In the context of enzyme-bound trifluoroberyllates, such promotion can be generally identified by co-ordination of one of the fluorine atoms to a protein-bound metal ion, typically magnesium. For example, in β-phosphoglucose mutase (PDB entry: **2WF8**), a BeF_3 is coordinated to Asp^8, while a catalytic magnesium bridges Asp^8 and one fluorine. This fluorine is thus identified as F^{1Be}. The prochiral fluorine atoms F^{2Be} and F^{3Be} are designated by CIP rules, as shown in the example (Fig. C1a). As there is no other fluorine in this structure, these labels can be abbreviated to F^1, F^2, and F^3, respectively.
(*Note that in PDB entry **2WF8**, these fluorine atoms were labelled F3, F1, and F2, respectively*).

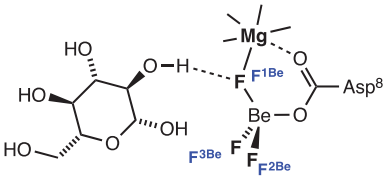

Fig. C1a

There is one example of a BeF_2 moiety bridging two anionic oxygen atoms. In this case, F^{1Be} and F^{2Be} will correspond to the (*pro-R)* and *(pro-S)* stereochemistry assigned by CIP rules. Thus, in UMPCMP kinase (PDB entry: **4UKD**), BeF_2 bonds to ADP O^{3B} and to UDP O^{3G} (Fig. C1b). Thus, the (*pro-R)* fluorine is F^{1Be} and the (*pro-S*) fluorine is F^{2Be}.[16] In this unique and rather complicated example, CIP rules give priority to $O^{5'}$ over $O^{5'''}$, since adenine (A) takes priority over uridine (U) (Section A1d).

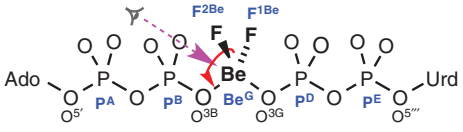

Fig. C1b

b) There may be less common species where there is no metal ion coordinating the trifluoroberyllate. In these, the hydrogen bonding priorities set out in **B3c** can be applied.

c) In an example of multiple metal coordination where the distances of separation from both metals to fluorine are less than the sum of the two van der Waals radii, the coordinating metal with higher atomic number will take priority.

5.2 C2. Trigonal bipyramidal phosphate transition state analogues – trifluoromagnesates and aluminum trifluorides

This requires a two-symbol code to describe (i) the identity of the fluorine relative to its congeners, and (ii) the identity of the core metal ion.

a) For AlF_3 (PDB code: **AF3**), MgF_3 (PDB code: **MGF**), and ScF_3 tbp transition state analogues (TSA), the three fluorine atoms are invariably equatorial with two axial oxygen ligands to the 5-coordinate metal. Priority

16 In the case of an (as yet unidentified) symmetrical species, the priority of the two equivalent fluorine atoms will be based on ligand coordination, as shown in Sections C2b and C3c below.

identification can be applied when one of the three fluorine atoms is promoted relative to the other two and directional priority for the two axial ligands is established. In the context of enzyme-bound trifluoromagnesates and aluminates, such promotion is readily identified by closest proximity of one fluorine to a protein-bound metal ion, typically a catalytic magnesium. The direction of viewing is determined by CIP priority of one of the apical (oxygen) atoms over the second, viewing down the priority O-metal bond. Thus, in the small G protein, Ras (PDB entry: **1OW3**), MgF_3 is axially coordinated to GDP via O^{3B} and to a water, and CIP priority gives $O^{3B} > OH_2$. Thus, the fluorine coordinated to the catalytic magnesium is designated F^{1Mg}. F^{2Mg} and F^{3Mg} are then identified in a clockwise progression from F^{1Mg} when viewed from O^{3B} to Mg (Fig. C2a).[17]

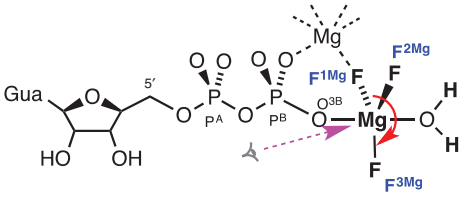

Fig. C2a

b) In case of multiple metal ion coordination where both distances of separation are less than the sum of the two van der Waals radii, the coordinating metal with highest atomic number will take priority. In cases where two fluorine atoms are coordinated to two equivalent metals, as for cAPK (PDB entry: **1L3R**), in which the tbp complex of ADP · MgF_3 is liganded to two catalytic magnesium atoms, F1Mg is prioritised as the fluorine coordinated to the magnesium of higher priority. Metal priority shall be determined by its amino acid coordination (see Section B3c). Viewing priority is determined by: $O^{3B} > O\text{-}Ser^{21\prime}$. In cAPK, one catalytic magnesium is coordinated to Asn^{171}, to Asp^{184}, and to a water; the second magnesium is coordinated to Asp^{184} and to two waters. Hence, the magnesium linked to Asn^{171} has priority and is coordinated to F^{1Mg}; F^{2Mg}, and F^{3Mg} follow in clockwise progression (Fig. C2b).

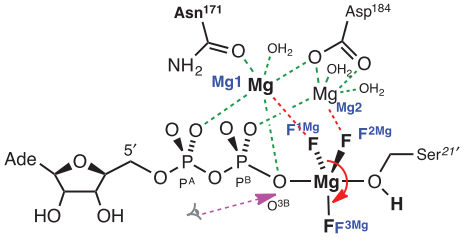

Fig. C2b

c) In the absence of fluorine coordination to a metal, hydrogen bonding to amino acids can be used to determine fluorine priority (see section B4c).[18]

d) A significant number of structures in the PDB (> 24) have a trigonal bipyramidal complex assigned as tetrafluoromagnesate$^{(2-)}$ (PDB ligand: **MF4**). The best resolved of these (PDB entry: **1WPG**, 2.30 Å resolution) has electron density and bond lengths that can be equally well assigned as a regular $Asp^{351}\text{-}CO_2^- \cdot MgF_3^- \cdot OH_2$ complex. This can be labelled as for C2a (above) using coordination to a catalytic Mg to give priority to F.[19]

17 **NB** These fluorine atoms are labelled F2, F1, and F3 respectively in PDB entry: **1OW3** (*Viewing indicated by magenta arrow).*

18 No example of a tbp complex of AF3 or MGF (PDB ligand identities for AlF_3 and MgF_3^- respt.) having a coordinating divalent metal at good resolution has been lodged in the PDB prior to December 2015).

19 No analytical work has been yet presented to identify the number of fluorides, *e.g.* by ^{19}F NMR.

5.3 C3. Octahedral phosphate transition state mimics

This requires a two-symbol code to describe (i) the identity of the fluorine relative to its congeners and (ii) the identity of the core metal ion.

a) For tetrafluoroaluminate, AlF_4^- octahedral TSA analogues (PDB ligand: **ALF**), the four fluorine atoms are invariably equatorial with two *trans*-oxygen ligands to the 6-coordinate aluminium. Priority identification can be applied by promoting one of the four fluorine atoms relative to the other three. In the context of enzyme-bound tetrafluoroaluminates, such promotion is invariably identified by closest proximity of one fluorine to a protein-bound metal ion, usually magnesium. The direction of viewing is determined by CIP priority of one of the apical oxygen atoms over the second and viewing down the priority O-metal bond. Thus, in the structure of βPGM (PDB entry: **4C4R**), the fluorine coordinated to the catalytic magnesium is identified as F^{1Al}, while F^{2Al}, F^{3Al}, and F^{4Al} follow in clockwise progression viewed from Asp^8–O^{d1}, which has priority over glucose-O^6. (**NB** *The corresponding PDB designations are F2, F1, F3, and F4, respectively*)[20] (Fig. C3a).

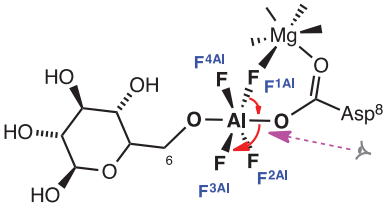

Fig. C3a

b) There are at least 3 examples of octahedral trifluoroaluminate complexes (in PDB to 2015) having three fluorines in equatorial positions with the fourth equatorial ligand identified as oxygen. An example of this is the transition state analogue for enzymatic hydrolysis of dUTP (PDB entry: **4DL8**). Axial priority for viewing is established by the CIP precedence of $O^{3A} > OWat^{401}$ (Fig. C3b). One fluorine is coordinated to two catalytic magnesium atoms and so is designated F^{1B}. A progression viewed in the priority direction then identifies the bridging oxygen as the second priority ligand, O^{1B}, with F^{2B} and F^{3B} completing the clockwise equatorial sequence.

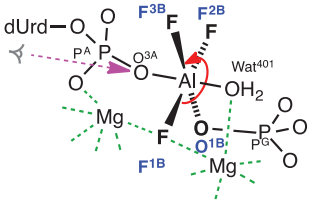

Fig. C3b

c) In case of multiple metal coordination, the coordinating metal with highest atomic number will take priority, where the distance of separation is less than the sum of the two van der Waals radii (as for Fig. C2b above).

d) In the absence of fluorine coordination to a metal, hydrogen bonding to amino acids will be used to determine fluorine priority. Thus, in the fructose 2,6-bisphosphatase reaction of the enzyme PFKFB3, an AlF_4^- complex with $His^{253}N^{e2}$ has been described (PDB entry: **3QPW**. Fig. C3c). This octahedral complex is completed by water coordination *trans* to the histidine nitrogen. The four fluorine atoms are coordinated F^1 to water, F^2 to Arg^{252} and Gln^{388}, F^3 to His^{387} and water, and F^4 to Arg^{252} and Asn^{259} (*fluorines numbered as in*

20 In cases where there are no other fluorines in the system, the Al designation may be omitted.

PDB **3QPW**). As F^{2} is coordinated to N^{ε} of Arg^{252} and F^{4} is coordinated to Arg^{252}-NH1, F^{2} takes priority since its H-bonding is to the nitrogen nearer to Cα of the lowest numbered coordinating amino acid. Hence, F^{1Al} is coordinated to Arg^{252} and Gln^{388} and the progression to F^{2Al}, F^{3Al}, and F^{4Al} proceeds clockwise as viewed from the water apex of the octahedral complex (CIP priority is O > N, magenta arrow) (Fig. C3c).[21]

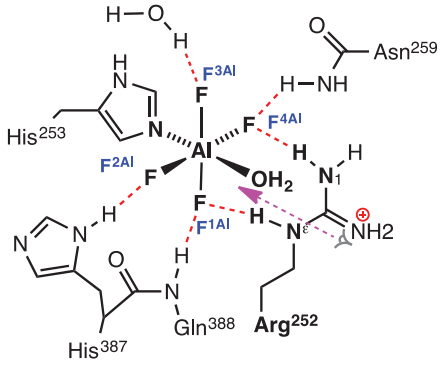

Fig. C3c

6 Recommendations for labelling vanadate and tungstate analogues of phosphates

6.1 C4. Vanadates

Orthovanadate, VO_4^{3-}, is encountered as an analogue of phosphate in a variety of forms. They are invariably trigonal bipyramidal and thus mimic a five-coordinate phosphoryl transfer process.

a) **Monosubstituted vanadate (V).** In isolation, vanadate (PDB ligand: **VO4**) can mimic the transition state for phosphoryl group transfer as a trigonal bipyramidal complex substituted by either one or two axial oxygen ligands that represent nucleophile and leaving group. A typical example is the Xac nucleotide pyrophosphatase/phosphodiesterase structure (PDB entry: **2GSO**) where the vanadate is axially coordinated to Thr^{90}. The three equatorial oxygen atoms are numbered O^{1V}, O^{2V}, and O^{3V}, with the axial oxygen O^{4V} being *trans* to the hydroxylic oxygen of Thr^{90} (Fig. C4a). The equatorial oxygen coordinated to two zinc ions takes priority and is O^{1V}. The direction of viewing is determined by the priority Thr^{90} oxygen > O^{4V} (magenta arrow). Thus, a clockwise progression identifies O^{2V} at the front and O^{3V} at the rear of the trigonal planar array.

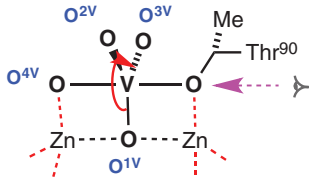

Fig. C4a

21 Coordination to an oxygen of an isolated water is ignored. This is because the presence or absence of water in a PDB structure may be a function of the resolution of the structure, and therefore may vary from one structure to another of the same protein-ligand complex. Also note the use of PDB style numbering for atoms in amino acids (which avoids the use of Greek symbols).

b) **Disubstituted vanadate.** A transition state analogue complex for phosphorylation of glucose 1-phosphate on O^{6} by α-phosphoglucomutase has vanadate linearly coordinated by oxygen-3 of Ser^{116} and by oxygen-6 of glucose 1-phosphate (PDB entry: **1C4G**). CIP priority analysis gives $O^{6G} > O^{3S}$. The three equatorial oxygen atoms take priority from O^{1V} by its coordination to cobalt, which substitutes for a native catalytic magnesium. Assignment of O^{2V} and O^{3V} follows a clockwise progression when viewed from O^{6G} (magenta arrow) (Fig. C4b).

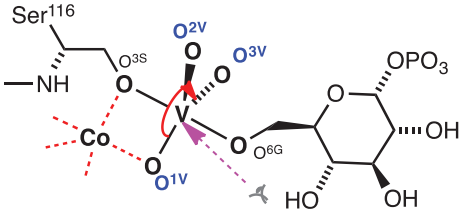

Fig. C4b

For the nucleoside-diphosphate kinase from *B. burgdorferi*, a vanadate transition state complex links ADP and His^{134} as axial ligands (PDB entry **4DZ6**). There is no catalytic metal to coordinate the three equatorial oxygen atoms. Thus, oxygen H-bonded to Lys^{13} takes priority as O^{1V} over oxygen O^{2V} H-bonded to Arg^{94}, while O^{3V} is not H-bonded to any amino acid. These assignments are in accord with those in the PDB entry.

c) **Trisubstituted vanadate.** Tyrosyl-DNA phosphodiesterase (Tdp1) is a DNA repair enzyme that catalyzes the hydrolysis of a phosphodiester bond linking a tyrosine residue to a DNA 3′-phosphate. Orthovanadate is central in a transition state analogue structure in which vanadium is linked to the tyrosine oxygen, to the 3′-oxygen of the scissile nucleotide, and to His^{262} of the enzyme (PDB entry: **1RFF**). Axial ligand priority is Tyr-O > $HisN^{\varepsilon 2}$. Equatorial ligand priority is assigned to Thd-$O^{3'}$. Hence, O^{2V} and O^{3V} follow in a clockwise progression when viewed from the Tyr-oxygen (Fig. C4c).

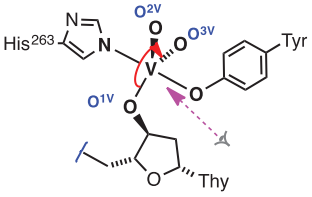

Fig. C4c

d) **Cyclic trisubstituted vanadate.** Trisubstituted vanadate provides a transition state analogue structure for hairpin ribozyme cleavage of a phosphodiester (PDB entry: **1M5O**). The axial $O^{2'}$ has CIP priority over the axial $O^{5'}$. Priority in the three equatorial oxygen atoms is taken by the ribose-$O^{3'}$ leading to assignment of O^{1V} followed by O^{2V} in a clockwise progression (Fig. C4d).

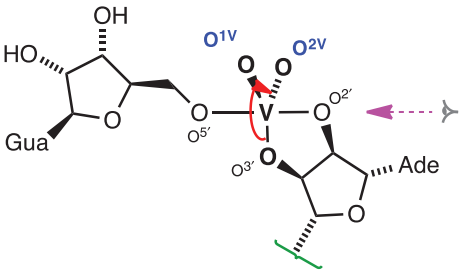

Fig. C4d

6.2 C5. Tungstates

Tungstate(VI) ion, $WO_4^{=}$ (PDB ligand code: **WO4***) is a mimic of tetrahedral phosphate in a small but significant range of structures in the PDB. In such systems, two oxygen atoms need to be assigned priority to enable the remaining two to be assigned by prochirality rules.*

a) **Isolated tungstate(VI) with two metal ions.** In a structure of purple acid phosphatase (PDB entry: **3KBP**), an isolated tungstate(VI) ion mimics phosphate. It is coordinated both to zinc and to iron. Zinc, with atomic number 30, takes CIP priority over iron (atomic number 26) and so the two tungstate oxygen atoms coordinated to these metal ions are labelled O^{1W} and O^{2W}, respectively (Fig. C5a). The remaining two tungstate oxygen atoms are now prochiral and can be labelled O^{3W} and O^{4W} by CIP rules described above.

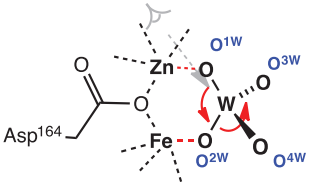

Fig. C5a

b) **Isolated tungstate(VI) with one metal ion**. In a structure of a tungstate complex of CheYN59D/E89R, the isolated tungstate(VI) ion is coordinated to manganese and several amino acids (PDB entry: **3RVS)**. Thus, O^{1W} is identified by its coordination to manganese. Coordination to oxygen gives precedence over coordination to nitrogen. Coordination to oxygen is only considered if the distance of the heavy atoms ≤ 3.0 Å (see Section B3c). Hence, O^{2W} is coordinated to Asp^{59} and takes precedence over the third oxygen that is coordinated to Thr^{87}. The remaining two tungstate oxygen atoms are now prochiral and can be labelled O^{3W} and O^{4W} by CIP rules described above (Fig. C5b).

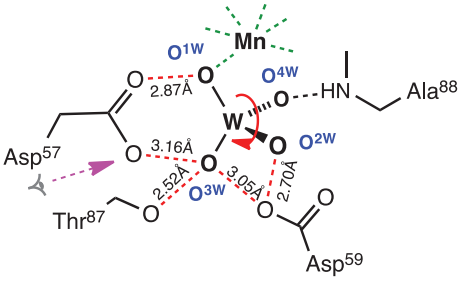

Fig. C5b

c) **Isolated tungstate(VI) with no metal ion**. For an isolated WO_4^{-} species, a similar procedure of prioritisation by amino acid coordination can be used to identify O^{1W} and O^{2W}. O^{3W} and O^{4W} can then be assigned by the prochirality procedure. Thus, in a structure of *Yersinia enterocolitica* PTPase complexed with tungstate (PDB entry **3F9A**), an isolated tungstate is encircled by a loop of amino acids 404–409, with three of its oxygen atoms coordinated to NHs. As isolated water coordination is ignored,[22] priority is given to coordination from Arg^{404} to O^{1W} followed by coordination from Val^{407} to O^{2W}. Hence, O^{3W} and O^{4W} are now prochiral and can be assigned using CIP rules (Fig. C5c).

22 Note that once coordination has reduced the number of non-prioritised oxygen atoms to two, this pair is assigned by application of CIP rules on prochirality.

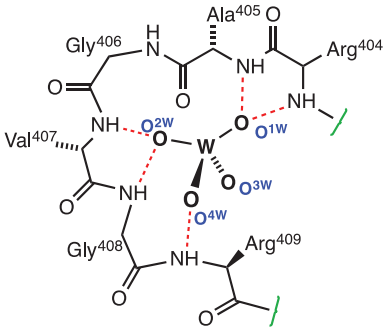

Fig. C5c

d) **Tungstate(VI) coordinated to a substrate ligand**. A compound example of tungstate as a dual analogue of phosphate is found in the structure of a protein of the histidine triad family in which adenosyl 5′-ditungstate (PDB ligand: **ADW),** an analogue of ADP, is coordinated to His^{112} (PDB entry: **1KPE**). This situation calls for labelling both tungstens and seven oxygen atoms, since the first tungstate is a trigonal bipyramidal TSA of P^A and the second tungstate is a tetrahedral analogue of P^B of ADP. Tungsten W^A is equatorially linked to the adenosyl 5′-oxygen and axially linked to His^{112}-N^{τ}.[23] As in the case of polyphosphates (Section B1c), the bridging oxygen to W^B is designated O^{3A}. This enables the assignment of the two prochiral equatorial oxygen atoms as O^{1A} and O^{2A} (when viewed in the axial direction O^{3A} to N^{τ}). For W^B, oxygen O^{3A} has highest CIP priority because it is coordinated to W^A. The oxygen coordinated to Gly^{105} takes precedence over the oxygen coordinated to Ser^{107} and is therefore identified as O^{1B}. This enables the prochiral pair of oxygen atoms to be assigned as O^{2B} and O^{3B} as shown (Fig. C5d).

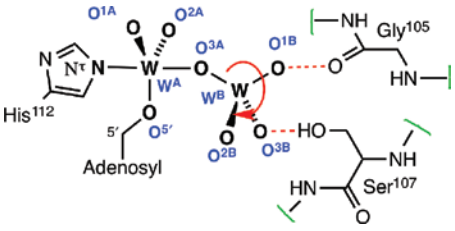

Fig. C5d

7 Summary

The recommendations presented here have been developed to describe molecules derived from orthophosphoric acid and its derivatives, analogues, and transition state analogues. We have built on the existing usage of nomenclature wherever possible, especially on Cahn-Ingold-Prelog Rules. We have introduced only one Rule to override CIP analysis, and that for the purpose of simplicity of use by non-experts in stereochemistry. In our experience, the recommendations have worked well for the most demanding species we have examined, *e.g.* **C3c** and **C5d**. However, we recognise that the recommendations may be equally relevant to other species with tetrahedral geometry, such as sulfates and sulfonamides, or with tbp or octahedral geometries. We also recognise that there may be existing, or as yet non-existent, structures that could require an extension of these recommendations, and we remain receptive for advice on such problems.

Funding: International Union of Pure and Applied Chemistry, (Grant/Award Number: 'USD 10,000').

23 The PDB description of N^{τ} is $N^{\varepsilon 2}$.

Appendix – A guide for the use of Cahn-Ingold-Prelog rules for prochirality

Short procedure for identification of paired non-bridging oxygen atoms (or paired fluorine atoms) using Cahn-Ingold-Prelog Rules for prochirality (enantiotopicity)

1. Two non-bridging oxygen atoms bonded to the same phosphorus are *enantiotopic* if promoting one of them from isotope-16 to isotope-18 generates the opposite enantiomer compared to promotion of the other.[24] This is illustrated for methyl phenylphosphonate (Fig. X1). Promoting the 'front' oxygen (Step **a**) gives molecule (**A**) where the phosphorus is a *stereogenic* centre and is labelled *R* in Cahn-Ingold-Prelog nomenclature. Promoting the 'rear' oxygen (Step **b**) gives molecule (**B**) where the stereogenic phosphorus is labelled *S*. This analysis is based on the CIP priority rule O(Me) > ^{18}O > ^{16}O > C; on viewing the face of the P-centered tetrahedron with the lowest priority ligand (C) at the rear (magenta arrow), a clockwise progression from high to low priority is designated *R* (as shown) and an anticlockwise progression is *S*. Because **A** and **B** are enantiomers, the two non-bridge oxygen atoms are enantiotopic. In extension, the paired, non-bridge oxygen atoms can be labelled (*pro-R*) for the front one (clockwise progression) and (*pro-S*) for the rear one (**C**, right).

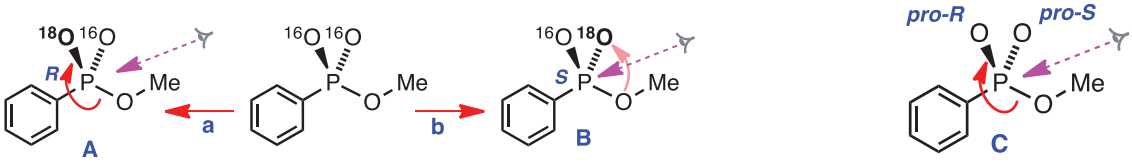

Fig. X1

2. Two non-bridging oxygen atoms bonded to the same phosphorus are *diastereotopic* if promoting one of them from isotope-16 to isotope-18 generates a different diastereoisomer compared to promotion of the other.[25] In the case of adenosine 5′-diphosphate (ADP), the two non-bridging paired oxygen atoms on P^A are diastereotopic. Promoting the 'front' oxygen to ^{18}O generates a new stereogenic centre at P^A (**D**; CIP label *S*), while promoting the 'rear' oxygen to ^{18}O generates a stereogenic centre with the opposite sense at P^A (**E**; CIP label *R*) (Fig. X2). **NB** the **D** and **E** stereoisomers are *not* mirror images because the stereochemistry of the D-ribose is unchanged. Since they are *not* enantiomers, they are termed diastereoisomers.[26]

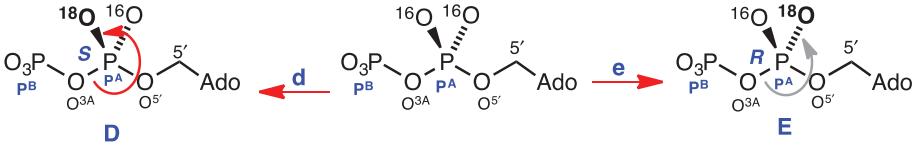

Fig. X2

3. We can now apply the priority rules described in Section X1 to label the non-bridging oxygen atoms on P^A in ADP. This analysis begins with the CIP priority rule that ranks di-coordinate oxygen above mono-coordinate oxygen. Thus, $O^{5'}$ and O^{3A} rank above the two non-bridging oxygen atoms. For these bridging oxygen atoms, relative priority is determined by the next atom in the chain: priority is given to the atom with the higher atomic number. In the case of ADP, the sub-adjacent atoms along the chain are P^B and $C^{5'}$. Hence, O^{3A} has priority over $O^{5'}$ as P has a higher atomic number than C. The CIP priority ranking is thus $O^{3A} > O^{5'} > {}^{18}O > {}^{16}O$ (Fig. X3).[27]

---

24 These oxygen atoms are spectroscopically and chemically non-equivalent in a chiral environment.

25 These oxygen atoms are spectroscopically and chemically non-equivalent in *any* environment.

26 The term diastereoisomer simply embraces all stereoisomers that are *NOT* enantiomers.

27 **NB** Labelling a non-bridge oxygen with ^{18}O only gives it priority over the ^{16}O oxygen. It does not change its priority relative to the non-bridging oxygen atoms.

Viewing the P-centered tetrahedron in stereoisomer (**D**) from the face with ^{16}O at the rear gives an anti-clockwise progression from high to low priority ligands (Fig. X3 left) and so P^A in **D** has *S*-configuration. Hence, the two paired-oxygen atoms in ADP can be labelled *pro-S* for the front one (*as its promotion to* ^{18}O *makes* P^A *an* S *chiral centre*) and *pro-R* for the rear one (*as its promotion to* ^{18}O *makes* P^A *an* R *chiral centre*) (Fig. X3 center). We can now apply CIP Rule 5 that gives *R* priority over *S*. Thus, the *pro-R* oxygen is labelled O^{1A} and the *pro-S* oxygen is labelled O^{2A} (Fig. X3 right).

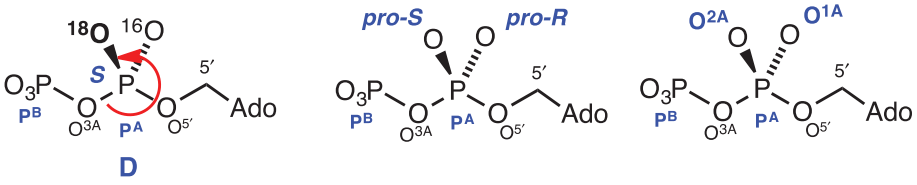

Fig. X3

4. The rigorous application of the CIP rules almost inevitably means that there are unexpected outcomes. For example, the stereochemistry of the non-bridging oxygen atoms at P^B in guanosine 5′-triphosphate (GTP) and in γ-thioGTP (**GSP**) have opposite assignments. For GTP, the rules for the in-chain atoms flanking P^B identify $O^{5'}$ bonded to $C^{5'}$, thereby taking priority over all oxygen atoms bonded to P^G (O^{1G}, O^{2G}, O^{3G}). Hence, the priority sequence for the four GTP oxygen ligands at P^B is $O^{3A} > O^{3B}$ and thus the front oxygen is *pro-R* and the rear oxygen is *pro-S* (Fig. X4a). Hence, the *pro-R* oxygen is labelled O^{1B} and the *pro-S* oxygen oxygen is O^{2B} (Fig. X4a).
By contrast, for GTPγS, the sulfur atom on P^G takes CIP priority over $O^{5'}$, with the consequence that O^{3A} takes priority over O^{3B} (Fig. X4b). The result is that in GTPγS (as presented) the rear oxygen is *pro-R* (and thus O^{1B}) and the front oxygen is *pro-S* (and thus O^{2B}) (Fig. X4b). This is the opposite 3D spatial outcome compared to GTP. We note that a like situation holds for GTPγF, but not for γ-amino-GTP.

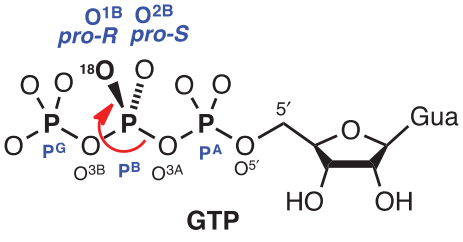

Fig. X4a

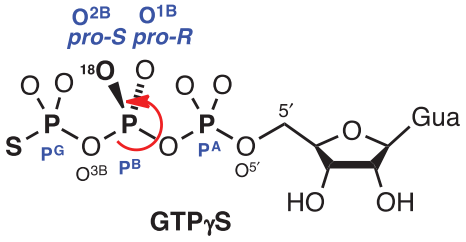

Fig. X4b

5. The above problem is readily resolved by a user-friendly solution.
The practical use of these interlinked analogues of nucleoside triphosphates, especially for ATP and GTP, calls for a simplified nomenclature system, designed to deliver comparable labels for diastereotopic oxygens on non-terminal phosphates across the series of phosphoramidate, phosphonate, phosphofluoridate, and phosphorothioate analogs of NTPs. That can be achieved by the use of a new nomenclature rule that gives priority to bridging oxygens in an extended polyphosphate chain **based simply on their position in the chain**, rather than on their often convoluted substituent atom priorities:
"Bridging atoms in polyphosphate chains should be given priorities that increase with separation from the priority nucleoside oxygen: $O^{n'} < O^{3A} < O^{3B} < O^{3G} < O^{3D} <$ *etc.*"
This rule, already described and used in Section 4B1d, is applied here to resolve the nomenclature discrepancy for O^{1B} and O^{2B} encountered in structures X4a,b (above). In the case of GTP analogs, having O^{3G} replaced by another heavy atom, the rule assigns priorities as: $O^{3B} > O^{3A} > O^{1B} > O^{2B}$ (Fig. X5a). This now

assigns the same descriptors to the prochiral oxygens at P^B as that in Fig. X4b (and the contrary to that for Fig. X4a).

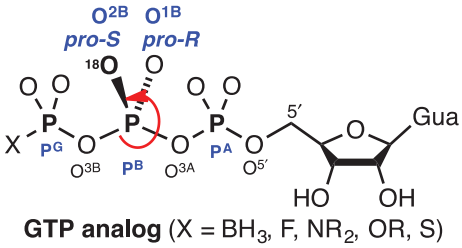

Fig. X5a

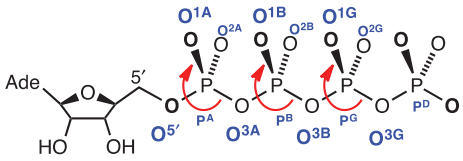

Fig. X5b

The simplicity of use of this rule is also illustrated in naming the diastereotopic pairs of oxygens on P^A, P^B and P^G for adenosyl-5′ tetraphosphate, p_4A (Fig. X5b). The priority order for the bridging oxygens, $O^{n'} < O^{3A} < O^{3B} < O^{3G}$, results in a consistent naming for the non-bridging oxygens relative to their spatial location as O^{1A}, O^{1B}, and O^{1G} (all frontal) and O^{2A}, O^{2B}, and O^{2G} (all rearwards) (Fig. X5b).

Membership of the sponsoring body

Membership of the IUPAC Organic and Biomolecular Chemistry Division Committee for the period 2014–2015 is as follows:

President: M. J. Garson (Australia); **Vice President:** M. A. Brimble (New Zealand); **Secretary:** A. Griesbeck (Germany); **Past President:** K. N. Ganesh (India); **Titular Members:** G. Blackburn (UK); A. Brandi (Italy); T. Carell (Germany); B. Han (China); F. Nicotra (Italy) N. E. Nifantiev (Russia); **Associate Members:** A. Al-Aboudi (Jordan); V. Dimitrov (Bulgaria); J. F. Honek (Canada); P. Mátyus (Hungary); A. P. A. E. Knight: Total internal reflection fluorescence microscopy 1319 Rauter (Portugal); Z. Xi (China/Beijing); **National Representatives:** Y.-M. Choo (Malaysia); O. M. Demchuk (Poland); M. Ludwig (Czech Republic); V. Milata (Slovakia); M. Olire Edema (Nigeria); P. M. Pihko (Finland); N. Sultana (Bangladesh); H. Vanik (Croatia); B.-J. Uang (China/Taipei); T. Vilaivan (Thailand).

Membership of the IUPAC Organic and Biomolecular Chemistry Division Committee for the period 2016–2017 is as follows:

President: M.A. Brimble (New Zealand); **Vice President:** F. Nicotra (Italy); **Secretary:** A. Rauter (Portugal); **Past President:** M. Garson (Australia); **Titular Members:** P. Andersson (Sweden); J. Clardy (USA); N. Nifantiev (Russia); G. Pandey (India); J. Scott (UK); Chen Xi (China): **Associate Members:** V. Bolzani (Brazil): T. Carell (Germany); E. Uggerud (Norway); O. Demchuk (Poland); N. M. Carballeira (Puerto Rico); B. Sener (Turkey); **National Representatives:** E. Naydenova (Bulgaria); Ken-Tsung Wong (China/Taipei); M. Ludwig (Czech Republic); S. Marque (France); A. Al-Aboudi (Jordan); I. Shin (Korea); K. Lammertsma (Netherlands); M. Ali Munawar (Pakistan); V. Milata (Slovakia); L. Mammino (South Africa).

References

[1] J. R. Knowles. *Annu. Rev. Biochem.* **49**, 877 (1980).
[2] G. Lowe. *Acc. Chem. Res.* **16**, 244 (1983).
[3] R. S. Cahn, C. Ingold, V. Prelog. *Angew. Chem.* **78**, 413 (1966).
[4] R. S. Cahn, C. Ingold, V. Prelog. *Angew. Chem. Int. Ed. Engl.* **5**, 385 (1966).
[5] G. A. Orr, J. Simon, S. R. Jones, G. J. Chin, J. R. Knowles. *Proc. Natl. Acad. Sci. USA* **75**, 2230 (1978).
[6] K.-F. Sheu, P. A. Frey. *J. Biol. Chem.* **253**, 378 (1978).

[7] R. L. Jarvest, G. Lowe. *J. Chem. Soc. Chem. Commun*. 364 (1979).
[8] S. R. Jones, L. A. Kindman, J. R. Knowles. *Nature* **275**, (5680), 564 (1978).
[9] M. W. Bowler, M. J. Cliff, J. P. Waltho, G. M. Blackburn. *New J. Chem*. **34**, 784 (2010).
[10] S. C. Kamerlin, P. K. Sharma, R. B. Prasad, A. Warshel. *Quart. Revs. Biophys*. **46**, 1 (2013).
[11] G. M. Blackburn, Y. Jin, N. G. J. Richards, J.P. Waltho. *Angewandte Chemie Int. Ed*. (2016). doi: 10.1002/anie.201606474.
[12] N. J. Baxter, L. F. Olguin, M. Goličnik, G. Feng, A. M. Hounslow, W. Bermel, G. M. Blackburn, F. Hollfelder, J. P. Waltho, N. H. Williams. *Proc. Natl. Acad. Sci. USA* **103**, 14732 (2006).
[13] N. J. Baxter, G. M. Blackburn, J. P. Marston, A. M. Hounslow, M. J. Cliff, W. Bermel, N. H. Williams, F. Hollfelder, D. E. Wemmer, J. P. Waltho. *J. Am. Chem. Soc*. **130**, 3952 (2008).
[14] Y. Jin, M. J. Cliff, N. J. Baxter, H. R. W. Dannatt, A. M. Hounslow, M. W. Bowler, G. M. Blackburn, J. P. Waltho. *Angew. Chem. Int. Ed*. **51**, 12242 (2012).
[15] J. Mishra, U. S. Bhalla. *Biophys. J*. **83**, 1298 (2002).
[16] M. A. G. Sillero, A. de Diego, E. Silles, F. Pérez-Zúñiga, A. Sillero. *FEBS Lett*. **580**, 5723 (2006).
[17] Nomenclature of Phosphorus-Containing Compounds of Biochemical Importance (Recommendations 1976) IUPAC-IUB Commission on Biochemical Nomenclature. *Proc. Natl. Acad. Sci. USA* **74**, 2222 (1977).
[18] International Union of Biochemistry and Molecular Biology, *Biochemical Nomenclature and related documents*, 2nd edition, Portland Press 1992, 109–114, 256–264, and 335 [ISBN 1-85578-005-4]. (see also IUPAC-IUB Joint Commission on Biochemical Nomenclature, *Pure Appl. Chem*. **55**, 1273–1280 (1983)).
[19] Abbreviations and symbols for the description of conformations of polynucleotide chains. Recommendations 1982. *Eur. J. Biochem*. **131**, 9 (1983).
[20] Newsletter of the Nomenclature Committees of the International Union of Biochemistry and Molecular Biology. *Eur. J. Biochem*. **122**, 437 (1982). [also ref 12, p. 265].
[21] IUPAC Nomenclature of Organic Chemistry, IUPAC Recommendations and Preferred Names 2013 (H. A. Favre, W. H. Powell, Eds.), Royal Society of Chemistry, Cambridge, UK (2014). (a) P-93.2.4 p. 1215, (b) P-93.3.3.5 p. 1222–1223, (c) P-93.3.3.7 p. 1225–1226, H. A. Favre, W. H. Powell *Royal Society of Chemistry* (2013). Corrections http://www.chem.qmul.ac.uk/iupac/bibliog/BBerrors.html; p. 432ff P-42.3-4 Phosphorus acids *etc.*; p. 915ff P-67 Phosphorus acids *etc.*; p. 992ff P-68.3 Phosphorus compounds; p. 1215ff P-93.2.4 Stereochemistry of phosphates *etc.*; p. 1420ff P-105 Nucleosides; p. 1425ff P-106 Nucleotides.
[22] *Nomenclature of Inorganic Chemistry: IUPAC Recommendations 2005*, *Royal Society of Chemistry* (2005), Corrections http://www.chem.qmul.ac.uk/iupac/bibliog/RBcorrect.html which has links to later corrections. PDF:http://www.iupac.org/nc/home/publications/iupac-books/books-db/book-details.html?tx_wfqbe_pi1[bookid]=5; p.180ff, IR 9.3.3.4, octahedral coordination; p184ff, IR 9.3.3.6, bipyramidal coordination; p. 189ff, IR 9.3.3.8, chiral octahedral; p. 190ff, IR 9.3.3.10, chiral bipyramid.