INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

How to Name Atoms in Phosphates, Polyphosphates and their Analogues, and Transition State Analogues for Enzyme-catalysed Phosphoryl Transfer Reactions

Journal:	Pure and Applied Chemistry
Manuscript ID:	
Manuscript Type:	Recommendation
Date Submitted by the Author:	*18 December 2015*
Complete List of Authors	Blackburn, G. Michael, Krebs Institute, University of Sheffield; Cherfils, Jacqueline, CNRS and Ecole Normal Supérieur, Cachan; Moss, Gerald P., Queen Mary University of London; Richards, Nigel J., Department of Chemistry, IUPUI, Indianapolis; Waltho, Jonathan P., Biosciences Institute, University of Manchester; Williams, Nicholas H., Chemistry Department, University of Sheffield; Wittinghofer, Alfred, Group for Structural Biology, Max-Planck-Institut für Molekulare Physiologie, Dortmund.
Keywords:	
Author-supplied Keywords	Phosphate nomenclature, recommendations, N, O, P atom labels, phosphate stereochemical naming, polyphosphates, phosphoryl transfer, atom labels for transition states.

December 18, 2015

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

ORGANIC AND BIOMOLECULAR CHEMISTRY DIVISION

HOW TO NAME ATOMS IN PHOSPHATES, POLYPHOSPHATES, THEIR DERIVATIVES AND MIMICS, AND TRANSITION STATE ANALOGUES FOR ENZYME-CATALYSED PHOSPHORYL TRANSFER REACTIONS

~~IUPAC Recommendations 2016~~[†]

G. Michael Blackburn,[a] Jacqueline Cherfils,[b] Gerald P. Moss,[c] Nigel J. Richards,[d] Jonathan P. Waltho,[e] Nicholas H. Williams,[f] Alfred Wittinghofer[g]

a) Department of Molecular Biology, Krebs Institute, University of Sheffield, S10 2TN, UK

b) Laboratoire de Biologie et Pharmacologie Appliquée, CNRS - Ecole Normale Supérieure de Cachan, Cachan, France

c) Queen Mary University of London, School of Biological and Chemical Sciences, London E1 4NS, UK

d) Department of Chemistry, Indiana University Purdue University Indianapolis, IL 46202, USA, and School of Chemistry, Cardiff University, Cardiff CF10 3AT, UK

e) Biosciences Institute, University of Manchester, M1 7DN, UK

f) Chemistry Department, University of Sheffield, Sheffield S10 7HF, UK

g) Group for Structural Biology, Max-Planck-Institut für Molekulare Physiologie, 44227 Dortmund, Deutschland

[†] *Prepared for publication in 2015 by* G. M. Blackburn and G. P. Moss

HOW TO NAME ATOMS IN PHOSPHATES, POLYPHOSPHATES, THEIR DERIVATIVES AND MIMICS, AND TRANSITION STATE ANALOGUES FOR ENZYME-CATALYSED PHOSPHORYL TRANSFER REACTIONS

(~~IUPAC Recommendations 2016~~)

Abstract: Procedures are proposed for the naming of individual atoms, N, P, O, etc., in phosphate esters, amidates, thiophosphates, polyphosphates, their mimics, and analogues of transition states for enzyme-catalysed 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 and 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 herein 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 clear to assimilate and convenient to use.

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

CONTENTS

P.O. 13757, Research Triangle Park, NC (919) 485-8700

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 Jeremy Knowles [1] and of Gordon Lowe [2] placed new demands on the nomenclature of the oxygen atoms of the transferring phosphoryl group, PO_3^-. 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] resolved the problem by labelling the oxygens (*R*p) and (*S*p) as appropriate. [4] The more advanced use of ^{16}O, ^{17}O, and ^{18}O bonded to the same phosphorus [5] led to the concept of *pro-pro-pro*-chirality at phosphorus, which was still capable of CIP identification. [2, 5] 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. [6, 7] Indeed, there are now over 1600 ligands in the PDB having a phosphoryl group component and they 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^-; PDB ligand code: **BEF**) is a ground state analogue for phosphate, with characteristic tetrahedral geometry when ligated to anionic oxygen. Tetrafluoroaluminate (AlF_4^-; PDB ligand code: **ALF**) is a mimic for concerted phosphoryl transfer in multiple enzymes, though it has octahedral geometry. Aluminium trifluoride (AlF_3; PDB ligand code: **AF3**) forms trigonal bipyramidal (tbp) TSA complexes that have the correct stereochemistry for a concerted PO_3^- group transfer but lack the ionic charge thereof. These two values converge in the relatively smaller number of trifluoromagnesate complexes (MgF_3^-; PDB ligand code **MGF**) 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. [8] The growth in use of these four types of MF_x complexes is illustrated in Figure 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 Figure 1. The double change from four coordinate tetrahedral PO_4 to five coordinate tbp O-MF_3-O and six 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 this development of these standardized naming conventions.

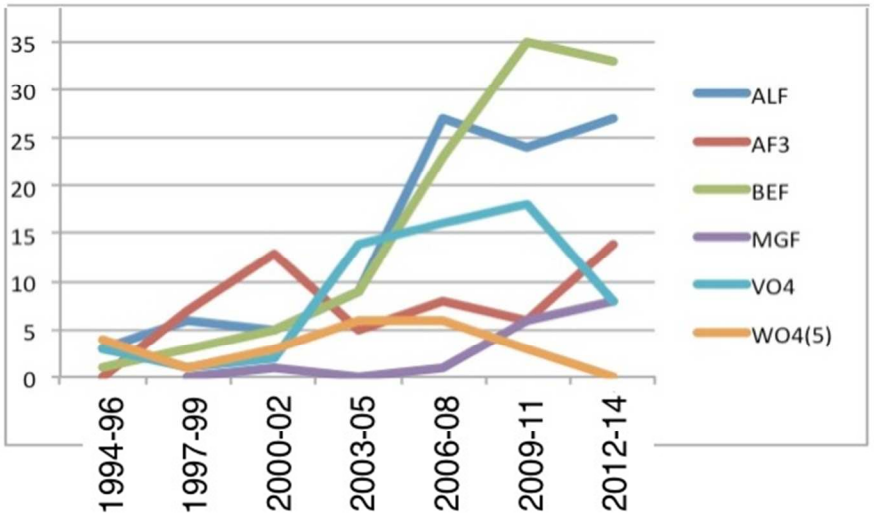


Figure 1. Protein structures published in the PDB for successive triennia containing the ligands designated as analogues of phosphoryl groups or their transition states.

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. [9] 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 oxygens of 5'-adenosyl 5'''-guanosyl P^1,P^4-tetraphosphate [10] and the 3 non-isotopically identifiable oxygens of the PO_3^- group of terminal phosphates. With those objectives accomplished, our recommendations could then be developed to incorporate the fluorine ligands of MF_x systems and also 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 oxygens 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. [11] 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 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. [12]

b) A document on the abbreviations and symbols for the description of conformation of polynucleotide chains, Recommendations 1982, was published in 1983. [13] In a related paper, it was proposed that the *pro-S* oxygen should be OP1 and *pro-R* should be OP2. [14] This is the *reverse* of the system proposed here and it is also contrary to CIP nomenclature that gives priority to *R* over *S* (CIP Rule 5). We have chosen to adhere to CIP priority Rule 5.

c) IUPAC Recommendations for preferred names of derivatives of phosphoric acid are pertinent. [15] They included 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. [16]

3. RECOMMENDATIONS FOR LABELLING PHOSPHORUS ATOMS IN PHOSPHATES

A. Labelling Phosphorus Atoms in Polyphosphate Species

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 PA, PB, PG, PD etc.[a] Hence adenosine 5'-tetraphosphate (PDB ligand: **AQP**) has phosphorus atoms labelled as PA, PB, PG, PD starting from the ribose 5'-oxygen [Fig. A1a].

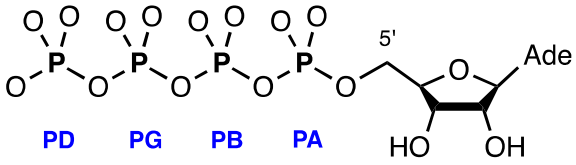


Figure A1a

b) 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**), PA is bonded to uridine-O5' and PB is bonded to O1" of glucose [Fig. A1b].[b]

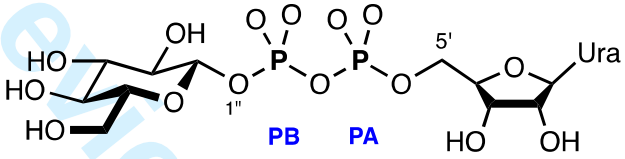


Figure A1b

c) A nucleic acid base takes priority over non-nucleic acid base (i.e. adenosine > nicotinamide riboside). Thus in NAD^+ (PDB ligand: **NAD**), PA is bonded to O5' of adenosine with PB bonded to O5''' of the nicotinamide riboside [Fig. A1c].

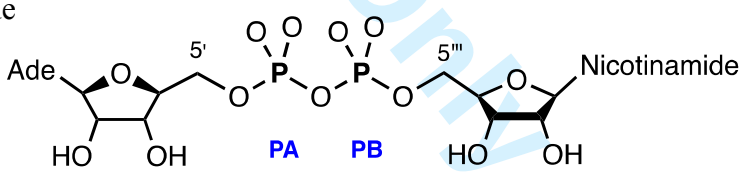


Figure A1c

[a] PDB usage currently always replaces P□ with PG as it does not use a Greek/Symbol font.

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

d) Nucleosides take priority in 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 PA, PB, PG, PD starting at the 5'-oxygen of the adenosine [Fig. A1d].

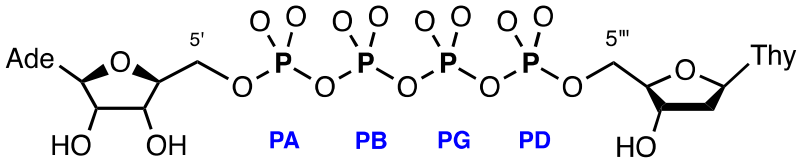


Figure A1d

Pentoses have priority D-ribose > L-ribose > 2-deoxy-D-ribose > 2-deoxy-L-ribose.[c] Thus a transition state for dAMP kinase should label the four phosphorus atoms PA, PB, PG, PD starting from the adenosine 5'-oxygen [Fig. A1e].

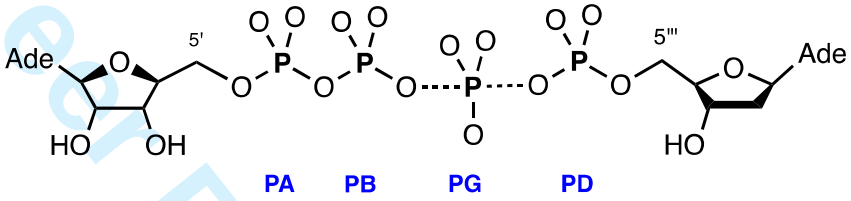


Figure A1e

e) 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 PA, PB and PG from the 5'-oxygen [Fig. A1f1].

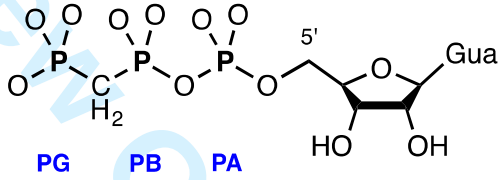


Figure A1f1

Likewise, for 2'-deoxyuridine 5'-α,β-imidotriphosphate (PDB ligand: **DUP**) phosphorus atoms should be named PA, PB and PG from the 5'-oxygen [Fig. A1f2].

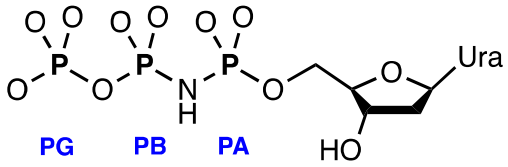


Figure A1f2

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

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 P1, P3, P4, P5, and P6 [Fig. A2a1].[d] For fructose 1,6-bisphosphate (PDB label: **FBP**) the phosphorus atoms should be labelled P1 and P6 [Fig. A2a2].

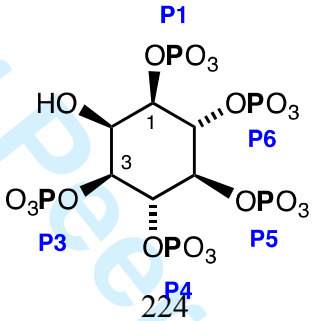


Figure A2a1

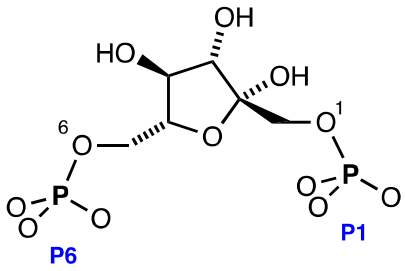


Figure A2a2

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 oxygens 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 PA5, PB5 and PG5, and the 3'-phosphorus atoms named PA3 and PB3 [Fig. A3a].

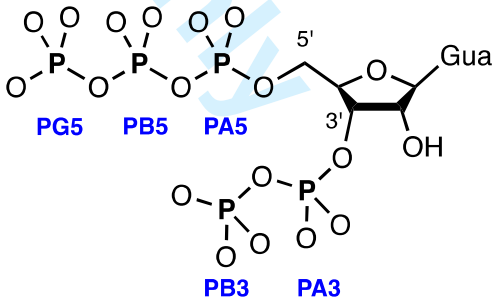


Figure A3a

[d] *cf.* R. F. Irvine & M. J. Schell, *Nature Rev. Molec. Cell Biol.* **2**, 327-338 (2001).

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 P1, P2, etc. should apply to single phosphorus entities while PAn, PBn will apply to diphosphates, as in PP-InsP$_5$ (PDB ligand: **I7P**) [Fig. A3b].[d]

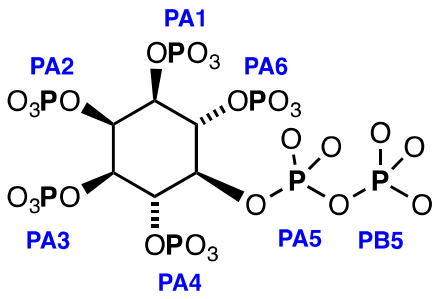


Figure A3b

4. RECOMMENDATIONS FOR LABELLING OXYGEN ATOMS IN PHOSPHATES

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 oxygens bonded to the same phosphorus, followed by a letter to indicate the parent phosphorus.

a) The oxygen linking PA to the carbon moiety of the molecule will retain its regular label. Thus in ATP, O5' bonds PA to the ribose [Fig. B1a1]. In Ap$_4$G, O5' bonds PA to adenosine while O5''' bonds PD to guanosine [Fig. B1a2]. [10]

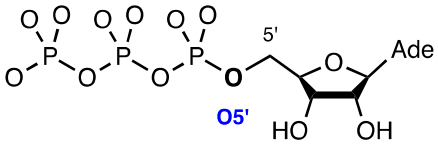


Figure B1a1

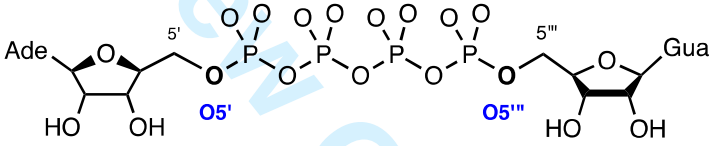
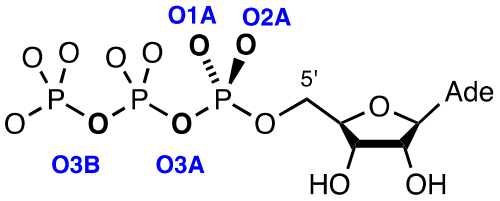


Figure B1a2

b) In each non-terminal phosphoryl group, the two non-bridging oxygens will be labelled 1 and 2 according to their CIP *pro-R*- and *pro-S*-chiralities respectively.[e] Hence in ATP, PA will have non-bridging oxygens O1A and O2A for the *pro-R* and *pro-S* oxygens respectively [Fig. B1b].



[e] This nomenclature is widely used in the PDB for oxygens on PA in nucleoside triphosphates but is rather variably used for oxygens on PB.

Figure B1b

c) In each non-terminal phosphoryl group (PO_3), the bridging oxygen bonding PX to P(X+1) in the chain should be numbered O3X. Hence in ATP, O3A joins PA to PB, and O3B joins PB to PG [Fig. B1b].

d) In chains containing a sulfur atom in a non-bridging, non-terminal position, the sulfur will take the name S1A (for substituent on PA), S1B (for substituent on PB), etc. The non-bridging oxygen then is named O2A, O2B, etc., and the bridging oxygen is O3A, O3B, etc., as above. This is shown for guanosine 5'-(*R*p)-α-thio-triphosphate (PDB ligand: **GAV**) [Fig. B1d].

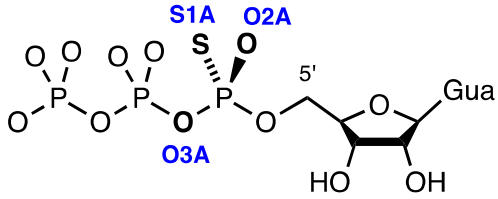


Figure B1d

e) In modified polyphosphate chains having two-atom bridges replacing an O3N (where N = A, B, etc.) the bridging atoms X and Y will be labelled X3A and Y4A progressively. Thus in β,γ-oxymethylene-ATP (Ado*POPO*CH_2*P*), the PB,PG-bridging atoms are O3B and C4B respectively [Fig. B1e].

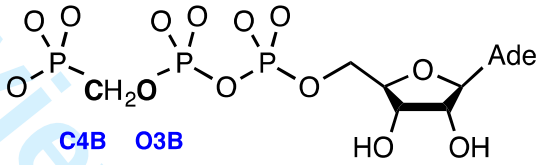


Figure B1e

f) In polyphosphate chains with a bridging oxygen replaced by carbon or nitrogen, the prochirality designations may change consequently. Thus in α,β-methylene adenosine 5'-triphosphate (PDB ligand: **APC**) [Fig.B1f], oxygens O1A and O2A are necessarily reversed relative to their designation in ATP [Fig. B1b].

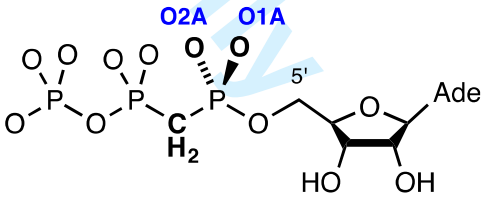


Figure B1f

B2. Non-terminal Phosphates in Molecules with Multiple Phosphate Chains

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

a) In each non-terminal phosphoryl group, the two non-bridging oxygens will be labelled 1 and 2 according to their CIP *pro-R* and *pro-S* chiralities respectively. Hence in ppGpp (PDB ligand: G4P), PA5 will have non-bridging oxygens O1A5 and O2A5 for the *pro-R* and *pro-S* oxygens respectively, and PA3 will have non-bridging oxygens O1A3 and O2A3 for the *pro-R* and *pro-S* oxygens respectively[f] [Fig. B2a1].

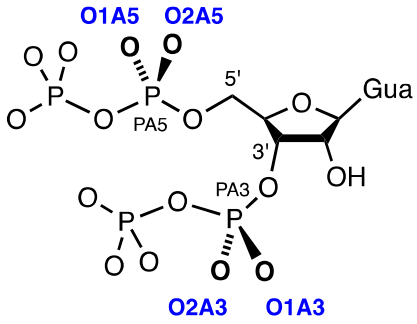


Figure B2a1

In NAD^+, the oxygens on PA5 will be labelled O1A5 and O2A5 for the *pro-R* and *pro-S* oxygens respectively, and the oxygens on PB5 will be labelled O1B5 and O2B5 for the *pro-R* and *pro-S* oxygens respectively [Fig. B2a2].

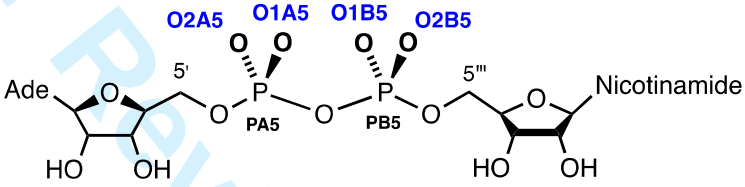


Figure B2a2

In ppIns5p, the oxygens on PA5 will be labelled O1A5 and O2A5 for the *pro-R* and *pro-S* oxygens respectively [Fig. B2a3].

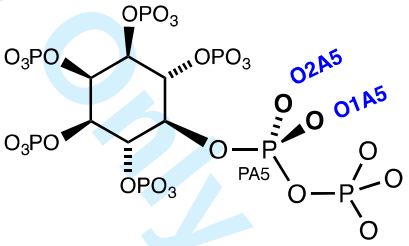


Figure B2a3

[f] For simplicity, the designation omits the prime symbol from e.g. O2A3'.

b) In each non-terminal phosphoryl group (PO_3), the bridging oxygen bonding PN to P(N+1) (where N = A, B, etc.) in the chain should be numbered O3Nx, where x designates the parent oxygen of the polyphosphate chain. Hence in ppGpp (PDB ligand: **P4G**), PA5 is joined to PB5 by O3A5, and PA3 is joined to PB3 by O3A3 [Fig. B2b].

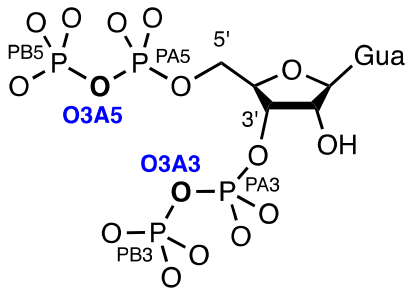


Figure B2b

B3. 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 oxygens 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 oxygens 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 oxygens are labelled O2G (*pro-R*) and O3G (*pro-S*) respectively [Fig. B3a].

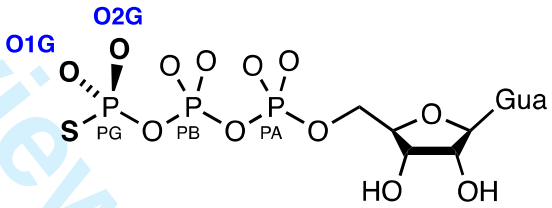


Figure B3a

b) Prochirality identification can be applied if one of the three oxygens is promoted relative to the other two. In the context of enzyme-bound nucleotides, such promotion can often be identified by co-ordination of the terminal phosphate to a protein-bound metal ion, typically magnesium. Thus for ATP bound in many kinases, the γ-phosphate is often coordinated from one of its three oxygens to magnesium. This oxygen is thus designated O1G. The remaining oxygens are now prochiral and can be identified in the priority series O3B > O1G > O2G > O3G. CIP rules then designate O2G as the *pro-R* oxygen and O3G as the *pro-S* oxygen, as illustrated for ATP bound in phosphoglycerate kinase [Fig. B3b; PDB entry: **1VJC**].

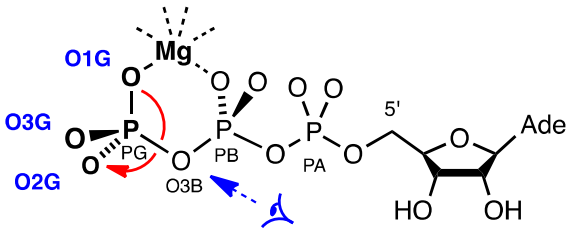


Figure B3b

c) In the absence of metal ion coordination to the terminal phosphate, hydrogen bond donation from amino acids in the protein provides a means of priority identification for O1N. Hydrogen bonds are considered only if they have a length ≤ 3.0 Å; 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 O1G in ATP, etc. If there is still ambiguity in the assignment, then backbone NH takes priority over sidechain NH.[g] This selection makes O2G and O3G prochiral and hence they can be assigned by application of CIP rules.[h] Thus in human bisphosphoglycerate mutase (PDB entry: **2A9J)**, the 3-phosphoglycerate has phosphoryl oxygen coordination from Arg^{100} and Arg^{116} to O1A, from Arg^{117} and Asn^{190} to O2A, and from Arg^{117} to O3A [Fig. B3c1]. After O1A is promoted by amino acid linkage priority, O2A and O3A are assigned by prochirality rules (O3 > O1A > O2A > O3A).

In the case of human protein tyrosine phosphatase **ptpn5** (C472S mutant), the tyrosine phosphate moiety is coordinated to residues in the loop Ala474-Arg478 (PDB entry: **2CJZ**). Consideration of hydrogen bonds ≤ 3.0 Å shows oxygen O1P coordinated to Gly^{476} and Ile^{477}; oxygen O2P coordinated to Ala^{474} and Arg^{478}; and oxygen O3P coordinated to Arg^{478}. Thus we can now designate O1A as being coordinated to the lowest numbered amino acid, Ala^{474} (it is labelled as O2P in **2CJZ**).[j] The oxygen atom priority is O4' > O1A > O2A > O3A, in which O2A and O3A are designated by CIP rules for prochirality as shown (O2A being *pro-R* and O3A is *pro-S*) [Fig. B3c2]. (**NB** There are hydrogen H-bonds from Ser^{472}(OH) to O2P and O3P but both are longer than 3.0 Å and thus are ignored).

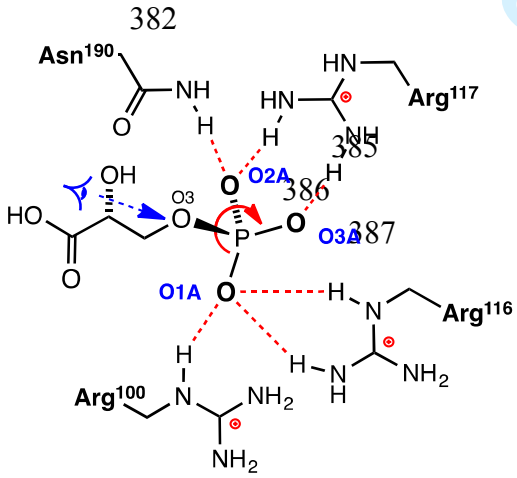


Figure B3c1

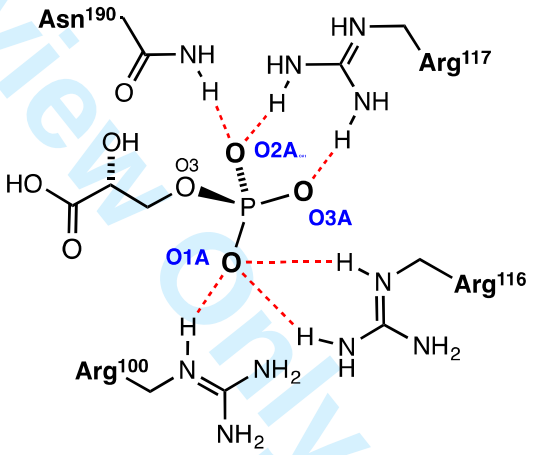


Figure B3c2

[g] 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.

[h] For CIP Rules see the IUPAC Blue Book p92. For the use of *pro-R* and *pro-S* see "Basic Terminology of Stereochemistry (IUPAC Recommendations 1996)" *Pure Appl. Chem.* **68**, 2193-2222 (1996).

[j] 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.

B4. Terminal Phosphates in Molecules with Multiple Phosphate Chains

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

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 oxygens of PA2 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 O1A2. The oxygen atom priorities for PA2 are thus O2 > O1A2 > O2A2 > O3A as shown [Fig. B4a].

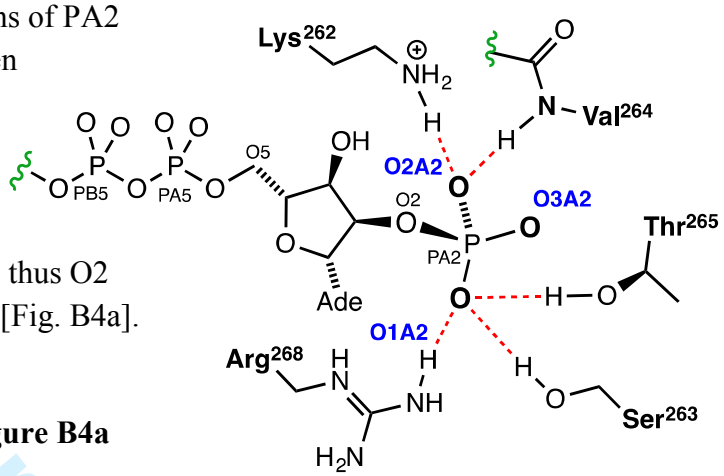


Figure B4a

B5. Isolated Single Phosphates

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

a) **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 oxygens are determined by H-bonds to amino acid residues. Ignoring H-bonds ≥ 3.0 Å, the structure identifies O1 coordinated to Ser^{30}(OH), O2 coordinated to Gly^{79}(NH), and O3 coordinated to Lys^{34}(NH_3^+). O4 is only coordinated to ligands at distances ≥ 3.0 Å (*oxygens numbered as in* **1Z0D**) (Fig. B4b left). Hence, the priority order is O1 > O3 > O2 > O4. Assigning the top two oxygen priorities as O1P and O2P respectively (Fig. B4b right) makes the two remaining oxygens a prochiral pair. Promoting the 'front' oxygen to ^{18}O gives phosphorus *S* chirality, thus identifying it as *pro-S*. By a similar analysis, the 'rear' oxygen is *pro-R*. Hence, the rear oxygen can be designated O3P and the front oxygen is O4P (Fig. B4b right) (**NB** *The PDB file assigns PA and PB to the GDP ligand*).

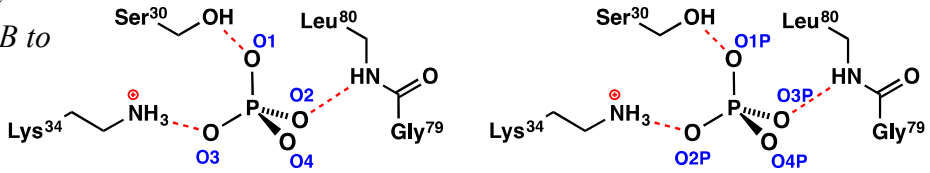


Figure B4b

5. RECOMMENDATIONS FOR LABELLING FLUORINE AND OTHER ATOMS IN PHOSPHATE TRANSITION STATE ANALOGUES

C1. Tetrahedral Phosphate Mimics – Trifluoroberyllates

These use a two-symbol code that may be expanded to four when there are additional fluorines 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 fluorines will follow the same rules as for the three oxygens in a terminal tetrahedral phosphate. Prochirality identification can be applied if one of the three fluorines is promoted relative to the other two. In the context of enzyme-bound trifluoroberyllates, such promotion can be generally be identified by co-ordination of one of the fluorines 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 F1Be. The prochiral fluorines F2Be and F3Be 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 F1, F2, and F3 respectively.
(*Note, that in PDB entry **2WF8**, these fluorines were labelled F3, F1, and F2 respectively).*

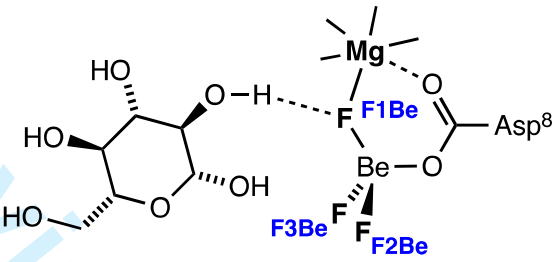


Figure C1a

There is one example of a BeF_2 moiety bridging two anionic oxygens. In this case, F1Be and F2Be 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 O3B, and to UDP O3G [Fig. C1b]. The (*pro-R)* fluorine is thus F1Be and the (*pro-S*) fluorine is F2Be.[k] In this unique and rather complicated example, CIP rules give priority to O5' over O5''' since adenine (A) takes priority over uridine (U) (Section A1d).

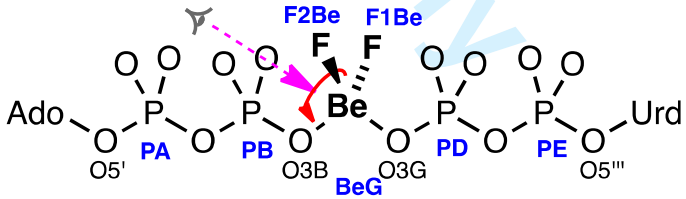


Figure C1b

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

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, and 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.

C2. Trigonal Bipyramidal Phosphate Transition State Analogues – Trifluoromagnesates and Aluminium 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 fluorines are invariably equatorial with two axial oxygen ligands to the 5-coordinate metal. Priority identification can be applied when one of the three fluorines 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 oxygens over the second and 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 O3B and to a water, and CIP priority gives O3B > OH_2. Thus the fluorine coordinated to the catalytic magnesium is designated F1Mg. F2Mg and F3Mg are then identified in a clockwise progression from F1Mg when viewed from O3B to Mg [Fig. C2a].[1]

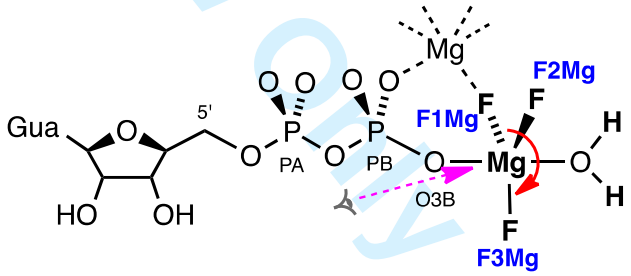


Figure C2a

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

b) In case of multiple metal ion coordination, and 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 fluorines 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 magnesiums, 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 O3B > O-$Ser^{21'}$. 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 thus coordinated to F1Mg; F2Mg and F3Mg follow in clockwise progression [Fig. C2b].

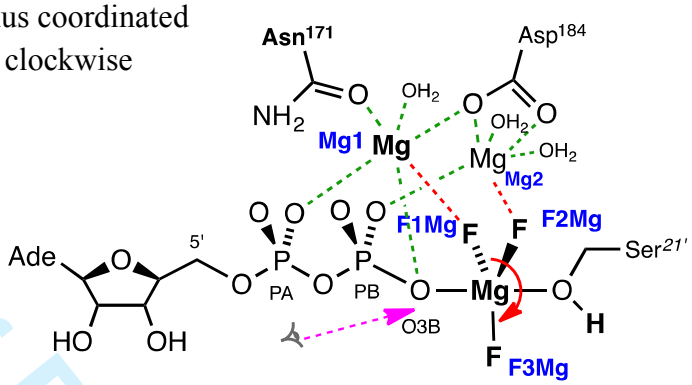


Figure C2b

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).[m]

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}-CO_2^-.MgF_3^-.OH_2 complex. This can be labelled as for C2a (above) using coordination to a catalytic Mg to give priority to F1.[n]

[m] 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).

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

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 fluorines are invariably equatorial with two *trans*-oxygen ligands to the 6-coordinate aluminium. Priority identification can be applied by promoting one of the four fluorines 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 oxygens 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 F1Al while F2Al, F3Al, and F4Al follow in clockwise progression viewed from Asp^8 □d□, which has priority over glucose O6. (**NB** *The corresponding PDB designations are F2, F1, F3, and F4 respectively*)[p] [Fig. C3a].

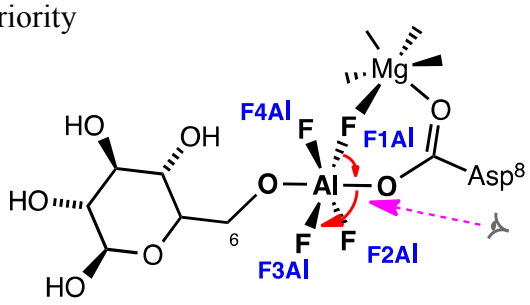


Figure C3a

b) There are (PDB to 2015) ≥ 3 examples of octahedral trifluoroaluminate complexes 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 O3A > $OWat^{401}$ [Fig. C3b]. One fluorine is coordinated to two catalytic magnesiums and so is designated F1B. A progression viewed in the priority direction then identifies the bridging oxygen as the second priority ligand, O1B, with F2B and F3B completing the clockwise equatorial sequence.

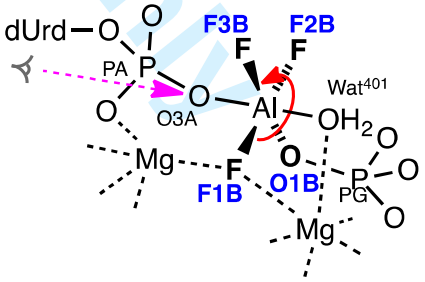


Figure C3b

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

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 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 $His^{253}N\square\square$ has been described (PDB entry: **3QPW**. Fig. C3c). This octahedral complex is completed by water coordination *trans* to the histidine nitrogen. The four fluorines 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} (*this fluorine numbering in superscript is as used in* **3QPW**). As F^2 is coordinated to Ne of Arg^{252} and F^4 is coordinated to Arg^{252}-N□1, F^2 takes priority as its H-bonding is to the nitrogen nearer to Cα of the lowest numbered coordinating amino acid. Hence F1Al is coordinated to Arg^{252} and Gln388 and the progression to F2Al, F3Al, and F4Al proceeds clockwise as viewed from the water apex of the octahedral complex (CIP priority is O > N, magenta arrow) [Fig. C3c].[q]

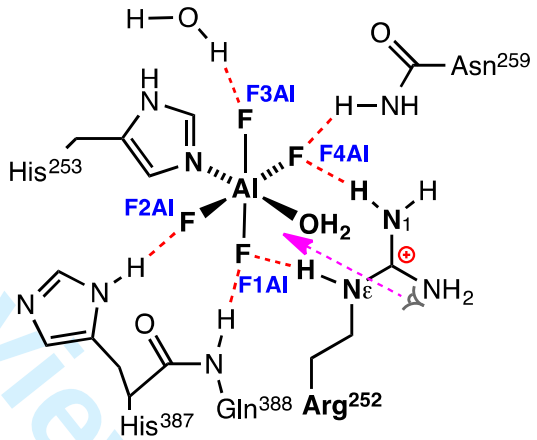


Figure C3c

[q] 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).

6. RECOMMENDATIONS FOR LABELLING VANADATE AND TUNGSTATE ANALOGUES OF PHOSPHATES

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 oxygens are numbered O1V, O2V, and O3V with the axial oxygen O4V being *trans* to the hydroxylic oxygen of Thr^{90} [Fig. C4a]. The equatorial oxygen coordinated to two zinc ions takes priority and is O1V. The direction of viewing is determined by the priority Thr^{90} oxygen > OV4 (magenta arrow). Thus a clockwise progression identifies O2V at the front and O3V at the rear of the trigonal planar array.

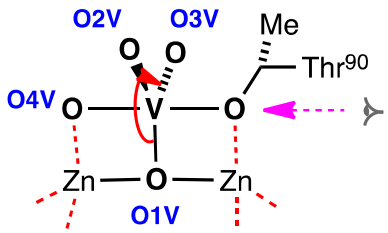


Figure C4a

b) **Disubstituted Vanadate.** A transition state analogue complex for phosphorylation of glucose 1-phosphate on O6 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 O6G > O3S. The three equatorial oxygens take priority from O1V by its coordination to cobalt, substituting for the native catalytic magnesium. Assignment of O2V and O3V follows a clockwise progression when viewed from O6G (magenta arrow) [Fig. C4b].

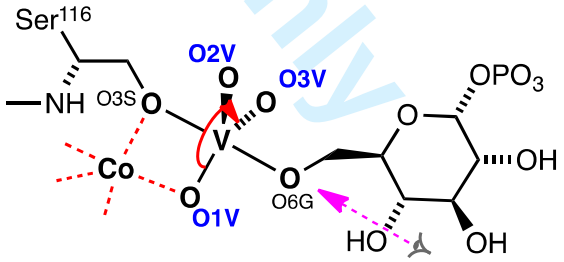


Figure C4b

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 oxygens. Thus, oxygen H-bonded to Lys^{13} takes priority as O1V over oxygen O2V H-bonded to Arg^{94}, while O3V 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□2. Equatorial ligand priority is assigned to Thd-O3'. Hence O2V and O3V follow in a clockwise progression when viewed from the Tyr-oxygen [Fig. C4c].

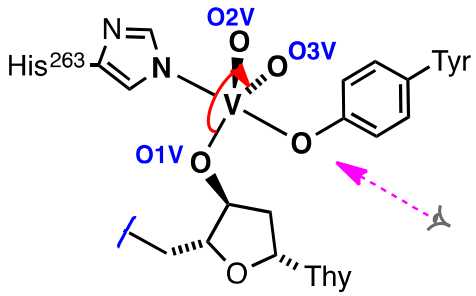


Figure C4c

c) **Cyclic Trisubstituted Vanadate.** Trisubstituted vanadate provides a transition state analogue structure for hairpin ribozyme cleavage of a phosphodiester (PDB entry: **1M5O**). The axial O2' has CIP priority over the axial O5'. Priority in the three equatorial oxygens is taken by the ribose O3' leading to assignment of O1V followed by O2V in a clockwise progression [Fig. C4d].

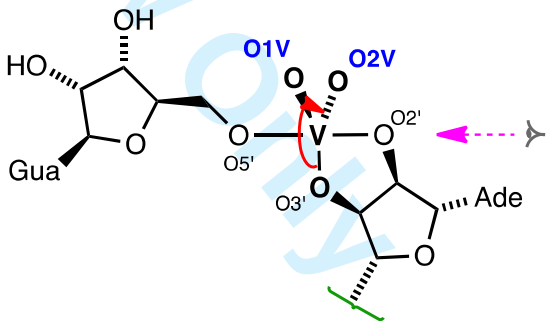


Figure C4d

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 oxygens need to be assigned priority to enable the remaining two to be assigned by prochirality rules.

b) **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 oxygens coordinated to these metal ions are labelled O1W and O2W respectively (Fig. C5a). The remaining two tungstate oxygens are now prochiral and can be labelled O3W and O4W by CIP rules described above.

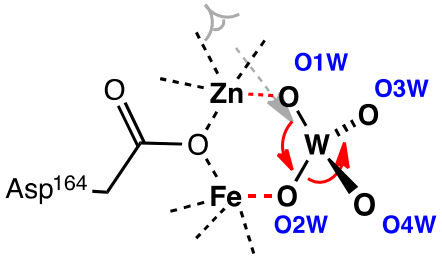


Figure C5a

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 O1W is identified by its coordination to tungsten. 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 O2W is coordinated to Asp59 and takes precedence over the third oxygen that is coordinated to Thr87. The remaining two tungstate oxygens are now prochiral and can be labelled O3W and O4W by CIP rules described above (Fig. C5b).

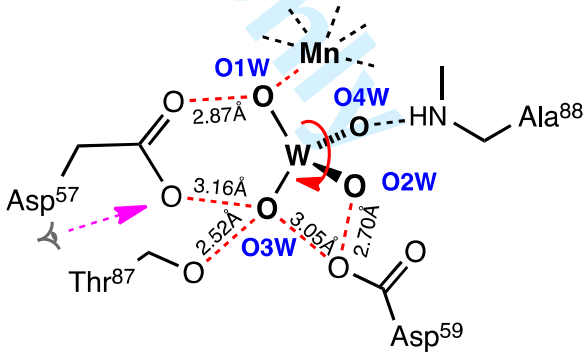


Figure C5b

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 O1W and O2W. Then O3W and O4W can 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 oxygens coordinated to nitrogens. As isolated water coordination is ignored,[r] priority is given to coordination from Arg^{404} to O1W followed by coordination from Val^{407} to O2W. Hence, O3W and O4W are now prochiral and can be assigned using CIP rules (Fig. C5c).[l]

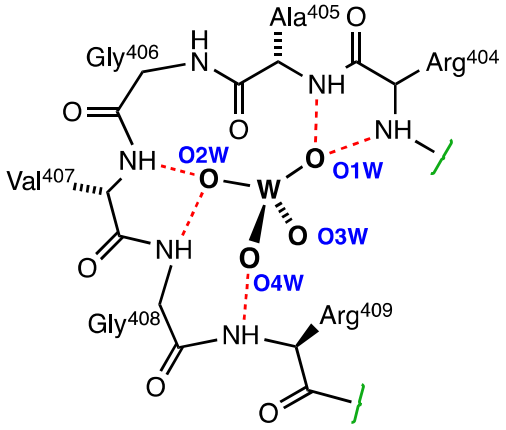


Figure C5c

c) **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 of both tungstens and of seven oxygens, since the first tungstate is a trigonal bipyramidal TSA of Pα and the second tungstate is a tetrahedral analogue of Pβ of ADP. Tungsten WA is equatorially linked to the adenosyl 5'-oxygen and axially linked to His^{112}-Nε2. As in the case of polyphosphates (Section B1c), the bridging oxygen to WB is designated O3A. That enables assignment of the two prochiral equatorial oxygens as O1A and O2A (when viewed in the axial direction O3A to Ne2). For WB, oxygen O3A has highest CIP priority because it is coordinated to WA. The oxygen coordinated to Gly^{105} takes precedence over the oxygen coordinated to Ser^{107} and is therefore identified as O1B. This enables the prochiral pair of oxygens to be assigned as O2B and O3B as shown (Fig. C5d)

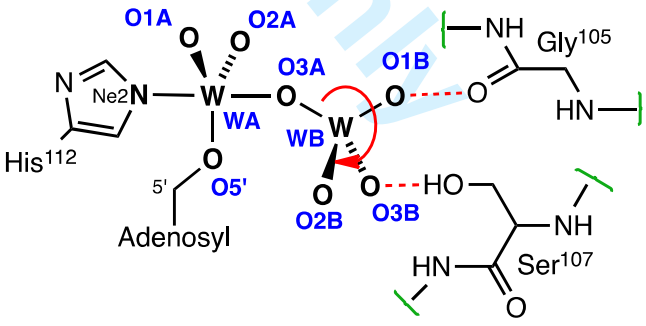


Figure C5d

[r] Note that once coordination has reduced the number of non-prioritised oxygens to two, this pair is assigned by application of CIP rules on prochirality.

7. SUMMARY

The recommendations presented here have been developed to describe molecules derived from orthophosphoric acid and its derivatives, analogues, and transition state analogues. In our hands, they have worked well for the most demanding species we have examined, e.g. **C3c** and **C5d**. However, we recognise that they 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-existant, structures that could require an extension of these recommendations, and we are receptive for advice on such problems.

8. REFERENCES AND NOTES

[1] J. R. Knowles. *Annu. Rev. Biochem.* **49**, 877-919 (1980).

[2] G. Lowe. *Accts Chem. Res.* **16**, 244-251 (1983).

[3] R. S. Cahn, C. Ingold, V. Prelog. (1966). *Angew. Chem. Internat. Ed. Engl.* **5**, 385-415 (1966).

[4] G. A. Orr, J. Simon, S. R. Jones, G. J. Chin, J. R. Knowles. *Proc. Natl. Acad. Sci. USA* **75**, 2230-2233 (1978); K-F. Sheu, P. A. Frey. *J. Biol. Chem*. **253**, 378-380 (1978); R. L. Jarvest, G. Lowe. *J. Chem. Soc. Chem. Commun.* 364-366 (1979).

[5] S. R. Jones, L. A. Kindman, J. R. Knowles. *Nature* **275**(5680), 564-565 (1978).

[6] M. W. Bowler, M. J. Cliff, J. P. Waltho, G. M. Blackburn. *New J. Chem*. **34**, 784-794 (2010).

[7] S. C. Kamerlin, P. K. Sharma, R. B. Prasad, A. Warshel. *Quart. Revs. Biophys.* **46**, 1-132 (2013).

[8] 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**, 14732-14737 (2006); 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. Amer. Chem. Soc.* **130**, 3952-3958 (2008); 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. Engl.* **51**, 12242-12245 (2012).

[9] J. Mishra, U. S. Bhalla. *Biophys. J.* **83**, 1298-1316 (2002).

[10] M. A. G. Sillero, A. de Diego, E. Silles, F. Pérez-Zúñiga, A. Sillero. *FEBS Lett.* **580**, 5723-5727 (2006).

[11] Nomenclature of Phosphorus-Containing Compounds of Biochemical Importance (Recommendations 1976) IUPAC-IUB Commission on Biochemical Nomenclature, *Proc. Natl. Acad. Sci. USA* **74**, 2222-2230 (1977).

[12] 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)).

[13] Abbreviations and symbols for the description of conformations of polynucleotide chains. Recommendations 1982, *Eur. J. Biochem*. **131**, 9-15 (1983).

[14] Newsletter of the Nomenclature Committees of the International Union of Biochemistry and Molecular Biology, *Eur. J. Biochem.* **122**, 437-438 (1982). [also ref 12, p. 265]

[15] IUPAC Nomenclature of Organic Chemistry, IUPAC recommendations and preferred names 2013, *Royal Society of Chemistry* (2013). (a) p-93.2.4 p. 1215-1216, (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; p432ff P-42.3-4 Phosphorus acids etc.; p915ff P-67 Phosphorus acids etc.; p992ff P-68.3 Phosphorus compounds; p1215ff P-93.2.4 Stereochemistry of phosphates etc.; p1420ff P-105 Nucleosides; p1425ff P-106 Nucleotides.

[16] *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; p180ff, IR 9.3.3.4, octahedral coordination; p184ff, IR 9.3.3.6, bipyramidal coordination; p189ff, IR 9.3.3.8, chiral octahedral; p190ff, IR 9.3.3.10, chiral bipyramid.

9. APPENDIX

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 oxygens 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.[s] 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 as anticlockwise progression is *S*. Because **A** and **B** are enantiomers, the two non-bridge oxygens are enantiotopic. In extension, the paired, non-bridge oxygens can be labelled (*pro-R*) for the front one (clockwise progression) and (*pro-S*) for the rear one (**C**, right).

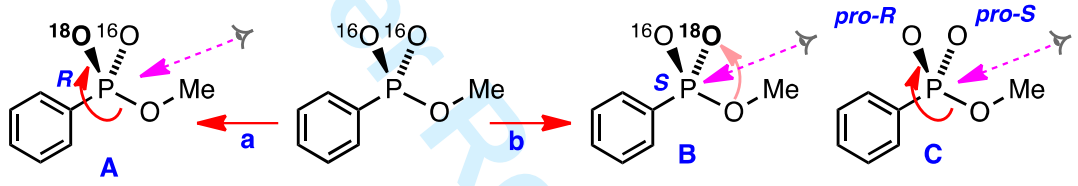
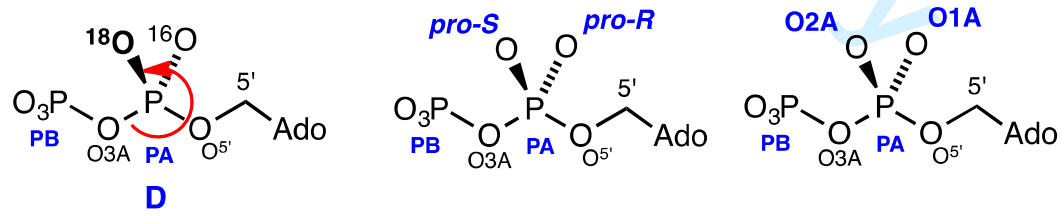


Figure X1

2) Two non-bridging oxygens 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.[t] In the case of adenosine 5'-diphosphate (ADP), the two non-bridging paired oxygens on PA are diastereotopic. Promoting the 'front' oxygen to ^{18}O generates a new stereogenic centre at PA (**D**; CIP label *S*) while promoting the 'rear' oxygen to ^{18}O generates a stereogenic centre with the opposite sense at PA (**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. As they are *not* enantiomers they are therefore termed diastereoisomers.[u]



[s] These oxygens are spectroscopically and chemically non-equivalent in a chiral environment.
[t] These oxygens are spectroscopically and chemically non-equivalent in *any* environment.
[u] The term diastereoisomer simply describes all stereoisomers that are *NOT* enantiomers

Figure X2

3) We can now apply the priority rules described in Section X1 to label the non-bridging oxygens on PA in ADP. This analysis begins with the CIP priority rule that ranks di-coordinate oxygen above mono-coordinate oxygen. Thus O5' and O3A rank above the two non-bridging oxygens. For these bridging oxygens, 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 PB and C5'. Hence, O3A has priority over O5' as P has a higher atomic number than C. The CIP priority ranking is thus O3A > O5' > ^{18}O > ^{16}O (Fig. X3).[v]

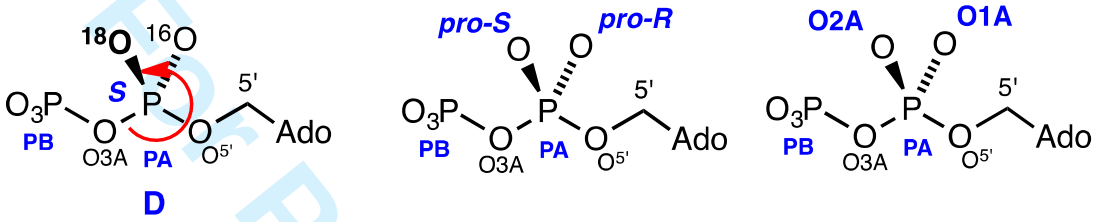


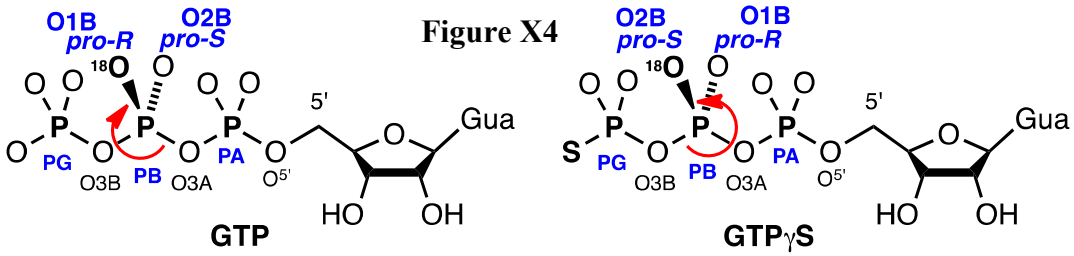
Figure X3

Viewing the P-centered tetrahedron in stereoisomer (**D**) from the face with ^{16}O at the rear gives an anticlockwise progression from high to low priority ligands (Fig. X3 left) and so PA in **D** has *S* chirality. Hence, the two paired-oxygens in ADP can be labelled *pro-S* for the front one (*as its promotion to ^{18}O makes PA an* S *chiral centre*) and *pro-R* for the rear one (*as its promotion to ^{18}O makes PA an* R *chiral centre*) (Fig. X3 center). We can now use CIP Rule 5 that gives *R* priority over *S*. Thus the *pro-R* oxygen is labelled O1A and the *pro-S* oxygen is labelled O2A (Fig. X3 right).

4) The accurate application of the CIP rules inevitably means that there are some unexpected outcomes. For example, the stereochemistry of the non-bridging oxygens at PB in guanosine 5'-triphosphate (GTP) and in γ□□□□□□□□ (**GSP**) have opposite assignments.

For GTP, the rules for the in-chain atoms flanking PB identify O5' bonded to C5' thereby taking priority over all oxygens bonded to PG (O1G, O2G, O3G). Hence, the priority sequence for the four GTP oxygen ligands at PB is O3A > O3B and thus the front oxygen is *pro-R* and the rear oxygen is *pro-S* (Fig. X4 left). Hence the *pro-R* oxygen is labelled O1B and the *pro-S* oxygen oxygen is O2B (Fig. X4 left).

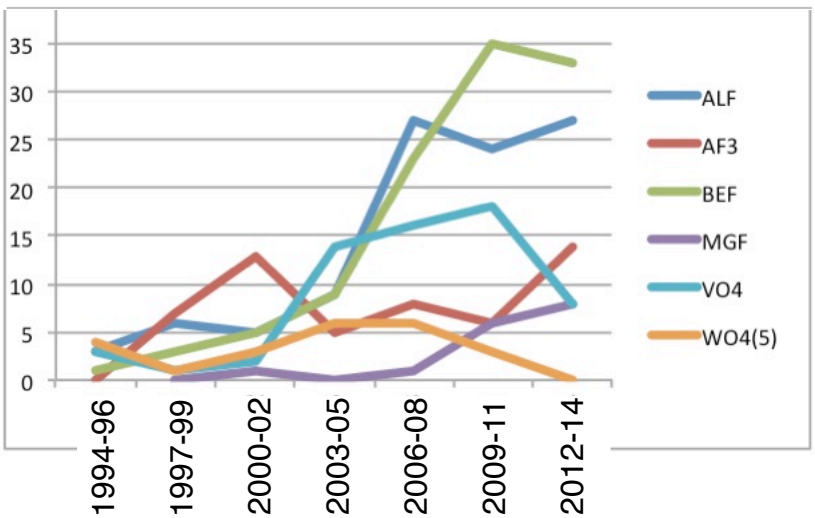
Figure X4

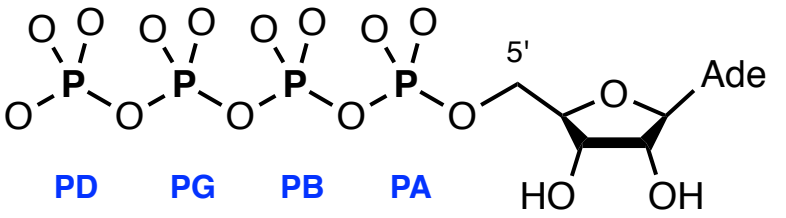


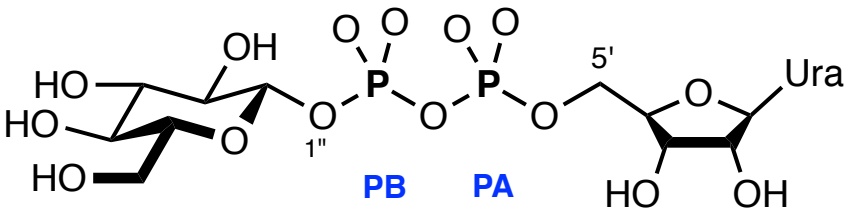
[v] **NB** Labelling a non-bridge oxygen with ^{18}O only gives it priority over the ^{16}O oxygen. It does not change its priority relative to the non-bridging oxygens.

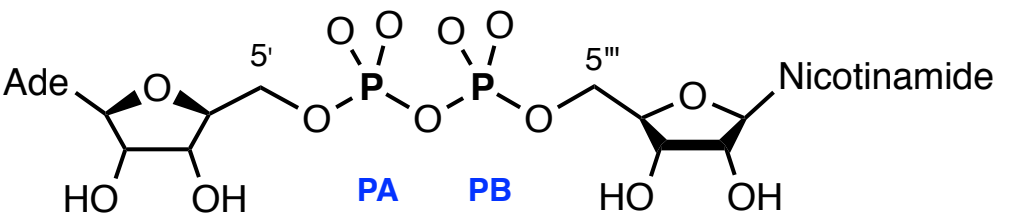
By contrast, for GTPγS, the sulfur atom on PG takes CIP priority over O5' with the consequence that O3A takes priority over O3B (Fig. X4 right). The result is that in GTPγS (as presented) the rear oxygen is *pro-R* (and thus O1B) and the front oxygen is *pro-S* (and thus O2B) (Fig. 4X right). This is the opposite 3D spatial outcome compared to GTP.

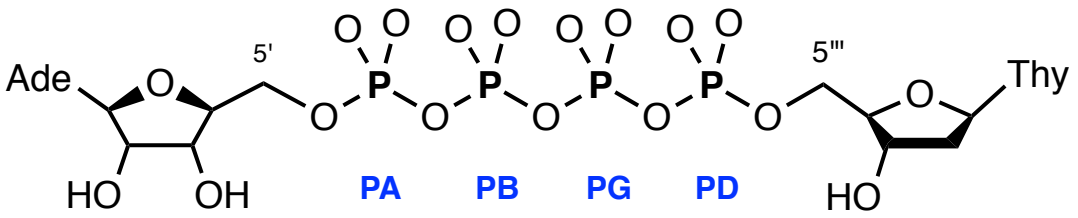
We can note that a similar situation will hold for GTPγF but not for γ-amino-GTP.

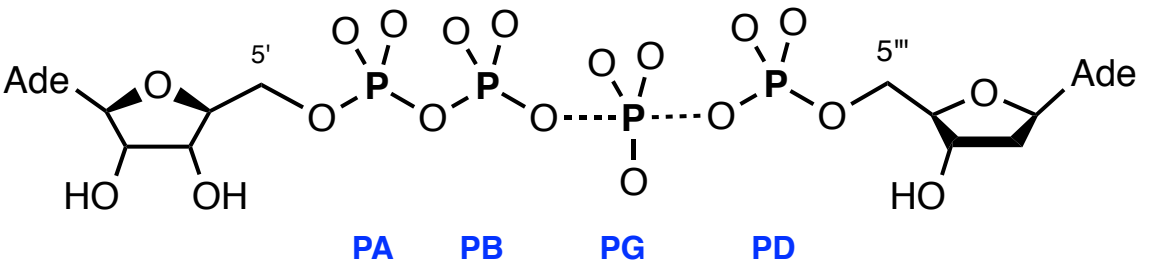

35
30
25
20
15
10
5
0
1994-96
1997-99
2000-02
2003-05
2006-08
2009-11
2012-14
ALF
AF3
BEF
MGF
VO4
WO4(5)

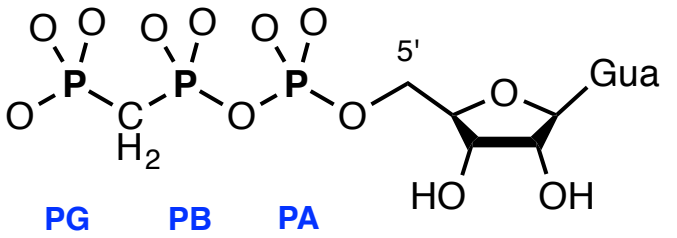

5'
Ade
HO
OH
PD
PG
PB
PA

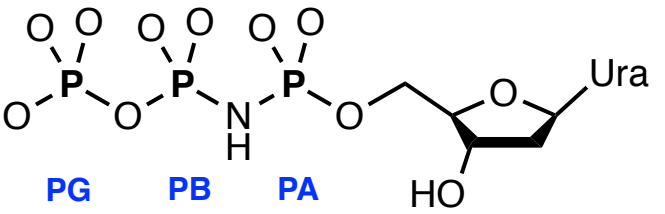

OH
HO
HO
HO
O
O
O
P
O
O
P
O
O
O
O
1"
5'
Ura
HO
OH
PB
PA

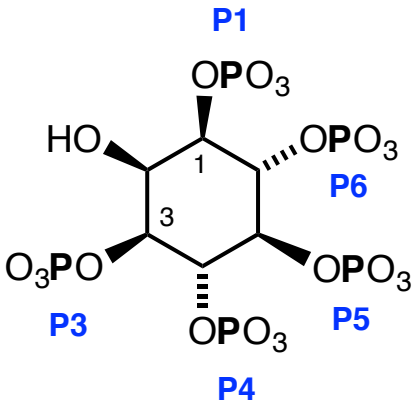

Ade
O
5'
O
O
P
O
O
P
O
O
O
5'''
O
Nicotinamide
HO
OH
PA
PB
HO
OH

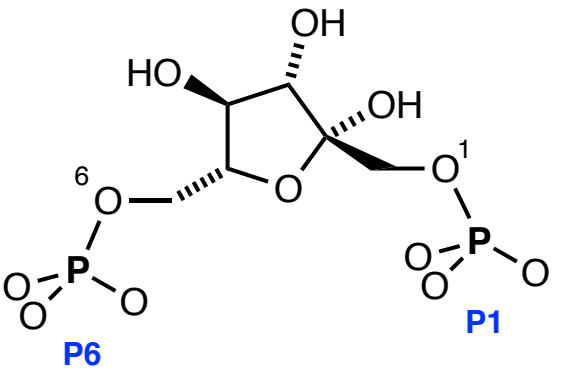

Ade
O
5'
O
P
O
P
O
P
O
P
O
5'''
O
Thy
HO
OH
PA
PB
PG
PD
HO

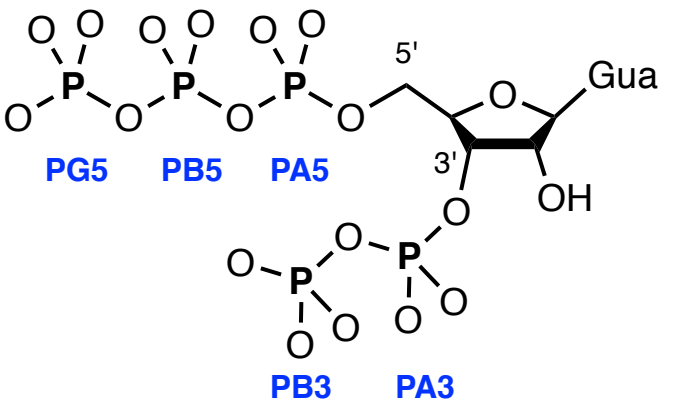

Ade
HO
OH
5'
5'''
Ade
HO
PA
PB
PG
PD

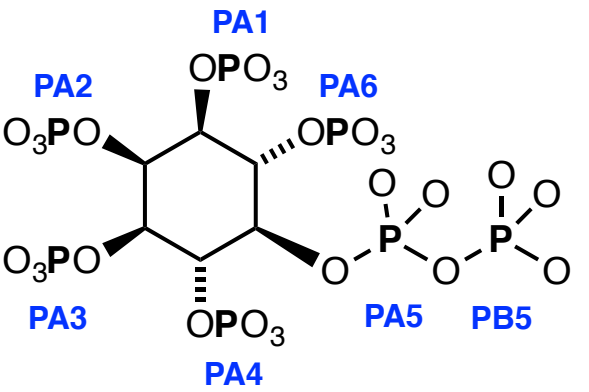

O
O
O
O
O
O
P
P
P
O
C
H2
O
O
5'
O
Gua
HO
OH
PG
PB
PA


O
O
O
O
O
O
P
P
P
O
O
O
N
H
O
O
Ura
HO
PG
PB
PA


P1
OPO3
HO
OPO3
1
P6
3
O3PO
OPO3
P3
P5
OPO3
P4


OH
HO
OH
6
O
O
1
O
P
O
O
O
P
O
O
P6
P1


PG5
PB5
PA5
PB3
PA3
5'
3'
Gua
OH


PA1
PA2
PA6
PA3
PA4
PA5
PB5