Metal Fluorides as Analogs for Studies on Phosphoryl Transfer Enzymes

Yi Jin[a], Nigel G. Richards[b], Jonathan P. Waltho[c], and G. Michael Blackburn[d]*
Abstract: In 1994, the protein structure of a transition state analogue for G1x, a small G protein, heralded a new field of research into the structure and mechanism of enzymes that manipulate transfer of the phosphoryl (PO\textsubscript{4}\textsuperscript{3-}) group. It was based on a protein complex of GDP and AlF\textsubscript{4} that mimicked the transition state for hydrolysis of GTP. The growing list of enzyme structures that embrace metal fluorides, MF\textsubscript{x}, as ligands that imitate either the phosphoryl group or a phosphate, now exceeds 80 per triennium. They fall into three distinct geometrical classes: (i) Tetrahedral complexes based on BeF\textsubscript{4}\textsuperscript{-}, mimic ground state phosphates, (ii) Octahedral complexes, primarily AlF\textsubscript{6}\textsuperscript{3-}, mimic “in-line” anionic transition state for phosphoryl transfer, and (iii) Trigonal bipyramidal complexes additionally mimic the tbp stereochemistry of the transition state and are represented by MgF\textsubscript{3}\textsuperscript{-} and putative AlF\textsubscript{3}\textsuperscript{3-} complexes. Their interpretation has provided a deeper mechanistic understanding of the behavior and role of phosphate monoesters in molecular biology. This review challenges the existence of AlF\textsubscript{3}\textsuperscript{3-} and MgF\textsubscript{3}\textsuperscript{-} as real species in protein complexes and questions the relevance for enzymes of physical organic chemistry and model studies that are water-based for phosphoryl group transfer. It proposes a new interpretation of the role of general acid-base catalysis.

1. Introduction

For a decade following their discovery, the atomic structures of proteins containing a metal fluoride (MF\textsubscript{x}) species were based primarily on geometric considerations. From 2003 onwards, this resulted in a growing uncertainty about their chemical constitution. Now, \textsuperscript{19}F NMR analysis of these complexes has been used firstly to analyze and identify their atomic composition, secondly to deliver experimental measurements of the electronic environment provided by the protein in conformations close to the transition state (TS), and lastly to identify a significant number of mis-assignments, thereby providing a corrective critique for past errors and future uncertainties.

There are now over 500 MF\textsubscript{x} structures in the Protein Data Bank (PDB) (Fig. 1). The molecular analysis of these structures has established a simple, logical, and rational understanding of the chemical constitution of transition state analog (TSA) and ground state analog (GSA) structures of MF\textsubscript{x} complexes. It provides a firm base for understanding enzymatic mechanisms for the catalysis of phosphate monoesters and anhydrides, notably ATPases, GTPases, kinases, mutases, phosphatases, and phosphohydrolases.\textsuperscript{[1]} They all employ “in-line” geometry, they are concerted, and they utilize tight control of hydrogen bonding in the active site complex to disfavor the formation of hydrogen bonds that would inhibit the chemical step in catalysis. Some of

*[a] Dr. Y. Jin, Department of Chemistry, University of York, York, YO10 5DD(UK)
[b] Prof. Dr. N.G. J. Richards School of Chemistry Cardiff University Cardiff, CF10 3AT (UK)
[c] Prof. Dr. J. P. Waltho Manchester Institute of Biotechnology Manchester, M1 7DN (UK)
[d] Prof. Dr. J. P. Waltho, Prof. Dr. G. M. Blackburn Department of Molecular Biology and Biotechnology, University of Sheffield Sheffield, S10 2TX (UK)
email: g.m.blackburn@sheffield.ac.uk

Michael Blackburn is Emeritus Professor of Biochemical Chemistry in Sheffield University and a founding member of the Krebs Institute. His undergraduate and post-doctoral career in Cambridge University led him in 1961 into the biological chemistry of phosphorus under the guidance of Prof. J.P. Waltho and G. M. Blackburn in the University of Sheffield, UK. In 2014, she moved to the group of Prof. G. J. Davies in YSB, University of York, UK. Her research interests involve mechanistic studies of disease-relevant phosphonyl transfer enzymes and carbohydrate processing enzymes using chemical, NMR, molecular biology, and crystallographic approaches.

Jonathan Waltho is the Gibson Professor of Biophysics at the University of Sheffield, and Professor of Structural Biology at the University of Manchester. He received his undergraduate degree from Durham University and his doctoral degree from Cambridge University. Following post-doctoral studies at SmithKline French and The Scripps Research Institute, he has tackled a wide range of problems in protein folding, misfolding, and protein-ligand interactions. His current research interests focus on structure, electronics and dynamics in phosphoryl, methyl and hydride transfer enzymes.
these structures have been starting points for multiple studies on enzyme mechanisms using QM-MM and DFT analysis. More controversally, the present analysis challenges the extrapolation to enzymes of physical organic chemistry, of model studies, and the role of general acid-base catalysis (GABC) of phosphoryl transfer in aqueous solution. structures, well supported by 19F NMR data and the best of many computational developments.

Phosphoric acid (H₃PO₄), its esters, amidates, and anhydrides share a common tetrahedral geometry based on a phosphorus (V) core linked near-symmetrically to four oxygens or nitrogens. Biological phosphoryl transfer reactions call for the relocation of a phosphoryl group, PO₃⁻, from a donor to an acceptor atom, typically N, O, or S and more rarely C or F. There are many reviews of this activity and its catalysis, but there is no consensus on whether the reactions are more associative (tight TS) or more dissociative (loose TS) in character, a description with a boundary value of 4.9 Å based on van der Waals considerations (Scheme 1).[1a] In either case, the TS for such phosphoryl group transfer will have trigonal bipyramidal (tbp) geometry, with axial dimensions that relate to the tight or loose nature of the TS. Because the primary database for MFₓ complexes is structurally driven, we review the separate groups of MFₓ protein complexes in terms of their geometry. This has the additional advantage of overriding ambiguities in the assignment of composition, as shown below.

2. Tetrahedral Phosphate Mimics, BeF₃⁻

Beryllium (II) forms stable fluorides in water that exist as a mixture of tetrahedral species including BeF₂⁻·H₂O, BeF₃⁻·H₂O, and BeF₂O⁻. Early NMR studies on fluoroberyllate complexes with ADP led to analysis of mixed fluoroberyllate-ADP species with myosin, and the first x-ray analysis of a fluoroberyllate protein structure was delivered in 1995 for an ADP·BeF₃⁻ complex with myosin (PDB: 1mm).[2] Since then, 122 fluoroberyllate complexes have been described, with 3 solved by NMR and 119 x-ray structures having resolutions ≥ 1.2 Å. The vast majority of these structures have a tetrahedral fluoroberyllate bonded to anionic oxygen. They divide into two principal groups: over 70 are coordinated to an aspartate carboxylate (including the 3 NMR structures) and around 50 are coordinated to a nucleotide terminal phosphate. Only 2 are coordinated to a histidine ring nitrogen.

2.1. Aspartyl fluoroberyllates

These structures share a common core, with bidentate coordination to an essential metal ion, generally Mg²⁺ and rarely Mn²⁺, from fluorine F₁ and the second carboxylate oxygen, OD₂, giving a near planar six-membered ring (Fig. 2). There are many reviews of this activity and its catalysis, but there is no consensus on whether the reactions are more associative (tight TS) or more dissociative (loose TS) in character, a description with a boundary value of 4.9 Å based on van der Waals considerations (Scheme 1).[1a] In either case, the TS for such phosphoryl group transfer will have trigonal bipyramidal (tbp) geometry, with axial dimensions that relate to the tight or loose nature of the TS. Because the primary database for MFₓ complexes is structurally driven, we review the separate groups of MFₓ protein complexes in terms of their geometry. This has the additional advantage of overriding ambiguities in the assignment of composition, as shown below.

Figure 1. Number of MFₓ structures published in the PDB triennially, vanadate data included for reference (data for 01/15 through 06/16 normalized by x2 to represent a triennial figure).

Figure 2. 17 Aligned aspartyl-trifluoroberyllate structures. BeF₃⁻ is locked in a 6-membered ring (center); Catalytic Mg²⁺ (rarely Mn²⁺) and an aspartate (usually Asp57) fuse a 13-membered ring to the fluoroberyllate ring with atoms from the adjacent two amino acids downstream (rear center). Octahedral coordination to Mg is completed by an additional aspartate (right), by 1-2 waters, but only twice by histidine (upper right). (Atom colors: fluorine, light blue; beryllium, yellow-green; nitrogen, blue; oxygen, red). In 7 structures, an isolated water (red spheres) is distantly related to one fluorine. (Electron densities presented in CCP4MG from mtz data in EDS and contoured at 1σ)

Scheme 1. Concerted phosphoryl transfer. Top, bond making precedes bond breaking (blue); center, bond breaking balanced by bond making (black); bottom, bond breaking in advance of bond making (red).
2.2. ADP·BeF$_3^-$ structures

There are 42 x-ray structures of BeF$_3^-$ complexes with ADP and 6 with GDP, which constitute isosteric mimics of ATP and GTP respectively. They are distributed among kinases, hydrolases, mutases, helicases, and small G proteins. Of the ADP-BeF$_3^-$ structures, 25 are resolved at ≤ 2.5 Å and 20 align remarkably well (Fig. 3). The beryllium is bonded to O3B and a catalytic Mg$^{2+}$ is coordinated to F1 and to O1B in a 6-membered ring. There is remarkable consistency in neighboring amino acids; an arginine and a lysine coordinate β and γ and balance the anionic charge of the nucleotide. By contrast, the adenine base occupies a range of remarkable consistency in neighboring amino acids; an arginine oxygen being ~3.4 Å from the beryllium, with a median “cone” angle of 158˚, and forming a H-bond to one of the fluorines. These waters lie well within the BeF$_3^-$ “cone” with their oxygen being ~3.4 Å from the beryllium, with a median “in-line” angle of 158˚, and forming a H-bond to one of the fluorines (2.82 ± 0.27 Å). As the axial O—Be—O distance is close to 5.1 Å, these waters are part of a Near Attack Conformation (NAC) that is intermediate between a ground state (GS) and a TS situation.[6]

The 6 GDP structures are very similar in structure to the ADP complexes but at rather lower resolution (SI Table 3). The BeF$_3^-$ complex for human phosphoglycerate kinase (hPGK) raises the question: “Where is the beryllium in the case of two fluorines?” The structure of the complex hPGK-ADP-BeF$_3^-$-3PG (PDB: 4axx, 1.74 Å resolution) places the Be atom 1.73 Å from the carboxylate oxygen and 2.85 Å from the ADP oxygen O3B. However, the three fluorines are on average 2.75 Å from the carboxylate oxygen and 2.96 Å from the ADP oxygen (Fig. 4A). As the sum of van der Waals radii for Be and O is 3.26 Å, these data suggest mixed occupancy with beryllium closer on average to the carboxylate.[7]

Various approaches to analogs of γ-phosphohistidine have been explored. Work on nicotinamide phosphoribosyltransferase (NAMPT) has structurally mimicked phosphorylation of pyrophosphate from C1 of ribose 5-phosphate (reactants in purple, products in silver, red curly arrow shows departure of phosphoryl oxygen). Structures of 2 overlaid complexes show BeF$_3^-$ bonded to Nc of His247 and one fluoride coordinating octahedral Mg$^{2+}$ (green sphere). C1′ of PRPP in reactant (purple sphere) moves 1.8 Å to bond the nicotinamide N1 (silver sphere), Reactant purple sticks, product silver sticks, Be in yellow-green.

2.3 Histidine trifluoroberyllates

Various approaches to analogs of γ-phosphohistidine have been explored. Work on nicotinamide phosphoribosyltransferase (NAMPT) has structurally mimicked phosphorylation of an active-site histidine using trifluoroberyllate. Crystal structures of NAMPT for reactant and product complexes (Fig. 4B) have a covalent His247·BeF$_3^-$·PRPP·Arg complex (Fig. 4B) have a covalent His247·BeF$_3^-$ reaction adduct.[8] In contrast to all other trifluoroberyllate structures, it is coordinated to one fluorine without any direct linkage to His247.[8]

2.4 A nucleotide beryllium difluoride structure

A solitary example of beryllium difluoride bridging ADP and UDP illuminates the activity of UMP/GMP kinase.[9] The 2.0 Å structure (Fig. 5A) has a tetrahedral beryllium bridging O3B of ADP to O1B of UDP. An essential Mg$^{2+}$ coordinates one fluorine, and O1B of ADP. The two diastereotopic fluorines show well-separated resonances in the $^{19}$F NMR (Fig. 2B). This stable mimic of Ap$_3$U is strongly coordinated to 4 arginines and 1 lysine and thus endorses the observation that nucleotide kinases are more strongly inhibited by Ap$_3$Nuc than by Ap$_3$Nuc on account of their additional negative charge.[9]
2.5 Conclusions

The significant ability of beryllium (II) fluorides to complete tetrahedral coordination by binding to an anionic oxygen makes them effective isosteric and electrostatic analogs of phosphate in a wide range of situations.[11] The bond lengths for Be–F and Be–O are close to those for P–O (1.6 ± 0.5 Å) and the dominant ionic character of the Be–F bond means that the fluorines readily accept H-bonds from a range of donors and/or coordinate to Group 2 metal ions. These mimics have been advantageously used to study changes in major conformation of proteins by crystallography, 1H NMR, and EM while studies on ADP·BeF3 have supported investigations of ATPases that drive various mechanical processes at a molecular level, particularly for myosin.[12] The bond lengths for Be–F and Be–O depending on the concentration of fluoride.[14] Their stability is a function of pH because aluminum forms insoluble Al(OH)3 above pH 7.5.[15] Aluminum and fluoride were discovered to stimulate the activity of small G proteins and the proposal that they could mimic the active GTP bound state was endorsed by 19F NMR analysis, which identified the formation of a GDP·AlF complex for Gαi.[17]

In 1994, crystal structures for tetrahedral GDP·AlF complexes of transducin α and a small G protein, Gαi2, appeared almost simultaneously, and were soon followed by an ADP·AlF complex for a myosin fragment.[18, 19] Since then, the number of such AlF3 complexes in the PDB (PDB ligand: ALF) determined by crystallography has grown steadily to reach 109 by March 2016 (Fig. 1, SI Table 4).

3. Octahedral MF3 Complexes

Aluminum (III) forms stable fluorides in water that exist as a mixture of octahedral species including AlF3·4H2O, AlF3·3H2O, AlF3·2H2O, and AlF3·H2O depending on the concentration of fluoride.[14] Their stability is a function of pH because aluminum forms insoluble Al(OH)3 above pH 7.5.[15] Aluminum and fluoride were discovered to stimulate the activity of small G proteins in the presence of GDP.[15] The PDB has 14 structures with a tetrafluoroaluminate bonded to an aspartyl oxygen. This mimics an aspartyl phosphate, known to be a transient species in the catalytic activity of these enzymes. They have an essential Mg2+ enclosed in a 6-membered ring, as seen for the corresponding BeF2 structure (Section 2.2), and all align very well on PDB: 2wt7 (Fig. 7, SI Table 6), showing commonality of remaining 4 ligands coordination the catalytic Mg2+.

These structures fall into two subsets: six members of the first group have a second aspartate next-but-one to the first, and it coordinates the oxygen that is the sixth aluminum ligand. The O–Al–O bonds are “in-line” (167.5 ± 7.0˚) with the aluminum midway between the two oxygens (separation 3.94 ± 0.11 Å). The Al–F bonds are 1.78 ± 0.02 Å (for the 6 best-resolved structures), independent of coordination to Mg. βPGM accounts for three of the six structures, the other three being a human mitochondrial deoxyribonucleotidase, a phosphoserine phosphatase (PSP), and a C-terminal domain phosphatase that operates on RNA polymerase II. In all of these, a catalytic aspartate accepts a short H-bond from the apical water/hydroxyl group (2.59 ± 0.05 Å) to complete the orientation of this oxygen for nucleophilic attack on the aspartyl phosphate.[18]

The second subset comprises ATPases involved in pumping Ca, Cu, and Zn. They use an aspartyl phosphate intermediate, whose TS for hydrolysis is mimicked by the octahedral AlF3. These have “in-line” O–Al–O bonds (163.8 ± 8.1˚) with aluminum midway between the two oxygens (O–O separation 3.92 ± 0.14 Å) and Al–F bonds 1.78 ± 0.02 Å. An axial water oxygen forms short H-bonds to an invariant glutamate (2.54 ± 0.1 Å) and a threonine carbonyl (2.57 ± 0.05 Å). These residues clearly orientate and

![Figure 5](image-url) Figure. 5 (A) Structure of BeF3 complexed to 2 nucleotides in U/C kinase. Beryllium (olive sphere) is bonded to oxygens of ADP (green) and UDP (purple) with one fluoride (light blue) coordinating an octahedral Mg2+ (green sphere). The tetrahedral complex is coordinated by 5 H-bonds to 4 amino acids. (gray sticks). (B) 19F NMR for the ADP·BeF3·UDP complex as above.

![Figure 6](image-url) Figure. 6. Structures of 14 aspartyl tetrafluoroaluminates superposed by Cx alignment. Aluminum is octahedrally coordinated to Asp-O4 (gray) forming a 6-membered ring with a catalytic magnesium and “in-line” with the acceptor oxygen, water (red sphere) or the hydroxyl group of a nucleoside or hexose reactant (colors). (Atom colors: fluorine, light blue; aluminum, gray; nitrogen, blue; oxygen, red; magnesium, green).
polarize the water for “in-line” attack on the aspartyl phosphate (Section 8.3).[20]

3.1.2. Nucleotide tetrafluoroaluminates, GDP

There are 46 x-ray structures of AlF<sub>4</sub> complexes with GDP that constitute isoelectronic but non-isosteric mimics of GTP in small G proteins, dynamins, ribosomal factors, kinases, ATPases, mutases, ion pumps, and helicases. Of these structures, 25 are resolved at ≤ 2.7 Å and align remarkably well (Fig. 7, SI Table 5). The aluminum is bonded to GDP by O3B and the catalytic Mg<sup>2+</sup> is coordinated to F1 and O1B in a 6-membered ring. There is remarkable consistency in neighboring amino acids, notably by a heptapeptide near the N-terminus, sequence XXXXGKS(T), whose serine hydroxyl coordinates magnesium trans to a fluorine. The guanosine base and ribose occupy a common conformation (Fig. 7) with the exception of Atlastin (PDB: 4ido). The geometry of the AlF<sub>4</sub> moiety is well defined, being regularly octahedral to 2.7 Å resolution, with an “in-line” O–Al–O angle 172.8° ± 7.1°, with aluminum midway between the axial oxygens that are 4.07 ± 0.23 Å apart (Table 1), and have AlF bonds 1.77 ± 0.28 Å. All the structures have an axial oxygen ligand (Fig. 7, red spheres) to aluminum that is trigonal planar with respect to two H-bond acceptors (ψ-dihedral 4.9° ± 2.9°) whose angle to the axial oxygen is 102 ± 6°(Fig. 8). One is the backbone carbonyl of a threonine, whose OG coordinates the magnesium (Fig. 7, upper right). The second is a glutamine side-chain carbonyl or a water (Fig. 7, lower right, red spheres).

Figure. 7 GDP tetrafluoroaluminate structures. 25 Structures are superposed on (PDB: 2gj8) by α-carbon atoms (primarily for the invariant heptapeptide, bottom to top center). AlF<sub>4</sub> is locked in a 6-membered ring (center) with catalytic Mg<sup>2+</sup> coordinating F1 and O3B. Octahedral coordination to Mg<sup>2+</sup> is provided by OB1, 2 trans-waters, a Thr hydroxyl (top right), and a Ser/Thr hydroxyl (top center). Phosphate oxygen coordination to a Lys18 (center) is standard. Location of guanines is regular (left rear) with two exceptions. Atom colors: fluorine, light blue; aluminum, gray; nitrogen, blue; oxygen, red; magnesium, green).

3.1.3. Nucleotide tetrafluoroaluminates, ADP

The 45 octahedral structures that have AlF<sub>4</sub> bonded to a terminal oxygen of ADP (O3B) include kinases, hydrolases, isomerases, myosins, helicases, transporter pumps, and nitrogenases. They mimic ATP and are relatively diverse in conformation. The 25 that are resolved at ≤ 2.5 Å have an axial O–Al–O distance of 4.04 ± 0.14 Å with an “in-line” angle of 170° ± 8°. The majority of the 45 have a water as the second oxygen ligand with the catalytic Mg<sup>2+</sup> also coordinated to one β-oxygen and a fluorine. This is illustrated for F1ATPase (PDB: 1h8e) (Fig. 9A). Three complexes have magnesium triply coordinated to OA, OB and F.

Overall, the aluminum is a little closer to O3B (1.97 ± 0.12 Å) than to the second oxygen (2.10 ± 0.12 Å), and Al–F bond lengths (for the 12 best-resolved structures) are 1.77 ± 0.04 Å. The variable general position of the fluorines relative to the catalytic Mg<sup>2+</sup> suggests that some compromise has been reached in fitting four fluorines into protein loci that have evolved to accommodate three electronegative oxygens.

3.1.4. Other tetrafluoroaluminates

Two structures have AlF<sub>4</sub> bonded to a histidine nitrogen, as illustrated for phosphoglycerate mutase (PDB: 2f90). This mimics phosphoryl transfer from His11 to OH-2 of 3PGA (Fig. 9B).
3.2 Trifluoroaluminates, AlF₃⁻

There are three examples of octahedral complexes where an aluminum trifluoride core is expanded to octahedral, six-coordination by having three oxygen ligands (SI Table 7). For the small G protein Rab5a, the mutation A30P results in the addition of the side chain hydroxyl of Ser29 to aluminum. For hPGK, the mutation K219A results in the addition of water to the aluminum. For a bacterial dUTPase, aluminum trifluoride takes the place of the β-phosphoryl group in dUTP and coordination to O3A, O3B, and to the water nucleophile completed the octahedral array (Fig.10). This structure provides a unique example where nucleophilic attack is not directed at a terminal phosphorus.¹²¹

4. Trigonal Bipyramidal MF₃.

4.1. Trifluoromagnesates MgF₃⁻

Magnesium does not form multiple stable fluorides in water. They are moderately soluble (2 mM) with a dissociation constant for MgF₂0₃⁻ estimated at 10⁻⁶ M.¹²² However, trifluoromagnesate protein complexes were first anticipated on the basis of magnesium-dependent fluoride inhibition studies that in 2002 led to the first identification of MgF₃⁻ in a tbp crystalline TSA complex for the small G protein RhoA/RhoGAP (Fig. 11A).¹²³ The PDB now lists 16 entries for this ligand (code MG) while a further 3 entries assigned as AlF₃⁻ have been shown by ¹⁹F NMR to be MgF₃⁻ complexes (SI Table 8).¹²⁴ Magnesium is regularly 6-coordinate and gives octahedral complexes with oxygen ligands. By contrast, trifluoromagnesate is 5-coordinate, trigonal bipyramidal in proteins, and has ideal characteristics to mimic the phosphoryl group as it is isoelectronic with PO₃⁻ and has similar geometry. Examples of its use include small and large molecule kinases, mutases, phosphatases, and hydrolases. Their complexes invariably involve coordination to one catalytic Mg²⁺ (two for some protein kinases), which are usually in a cyclic 6-membered ring structure, as shown for aspartyl phosphate mimics (Fig.11B). They have an axial O-Mg-O distance of 4.19 ± 0.08 Å with an in-line angle 171.4˚ ± 3.9˚. The axial Mg–O bonds are 2.13 ± 0.10 Å with Mg-F bonds of 1.83 ± 0.06 Å, compared to computed non-bridging P–O bonds of 1.52 ± 0.02 Å.¹²⁵

4.2. Aluminum Trifluoride, AlF₃⁻

The first example of an aluminum trifluoride complex was presented in 1997 for the tbp complex in the active site of a dinucleotide kinase (PDB: 1kdn), shortly to be followed by a study on Ras/RasGAP with a GDP complex.¹²⁶ There are now 56 examples of structures that report an aluminum trifluoride core. Of these, three are octahedral (Section 3.2), and three have been shown by ¹⁹F NMR to be trifluoromagnesates (see Sections 4.1 and 7.2). Of the remainder, only two alkaline phosphatase structures may be identified as having a tbp aluminum trifluoride core (Fig. 12). In mutant P300A (PDB: 1kh5) two catalytic Zn²⁺ ions share one fluoride while Ser102 and a zinc-coordinated water provide the axial ligands for the tbp aluminum. It has an apical O–Al–O distance of 3.80 Å and Al-F bonds of 1.75 Å characteristic
of the AlF₃⁻ complexes described above (Section 3.1, SI Table 7). What is the situation for the remaining 48 AlF₃₀ complexes?

Figure. 12 Upper: Structure of the catalytic center for alkaline phosphatase complexed to AlF₃ (PDB: 1kh5). Lower: Cartoon of the coordination organization in the active site with transferring phosphoryl group (green) and nucleophilic water (red).

The influence of pH on the transition between octahedral and tbp structures of AlF₃ complexes in protein crystal structures for phosphoryl transfer enzymes was proposed to involve a switch from AlF₃⁻ to AlF₃₀ at elevated pH. However, studies on the pH dependence of aluminum hydroxide solubility supported an alternative interpretation. Aluminum hydroxide precipitates at pH ≥ 8, hence it will be replaced by magnesium in the protein complexes, with a consequent change to tbp geometry. That conclusion has now been validated by pH-dependent F NMR analysis or crystallization in an aluminum-free medium will justify reassignment of some, or many, of these complexes as trifluoromagnesates (SI Table 9). In structural terms, the dimensions of the tbp complexes closely reflect those of known trifluoromagnesates: axial O-M-O bonds 4.29 ± 0.39 Å, and M-F bonds 1.75 ± 0.12 Å. It is therefore likely that ¹⁹F NMR analysis or crystallization in an aluminum-free medium will justify reassignment of some, or many, of these complexes as trifluoromagnesates (SI Table 9).

Figure. 13 Structure of hPPIP5K2 (PDB: 2q9p) to show the “MgFx” cluster adjacent to phosphates 4 and 5 of Ins6P.

5. ¹⁹F NMR Studies on MFₓ

The inclusion of metal fluoride moieties within protein complexes has opened up the opportunity to use ¹⁹F NMR measurements to examine the environment in which phosphate groups reside within the protein. The ¹⁹F isotope has 100% natural abundance and a very high gyromagnetic ratio (25.18 x 10⁻¹ T⁻¹s⁻¹), leading to very high sensitivity NMR spectra. Hence, metal fluoride species can be detected at low protein concentrations, and in large molecular weight complexes.

5.1 Chemical shifts

The chemical shifts of ¹⁹F resonances provide a key measure of interactions between MFₓ moieties and their protein hosts. They are reliable reporters of the electronic environment in the vicinity
of the fluorine nuclei and, by extrapolation and calculation, they act as reporters of the electronic environment experienced by the phosphoranyl oxygen atoms in the TS.\textsuperscript{[20, 36]} \(^{19}\text{F}\) resonances display a high degree of dispersion and are predictable with good precision from quantum calculations of electronic distribution.\textsuperscript{[37]}

The average chemical shifts of resonances from AlF<sub>3</sub>, MgF<sub>2</sub>, and BeF<sub>2</sub> species differ (-154, -156, and -169 ppm, respectively), but a wide spread of individual shifts is observed in complexes with proteins. In similar βPGM complexes, for example, the average chemical shifts are -138 (AlF<sub>3</sub>), -153 (MgF<sub>2</sub>) and -160 (BeF<sub>2</sub>) ppm.\textsuperscript{[13, 24c]} This distribution is strongly affected by the vicinity of H-bond donors, as shown clearly in a comparison of the G6P and the 2-deoxyG6P complexes of βPGM. In the βPGM-MgF<sub>2</sub>-2deoxyG6P TSA complex one fluorine loses its H-bond partner and its resonance moves substantially upfield (-18.1 ppm). (NB: \(^{19}\text{F}\) chemical shifts are quoted relative to trifluoroacetic acid as reference)

The high sensitivity of \(^{19}\text{F}\) chemical shifts to the surrounding environment can be used to show how enzymes control the influence of changes of protonation state. Thus, for βPGM it was observed that \(^{19}\text{F}\) chemical shifts are invariant over the pH range 6.5 – 9.5, indicating that any changes in protonation state of the protein has no detectable influence on the environment of the TS complex. Characteristic average chemical shift values for different MF<sub>x</sub> species have been identified that millimolar fluoride is sufficiently effective at leaching Al<sup>3+</sup> from glass, including borosilicate glass, to transform MgF<sub>2</sub>-complexes into AlF<sub>3</sub>-complexes unless an aluminum chelator such as deferoxamine is present.

5.2 Chemical exchange

It is observed, particularly in the AlF<sub>3</sub>-complexes of some enzymes (including many early NMR studies of these complexes), that individual \(^{19}\text{F}\) resonances coalesce to a single resonance as a result of rapid chemical exchange.\textsuperscript{[23b, 38]} Resolved resonances of similar complexes have chemical shift differences of up to 10 kHz, and this shows that in some AlF<sub>3</sub>-complexes the interchange of fluorines greatly exceeds this rate. All MgF<sub>2</sub>-complexes of wild-type enzymes reported to date have resolved \(^{19}\text{F}\) resonances, and hence much slower rates of fluorine interchange. For BeF<sub>2</sub>-complexes, the spectra show evidence of faster exchange rates than for MgF<sub>2</sub>-complexes.\textsuperscript{[13]}

5.3 NOEs

Proton distribution in the vicinity of fluorine nuclei in the MF<sub>x</sub> moiety can be assessed through the quantitation of \(^{19}\text{F}--\text{H}\) NOEs. This approach has been used to determine solution structures of βPGM-MgF<sub>2</sub>-G6P TSA and βPGM-AlF<sub>3</sub>-G6P TSA complexes and so resolve a controversy concerning a reported pentaoxo phosphorane for this enzyme (Section 7.1).\textsuperscript{[24a, 24c]}

Traditionally \(^{19}\text{F}--\text{H}\) NOEs are difficult to quantify owing to the effects of spin diffusion between \(^{1}\text{H}\) nuclei as the \(^{19}\text{F}--\text{H}\) NOE builds but, for MF<sub>x</sub> complexes, the primary NOEs are to exchangeable protons. Hence \(^{1}\text{H}--\text{H}\) spin diffusion can be suppressed by using a perdeuterated enzyme in a protonated buffer. Resonance assignment of the exchangeable \(^{1}\text{H}\) nuclei in the protein allows unambiguous assignment of individual \(^{19}\text{F}\) resonances.

5.4 SIIS – solvent induced isotope shifts for \(^{19}\text{F}\) NMR

Proton distributions in the vicinity of fluorine nuclei can be assessed independently of \(^{19}\text{F}--\text{H}\) NOEs on the basis of solvent induced hydrogen/deuterium primary isotope shifts (SIIS) of the \(^{19}\text{F}\) resonances. For H-bonds to fluorine present between MF<sub>x</sub> moieties and proteins, F---H---N and F---H---O, the magnitudes of the isotope shifts reflect local proton densities because of the through-space transmission of electric field differences between N---H and N---D bonds.\textsuperscript{[39]} For example, in the βPGM-MgF<sub>2</sub>-G6P TSA complex (Fig. 14), the three fluorines are coordinated by three protons (in a distorted tetrahedral arrangement), two protons (in a trigonal arrangement) and one proton, giving sum SIIS values of 1.6 ppm, 1.4 ppm, and 0.9 ppm, respectively. Comparing the G6P and the 2-deoxyG6P TSA complexes of βPGM, the sum SIIS value of one fluoride ion for the latter complex falls to close to zero (0.2 ppm), indicating that loss of the hexose 2-OH group leaves this fluorine virtually devoid of H-bonds.\textsuperscript{[38]} The consequence of the removal of this hydroxyl group on the whole TSA complex is also observable as the other two fluorines move closer to their hydrogen bonding partners, as shown by small increases in their sum SIIS values (to 1.7 ppm and 1.5 ppm).

![Figure 14 1D \(^{19}\text{F}\) NMR spectra of \(\beta\)PGM complexes with BeF<sub>2</sub> (upper), MgF<sub>2</sub> and G6P (middle), and AlF<sub>3</sub> and G6P (lower). The \(^{19}\text{F}\) resonance at -119 ppm in each spectrum is from free F<sup>−</sup> ions, while those between -160 and -170 ppm (upper spectrum) are from unbound BeF<sub>2</sub> species and those between -150 and -160 ppm (lower spectrum) are from unbound AlF<sub>3</sub> species. The middle spectrum contains 3 small peaks from a second MgF<sub>2</sub>-bound protein conformation.](Image 310x273 to 548x522)

5.5 Scalar couplings across hydrogen bonds

Details of the coordination of the MF<sub>x</sub> moiety by the protein is further shown in scalar couplings between nuclei involved with N---H---F H-bonds. \(^1\text{H}J\) and \(^2\text{H}J\) couplings have been reported for individual H---F pairs, with values up to 59 and 36 Hz, respectively.\textsuperscript{[26]} The magnitudes of both scalar couplings correlate closely with distances measured from crystal structure analysis. Hence, as well as reporting on the interaction across
individual hydrogen bonds, scalar couplings provide an
independent means of assigning $^{19}$F resonances, and cross-
validating solution and crystal behavior.

5.6 Conclusions

NMR measurements of $^{19}$F nuclei in the active site of MF$_x$ TSA
complexes provide a picture of the relationship between charge
distribution of the mimic for phosphoryl group transfer and the
enzyme. The good relationship between $^{19}$F chemical shifts and
SISS values illustrates the dominant influence that the very local
H-bonding groups have on shaping charge density on MF$_x$ moieties. Moreover, the strong correlation between the observed
NMR parameters and coordinates determined for numerous
proteins in the crystalline state is a vital link showing that atomic
positions determined at high resolution in the solid phase very
closely reflect solution behavior.

6. Computational Analyses of MF$_x$ Complexes

There have been relatively few computational studies of MF$_x$
complexes within protein binding sites as such. Instead, these
GSA and TSA structures have been widely used as starting points
for a very large number of calculations by switching the MF$_x$ atoms
into PO3 while retaining the same geometry. Such computation
has been applied to the molecular mechanisms of a variety of
enzymes catalyzing phosphoryl transfer reactions, particularly
the small GTPases, which play critical roles in cell signaling and
regulation, and to cAPK. Theoretical methods provide
considerable insights into the distribution of electrons within
molecules, and the energies of protein/ligand interactions that
mediate binding and TS stabilization. In some cases, accurate
structures can be obtained and used to resolve the nature of MF$_x$
species in X-ray crystal structures of relatively low resolution.
Such calculations, briefly described below, yield useful
information on the extent to which MF$_x$ moieties resemble ground
states or TSs in enzyme-catalyzed phosphoryl transfer.

6.1 Computational Methods.

The principal approach to obtaining the properties of MF$_x$
complexes has been the use of Density Functional Theory (DFT),
given the ability of this method to yield accurate structural
properties. Numerous reviews are available that detail the
theoretical principles underlying DFT together with its limitations,
which include problems in modeling dispersion interactions and
activation energy barriers in chemical reactions. One important
advantage of DFT, is that molecular systems composed of
relatively large numbers of atoms can be treated completely
quantum mechanically, allowing considerable insight into the
electrostatic properties of MF$_x$ complexes and how these might be
perturbed by being in a protein environment. The general strategy
has been to build active site models composed of the MF$_x$
complex and residues that interact directly with the complex and
surrounding molecules, such as ATP and GDP. Larger models
can also be built that include "second shell" residues, which form
H-bonds to the initial set of inner residues. An alternative
approach, which avoids the need to place artificial coordinate
restraints on atoms in the QM region, the complete system is
modeled using QM/MM methodologies. Here the QM region is
embedded in the rest of the protein and solvent, with the
additional atoms (in an MM region) being described by classical
potential energy functions that depend on "force field" parameters.
Various methods can then be used to "couple" the QM and MM
regions. The advantage of the QM/MM approach, which also
permits the inclusion of electrostatic effects arising from the
protein and solvent environment, lies in the elimination of "edge
effects" at the boundaries of the QM region arising from
coordinate restraints. In addition, the relative simplicity used to describe the MM region allow the use of MD simulations
to obtain free energy estimates for the system, which are not
reliably obtained by analysis of the geometry-optimized QM
active site models.

6.2 BeF$_3$ complexes

As discussed in Section 2, beryllium fluoride complexes resemble
GS phosphate groups when bound to nucleophilic groups or
dinucleotides. The extent to which such tetrahedral complexes
mimic phosphate moieties was explored using QM calculations of
BeF$_3$-complexed to the catalytically important aspartate side
chain of βPGM in the presence and absence of βG6P, a substrate
for the enzyme. Large models, consisting of the BeF$_3$ complex
and 29 residues surrounding the active site, were obtained from
crystal structures of these complexes and structurally optimized
using B3LYP and 6-31G basis set, with the inclusion of d
polarization functions for the fluoride ions. As usual, the outer
atoms in these models were constrained to their crystallographic
coordinates. Atomic charges were then computed using the
 Mulliken formulation in order to minimize computational expense.
The results showed that the beryllium and fluoride ions carry
about 60% and 75% of the charges expected for phosphorus and
oxygen atoms in a phosphate group. Hence, although the total
charge of the BeF$_3$-moiety is identical to that of the reactive
intermediate in the enzyme-catalyzed reaction, the internal
separation of charge is scaled down.

6.3 MgF$_3$ complexes.

There is ample evidence that the MgF$_3$-ion is an excellent stable
analog of the TS for phosphate transfer in a number of enzyme-
catalyzed reactions (Section 4). Early DFT calculations were
performed to investigate the claim that X-ray crystallography had
revealed the structure of a phosphorane intermediate in the
reaction catalyzed by βPGM, and validated the correction that the
tbp complex was MgF$_3$- (Section 7.1). The calculated distances
for a MgF$_3$-anion were consistent with those seen in the crystal
structure. Subsequent high-level QM/MM calculations have
supported this conclusion, and have shown that it also holds for
phosphoryl transfer catalyzed by UTPase. QM/MM studies followed
that sought to demonstrate that MgF$_3$- was present in
medium-resolution X-ray crystal structures of the Ras/RasGAP
complex rather than the isoelectronic AlF$_3$. The QM region was
modeled using standard Hartree-Fock ab initio calculations, which
generate the effects of electronic correlation. Nonetheless, this
level of QM theory was sufficient to show that calculated distances
and angles for the MgF$_3$-complex were in much better agreement
with the crystal structure for the Ras/RasGAP-GDP-MF$_x$ complex
than those computed for either AlF$_3$ or AlF$_5$. This was an important
result because the electron density observed for the MF$_x$ species
in the Ras/RasGAP-GDP-MF$_x$ structure (PDB: 1wq1) was

\[ 1wq1 \]
inadequate to permit an unambiguous assignment of the ion. More recent work has sought to establish the extent to which MgF$_3$ resembles the phosphoryl moiety undergoing transfer in the TS for GTP hydrolysis catalyzed by the RhoA-RhoGAP-GDP complex. Specifically, this study, which employed DFT calculations on a very large active site model, demonstrated that the observed $^{19}$F chemical shifts for the RhoA-RhoGAP-GDP-MgF$_3$: complex can indeed be interpreted as indirect measures of the relative electron densities of the cognate oxygen atoms in the "true" TS for attack of water on the terminal phosphate of GTP.

6.4 AlF$_3$ complexes.

Notwithstanding the questions raised about the validity of designating many tbp MF$_x$ complexes as AlF$_3$, their structures, notably for Ras and for cAPK, have been used as starting points for many computations. The success of these computations lies in the simplicity of the transformation of AlF$_3$ into PO$_4^-$ without regard to the change in charge involved. Only the tbp geometry matters.

6.5 Conclusions

Taken overall, the number of computational studies on the electronic structure and steric properties of protein-bound MF$_x$ complexes has been relatively limited. There has also been limited evaluation of their resemblance to TS structures calculated using either QM or QM/MM methods for a range of enzymes, and their dynamic behavior within the active site remains poorly explored. This is surprising given the clear differences in the $^{19}$F NMR spectra reported for BeF$_3^-$, MgF$_3^-$ and AlF$_3^-$ containing complexes (Section 5).

MF$_x$ complexes have necessarily provided valuable starting points for numerous QM and QM/MM studies of the mechanism(s) of phosphoryl transfer, with particular focus on the Ras-RasGAP-GDP-MF$_x$ structure (1wq1) as a basis for efforts to model the structure and energetics of the TS for Ras-catalyzed GTP hydrolysis. This choice has not, however, led to a consensus view of the mechanism. For example, extensive QM/MM calculations by some groups consistently predict a MF$_x^-$ structure for the TS for GTP hydrolysis catalyzed by the RhoA-RhoGAP complex.

7. Sorting the Sheep and the Goats

Studies on MF$_x$ transcend the boundary between protein crystallography and biomolecular chemistry. As a result, many situations exist which can benefit from closer integration of the available experimental and computational approaches. Several examples have been identified where electron density data can be reassigned by a broader approach to its interpretation, while this review has identified new examples capable of reinterpretation. This is notably where the electron density maps are insufficiently resolved to make their interpretation unambiguous in the absence of a chemical evaluation. We briefly highlight two cases that are fully documented and one that is susceptible of reinterpretation.

7.1. MgF$_3^-$ misidentified as PO$_4^-$

The 2003 publication of a tbp complex in the active site of JPGM as a pentaoxyphosphorane received immediate attention, and re-examination (2004, 2008, 2007). A combination of computation (Section 6.3) and $^{19}$F NMR analysis (Section 5) established that it is accurately interpreted as a trifluoromagnesate complex (Fig. 15A) (2008, 2006). A later in-depth QM/MM analysis calculated both the reaction path of the phosphorylation step (using PO$_4^-$) and also the geometry of a complex with the MgF$_3^-$ TSA. It concluded that trifluoromagnesate is a good mimic of the true TS, which has pentacoordinate phosphorane character.

7.2. MgF$_5^-$ misidentified as AlF$_3^-$

An authoritative and extensive study on protein kinase A (cAPK) included the description of a tbp complex for the phosphorylation of a target serine peptide by ATP. $^{19}$F NMR established the major presence of MgF$_5^-$ in the complex along with some octahedral AlF$_3^-$, showing that charge balance predominates over geometry in selection of the TS analog (Section 4.2, Fig. 13B). A later in-depth QM/MM analysis calculated both the reaction path of the phosphorylation step (using PO$_4^-$) and also the geometry of a complex with the MgF$_5^-$ TSA. It concluded that trifluoromagnesate is a good mimic of the true TS, which has pentacoordinate phosphorane character.

7.3. MgF$_5^-$ misidentified as MgF$_4^-$

We have explained that it is exceptional to find magnesium in the form of tetrahedral tetrafluoromagnesate, MgF$_4^-$ (Section 4.2). Of 28 examples of this tetrahedral ligand listed in the PDB, the best resolved (2.40 Å, PDB: 2xxe) is for an ATPase ion from shark. Its electron density map does not unambiguously support interpretation as a magnesium-coordinated tetrahedral MgF$_4^-$.

We have therefore re-refined the data to show an alternative interpretation of a tetrahedral ligand covalently bonded to the essential Asp376 (Fig. 15C). This has an axial O-Mg-O distance of 3.85 Å, an in-line angle of 171.3°, and Mg-F bonds 1.86 Å. It is likely that this analysis could be applied to some or all of these tetrahedral complexes, although electron density is not available for the majority of them.
Figure. 15 The electron densities for the original interpretation and re-interpretation based on the unbiased omit map $F_o - F_c$. (A) The original map for pentaoxyphosphorane in PDB: 1o08 (left) and the unbiased omit map for MgF$_3$- in PDB: 2wf5 (right). (B) The original map for AlF$_30$ in PDB: 1l3r (left) and the unbiased omit map for the reinterpretation mixed occupancy for MgF$_3$-/AlF$_4$- at 70/30 ratio. (C) The original map for MgF$_{42}^-$ in PDB: 2zxe (left) and the omit map (right) for an alternative MgF$_3^-$ and water in the same density. All the unbiased $F_o - F_c$ omit maps in magenta are contoured at 3$\sigma$ for the metal fluoride moiety before their inclusion in the model and the 2$F_o - F_c$ maps in blue are contoured at 1$\sigma$.

8. Fundamentals of Phosphoryl Transfer revealed by MF$_x$

8.1 Protein conformation – H-bonded and aligned NACs

The accessibility of high-resolution structures and solution NMR measurements for multiple MF$_x$ complexes allows a detailed picture to be developed of many of the steps involved in catalysis. $\beta$PGM is a very good example where data are available for the apo-enzyme, the BeF$_3^-$ mimic of the phosphoenzyme (EP), the BeF$_3^-$ mimic of the EP complexes with both substrates (G6P and $\beta$G1P), and the corresponding MgF$_3^-$ and AlF$_4^-$ TSA complexes for each reaction. From them the development of the TS complex can be mapped out (Fig. 16). These data reveal how the EP down-regulates hydrolysis by disfavoring water from occupancy of a position to attack the phosphate. The EP undergoes domain closure in the presence of substrate but to alternative NACs.$^{[13]}$ The first is a more stable complex where the substrate H-bonds with the target phosphate, and which interconverts with a second, less stable complex where the substrate is aligned for attack. The latter NAC develops into the TS. This mutase operates on each of its two substrates in two consecutive reactions. A comparison of its behavior with the two substrates reveals that the protein conformation is conserved in the TSs of the two chemical steps, and the enzyme responds to the step change in substrate geometry by utilizing water molecules as spacers in one reaction, and leaving the transferring phosphate group depleted in H-bond partners in the other.$^{[35b]}$

Figure. 16 Progression of $\beta$PGM active site from ground state (top, magenta) to transition state (bottom) (rainbow coloring shows pairwise progression). Left track: Step 1 pathway via phosphoenzyme (EP) (orange) to NAC1 (yellow) to TS1 for phosphorylation of glucose $\beta$-1-phosphate (green). Right track: Step 2 pathway via NAC1 (grape) to NAC2 (cyan) to TS2 (deep blue). Domain closure (EP to NAC) is linked to conformational adjustment of catalytic Asp10 to provide GABC for the glucose-OH group.

8.2 Charge Balance – Neutralize the "Anionic Shield"

The concept of charge balance was provoked by the observation that Ap$_5$A (5–ve charges) is a better inhibitor of adenylate kinase than is Ap$_4$A (4–ve charges).$^{[10a]}$ The true TS (6–ve charges) is thus better mimicked by Ap$_5$A. The concept says that enzymes complement the excess anionic charge on TSs for phosphoryl transfer by cationic Mg$^{2+}$ and side-chain residues in the immediate vicinity of the transferring phosphorus atom. Studies on hPGK validated this concept by demonstrating that hPGK prioritizes anionic charge over geometry in selection of MF$_x$ for TSA complex
formation. Based on the geometry of MF₄ complexes for a wide range of phosphoryl transfer enzymes, it was demonstrated that charge balance is maintained within a sphere of up to 15 Å around the transferring phosphorus even when that borders on bulk water (Fig. 17). A classic example is that of cAPK where charge balance is only achieved by the incursion of the substrate peptide with three +ve charges into a 13.5 Å sphere. This concept has been endorsed in a DFT study on cAPK, that found the order of affinity to the enzyme is MgF₃ > AlF₄ > AlF₃ while it confirmed charge balance out to 8 Å from the reaction center.

The overwhelming majority of x-ray structures in this review on MF₄ as a TSA for phosphoryl transfer show that there are commonly only two situations where water comes within 4 Å of the transferring phosphorus. First, this is where the water is an isolated nucleophile for the hydrolysis of ATP, GTP or an aspartyl phosphate. Secondly, it is where water features as a ligand for coordination to a catalytic Mg²⁺ that is itself coordinated to the transferring phosphorus group. Thus, for 10 well-resolved ADP-AlF₄⁻ complex structures, the average distance to the next nearest non-specific water is 4.3 ± 0.7 Å. It is also evident that water is more clearly excluded from the catalytic center in MF₄ structures of TSA complexes than in the corresponding NAC structures. Thus for 12 small G proteins the next nearest water is 6.6 ± 0.2 Å for GDP-AlF₄⁻ TAs but 4.22 ± 0.1Å for NACs. The unavoidable conclusion is that model studies on the hydrolysis of ATP, GTP in water cannot reliably be extrapolated to enzyme catalysis. A prime reason for exclusion of water is the control of H-bonding to neutral OH nucleophiles. Without exception all of these show proximity to a H-bond acceptor, often an aspartate carboxylate. While this interaction has historically been interpreted as providing GABC, recent computational analyses have shown that proton transfer occurs.

**8.4 Desolvation – Activate the nucleophile and the electrophile**

The importance for catalysis of the exclusion of water from the active site of phosphoryl transfer enzymes historically has proponents and opponents. The overwhelming majority of x-ray structures in this review on MF₄ as a TSA for phosphoryl transfer show that there are commonly only two situations where water comes within 4 Å of the transferring phosphorus. First, this is where the water is an isolated nucleophile for the hydrolysis of ATP, GTP or an aspartyl phosphate. Secondly, it is where water features as a ligand for coordination to a catalytic Mg²⁺ that is itself coordinated to the transferring phosphorus group. Thus, for 10 well-resolved ADP-AlF₄⁻ complex structures, the average distance to the next nearest non-specific water is 4.3 ± 0.7 Å. It is also evident that water is more clearly excluded from the catalytic center in MF₄ structures of TSA complexes than in the corresponding NAC structures. Thus for 12 small G proteins the next nearest water is 6.6 ± 0.2 Å for GDP-AlF₄⁻ TAs but 4.22 ± 0.1Å for NACs. The unavoidable conclusion is that model studies on the hydrolysis of ATP, GTP in water cannot reliably be extrapolated to enzyme catalysis. A prime reason for exclusion of water is the control of H-bonding to neutral OH nucleophiles. Without exception all of these show proximity to a H-bond acceptor, often an aspartate carboxylate. While this interaction has historically been interpreted as providing GABC, recent computational analyses have shown that proton transfer occurs.

---

**Figure 17** Upper: Charge balance for kinases cAPK and CDK2 showing distortion for “AlF₄⁻” assignment. Lower: Charge balance for a range of phosphoryl transfer proteins with insert showing radial nature of charge balance calculation.

**Figure 18 Upper**: Structures aligned (backbone Cα) for phosphoryl transfer by human hPPIP5K2. Reactants (red), transition state (yellow) and product (green) complexes show “in-line” transfer of phosphoryl group from ADP (right) to Ins6P (left) with near superposition of the three equatorial oxygens of the tbp in side and orthogonal front view. **Lower**: Approach of reactants by 0.4 Å places three equatorial oxygens in TS locations enabling phosphorus to move 1.2 Å through the core of the tbp complex to effect phosphoryl transfer.
late in the TS, as especially demonstrated for the small G protein, RhoA (Section 8.5).\textsuperscript{20, 50b, 64}

Such H-bonding activity is equally evident in PGM, PSP, and phosphoglycerate mutase. Its purpose is primarily to orientate the oxygen for nucleophilic attack by enabling orbital overlap and denying H-bonding from the OH group to the anionic oxygens of the electrophilic phosphoryl group. Strong support is provided from a study on RNase A in which His12 and His119 were independently replaced by 4-fluorohistidine, pKa 3.5. The artificial mutants delivered an unchanged kcat but with greatly modified pH profiles.\textsuperscript{365} This result is consistent with these histidines delivering H-bonding for nucleophile orientation rather than for GABC. While our interpretation is founded on MF\textsubscript{x} complexes in proteins, it may well also apply to enzyme mechanisms for C-O cleavage, typically glycosidases, which have hitherto been designated as using GABC.

8.5 GTP hydrolysis depends on controlling H-bonds

Small G proteins accelerate the hydrolysis of bound GTP to GDP by 10\textsuperscript{11} using a mechanism whose details have been very controversial.\textsuperscript{30, 48} In particular, LFER and KIE studies have supported a proposal that the hydrolysis of GTP in water is a dissociative process.\textsuperscript{66} This analysis has been extrapolated to the Ras-catalyzed reaction,\textsuperscript{67} and recently endorsed by KIE analysis as having a loose TS.\textsuperscript{68} Aqueous hydrolysis has been developed in a QM study that has invoked a second water molecule to assist in proton transfer in the TS.\textsuperscript{52, 56} This proposition has been developed into a “two water” mechanism for enzymatic hydrolysis of GTP based on a modest resolution structure for Ras at 2.5 Å resolution (PDB: 1wq1) which has a less closed assembly of residues in the TS.\textsuperscript{69}

What is the evidence from MF\textsubscript{x} studies? To date, over 30 octahedral and ttp x-ray structures of GDP-MF\textsubscript{x}, TSA complexes can be superposed to show that water attacks P\textsubscript{Y} “in-line” (Fig. 19A) in trigonal coordination with H-bonds donated to Thr37 and Gin63 (RhoA numbering), and in a compact TS.\textsuperscript{20} Moreover, there is no second water in any of the high-resolution TSA structures, the next nearest water being 6 Å distant from P\textsubscript{Y} (excepting two waters coordinating the catalytic Mg\textsuperscript{2+}). The \textsuperscript{19}F NMR spectrum of a RhoA/RhoGAP-GDP-MgF\textsubscript{3}–TSA complex has identified F\textsubscript{1} as the most shielded fluorine and DFT computation extends that analysis to O1G as the most electronegative oxygen. High-level QM calculations, using 87 heavy atoms drawn from 17 amino acids, show that, for RhoA/RhoGAP, the MgF\textsubscript{3}–complex accurately mimics the true TS for phosphoryl transfer. It involves neither torsional phosphate strain nor GABC, and has an “in-line” angle of 175˚ with an O–P–O distance of 4.27 Å in a tight TS. The primary barrier to GTP hydrolysis appears to be the propensity of water to H-bond to an oxygen on the terminal phosphoryl group, as shown for 18 structures of small G proteins with GPPNP that have the water H-bonded to O2G (Fig. 19B). This denies orbital overlap between nucleophile and electrophile. Thus, the core of the catalytic mechanism is the orientation of both protons on the water away from GTP by passive H-bonds. It enables the nucleophilic oxygen to effect occupied orbital overlap with the antibonding orbital of P\textsubscript{Y} (Fig. 19C). GABC is not needed in GTPases, as has been confirmed by computation showing that the protons remain on oxygen in the TS.\textsuperscript{20}

Figure. 19 (A) Catalytic site for 8 small G proteins in ttp GDP-MF\textsubscript{x} complexes (green). Nucleophilic water complexed to M (2.1 Å) in-line and H-bonded to Thr37 and Gin63. (B) Catalytic site for GSA structures of 18 small G proteins with GPPNP (blue) H-bonded to water at 3.4 Å separation in NAC complexes.\textsuperscript{20} (C) Cartoon showing change in water orientation from GS to intermediate stage and to TS through completion of the H-bond network by GAP protein.\textsuperscript{20}

9. Conclusions

The three primary MF\textsubscript{x} species are trifluoroberyllate, tetrafluoroaluminate, and trifluoromagnesate. Structural, spectroscopic, and computational methods have combined to validate their use as surrogates for the phosphoryl group in ground state and transition state analog complexes for a wide variety of enzymes. The results achieved through their use have delivered detailed of phosphoryl transfer at the atomic level and supported investigations of protein folding and aggregation for tertiary structure problems. However, their use has been overwhelmingly committed to studies on terminal, dianionic phosphates and their reactions, with hardly any incursion into phosphate diester chemistry. That remains a major challenge for the future.

Acknowledgements

The authors thank Drs. Nicola Baxter, Matthew Bowler, Matthew Cliff and Robert Molt for valuable discussions and input during the preparation of this review. This work was supported by the BBSRC and University of Sheffield, UK, and by Cardiff University, UK. Y. J. is funded by ERC Advanced Grant AdG-322942.

Keywords: Metal fluorides • transition state analogs • phosphoryl transfer • enzyme mechanisms • \textsuperscript{19}F NMR spectroscopy • DFT analysis

References:

[References...]

Image 310x535 to 548x758


E. Kowalski, A. Schuller, R. Green, 4524-4529.


<table>
<thead>
<tr>
<th>Protein</th>
<th>PO₃⁻ Donor</th>
<th>PO₃⁻ Acceptor</th>
<th>PDB1 Reactant complex</th>
<th>PDB2 TSA complex</th>
<th>PDB3 Product complex</th>
<th>Pᵣ₋Pₚ dist Å</th>
<th>Oᵣ₋Oₚ dist Å OG¹</th>
<th>Oᵣ₋Oₚ dist Å OG²</th>
<th>Oᵣ₋Oₚ dist Å global</th>
<th>Od₋Oa distance TSA</th>
<th>Od₋Oa distance reactant</th>
<th>Od₋Oa distance product</th>
<th>Od₋Oa distance global</th>
<th>Oᵣ₋Pᵣ₋Oₚ angle</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ecoAcid Pase</td>
<td>AspP</td>
<td>Water</td>
<td>2heg 1my</td>
<td>1.43</td>
<td>0.48</td>
<td>0.54</td>
<td>0.45</td>
<td>0.49</td>
<td>5.0</td>
<td>4.21</td>
<td>4.50</td>
<td>4.57</td>
<td>170.23</td>
<td>1.20 ± 0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2hf7</td>
<td>1.24</td>
<td>0.59</td>
<td>1.00</td>
<td>0.66</td>
<td>0.75</td>
<td>4.53</td>
<td>4.17</td>
<td>4.71</td>
<td>4.47</td>
<td>173.20</td>
<td>0.24 ± 0.46</td>
<td></td>
</tr>
<tr>
<td>AK</td>
<td>ATP</td>
<td>AMP</td>
<td>1ank 4cf7</td>
<td>1.06</td>
<td>-0.50</td>
<td>0.51</td>
<td>0.26</td>
<td>0.09</td>
<td>4.52</td>
<td>4.28</td>
<td>4.33</td>
<td>4.30</td>
<td>162.18</td>
<td>0.39 ± 0.49</td>
<td></td>
</tr>
<tr>
<td>hPGK</td>
<td>ATP</td>
<td>SerOH</td>
<td>1rdq 113r</td>
<td>1.21</td>
<td>0.23</td>
<td>0.58</td>
<td>0.59</td>
<td>0.15</td>
<td>4.55</td>
<td>4.27</td>
<td>4.54</td>
<td>4.58</td>
<td>170.91°</td>
<td>1.30 ± 0.14</td>
<td></td>
</tr>
<tr>
<td>βPGM</td>
<td>AspP</td>
<td>G1P</td>
<td>2wfb 2wf8</td>
<td>1.30</td>
<td>0.55</td>
<td>0.58</td>
<td>0.22</td>
<td>0.45</td>
<td>n/a</td>
<td>n/a</td>
<td>4.20</td>
<td>4.41</td>
<td>4.30</td>
<td>176.45</td>
<td>1.43 ± 0.59</td>
</tr>
<tr>
<td>hPP1P5K2</td>
<td>ATP</td>
<td>InsP7</td>
<td>3t9c 3t9e</td>
<td>1.36</td>
<td>0.40</td>
<td>0.50</td>
<td>0.58</td>
<td>0.49</td>
<td>4.66</td>
<td>4.20</td>
<td>4.66</td>
<td>4.84</td>
<td>167.13°</td>
<td>1.20 ± 0.18</td>
<td></td>
</tr>
<tr>
<td>PSP</td>
<td>AspP</td>
<td>SerOH</td>
<td>117n 1j97</td>
<td>0.98</td>
<td>0.18</td>
<td>0.48</td>
<td>0.28</td>
<td>0.00</td>
<td>5.07</td>
<td>4.24</td>
<td>5.45</td>
<td>4.79</td>
<td>173.93</td>
<td>1.10 ± 0.43</td>
<td></td>
</tr>
<tr>
<td>Rab11a</td>
<td>GTP</td>
<td>Water</td>
<td>1oiw 1grm</td>
<td>1.10</td>
<td>0.43</td>
<td>0.48</td>
<td>0.76</td>
<td>0.24</td>
<td>n/a</td>
<td>4.39</td>
<td>4.68</td>
<td>4.55</td>
<td>157.49°</td>
<td>1.30 ± 0.14</td>
<td></td>
</tr>
<tr>
<td>Ras</td>
<td>GTP</td>
<td>1ctq</td>
<td>1wq1 1xd2</td>
<td>1.39</td>
<td>0.65</td>
<td>0.81</td>
<td>1.15°</td>
<td>0.73</td>
<td>6.62°</td>
<td>4.45</td>
<td>4.67</td>
<td>4.61</td>
<td>165.13°</td>
<td>1.39 ± 0.14</td>
<td></td>
</tr>
<tr>
<td>RhoA.GAP</td>
<td>GTP</td>
<td>Water</td>
<td>1a2b 1ow3</td>
<td>0.93</td>
<td>-0.66</td>
<td>0.38</td>
<td>0.53</td>
<td>0.08</td>
<td>5.24</td>
<td>4.19</td>
<td>4.44</td>
<td>4.62</td>
<td>172.38°</td>
<td>1.20 ± 0.18</td>
<td></td>
</tr>
</tbody>
</table>

Mean ± SD:
- 1.20 ± 0.24 ± 0.39 ± 0.48 ± 0.37 ± 4.80 ± 4.26 ± 4.55 ± 170.2° ± 4.6°

(a) Refined (by Dr Matt Bowler) as MgF₂⁻ on the basis of ¹⁹F NMR analysis
(b) Data in italics is ≥ 2 S.D. from the mean and so omitted from analysis
(c) Unreleased
(d) Order for the three O−O distances is clockwise (with catMg behind) and O1G coordinated to magnesium

References:
[1]