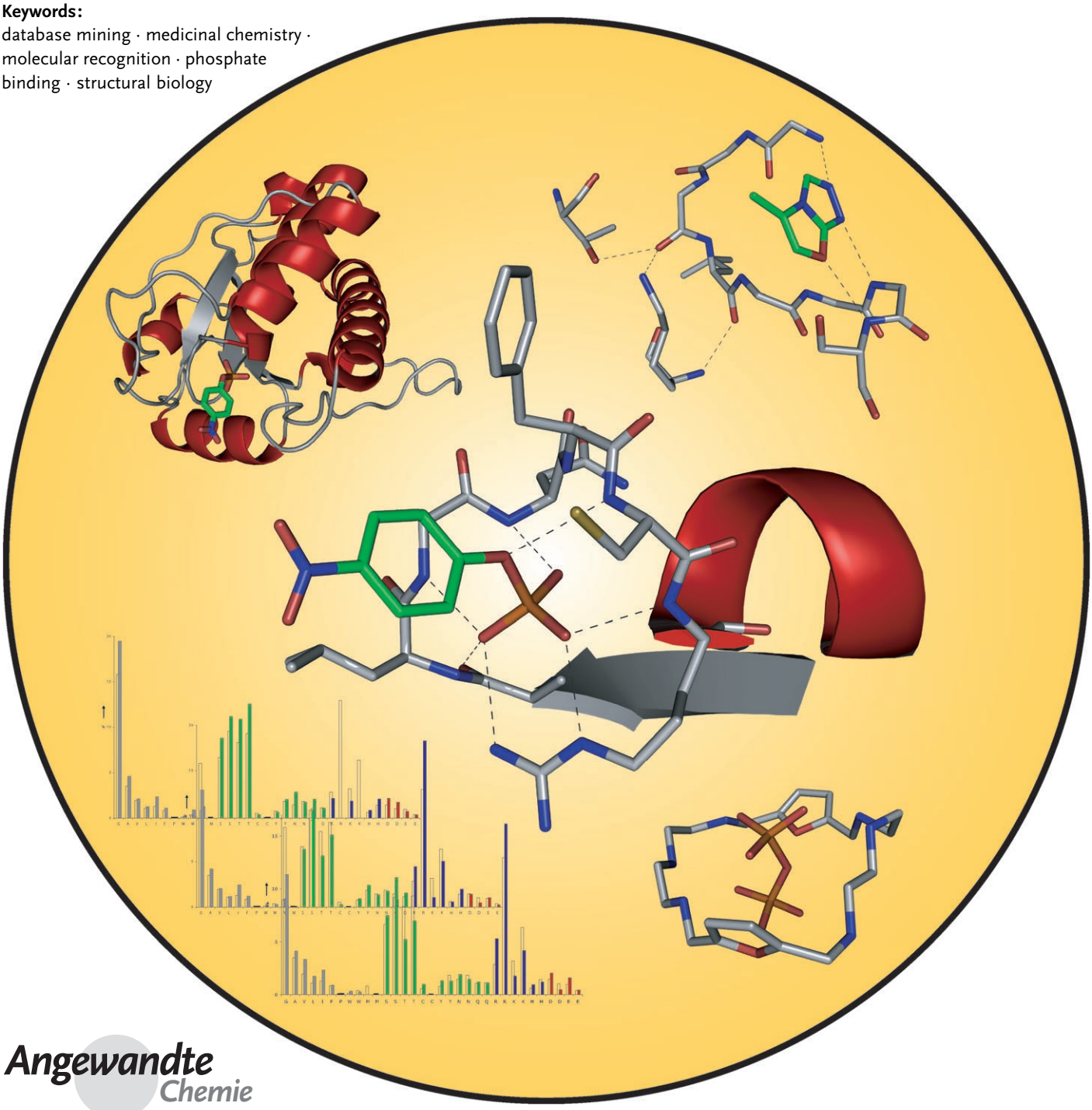
Molecular Recognition

DOI: 10.1002/anie.200603420

Phosphate Recognition in Structural Biology

*Anna K. H. Hirsch, Felix R. Fischer, and François Diederich**

Keywords:
database mining · medicinal chemistry · molecular recognition · phosphate binding · structural biology



Drug-discovery research in the past decade has seen an increased selection of targets with phosphate recognition sites, such as protein kinases and phosphatases, in the past decade. This review attempts, with the help of database-mining tools, to give an overview of the most important principles in molecular recognition of phosphate groups by enzymes. A total of 3003 X-ray crystal structures from the RCSB Protein Data Bank with bound organophosphates has been analyzed individually, in particular for H-bonding interactions between proteins and ligands. The various known binding motifs for phosphate binding are reviewed, and similarities to phosphate complexation by synthetic receptors are highlighted. An analysis of the propensities of amino acids in various classes of phosphate-binding enzymes showed characteristic distributions of amino acids used for phosphate binding. This review demonstrates that structure-based lead development and optimization should carefully address the phosphate-binding-site environment and also proposes new alternatives for filling such sites.

From the Contents

1. Introduction

Over the past years, we have pursued a multidimensional approach aimed at deciphering molecular recognition principles in chemical and biological systems. This approach includes structural investigations with proteins and synthetic receptors, biological assays and host-guest binding studies, mining of the Cambridge Structural Database (CSD) and the Protein Data Bank (PDB), theoretical calculations, and takes into account results from gas-phase investigations. The power of this multidimensional approach, which generates knowledge that greatly facilitates structure-based drug design and lead optimization, has already been documented with the analysis of interactions with aromatic rings,[1] orthogonal multipolar interactions,[2] and cation–π interactions in aromatic boxes at enzyme active sites.[3] In these investigations, we developed substantial competence in the nontrivial mining of the PDB, which also greatly benefits this study.

We recently became interested in new antimalarial targets[4] and started the development of inhibitors against IspF (2*C*-methyl-D-erythritol-2,4-cyclodiphosphate synthase, *ybgB*), one of the seven enzymes in the non-mevalonate pathway.[5] This pathway is used for isoprenoid biosynthesis by the malarial plasmodium (and other) parasites but not by humans. On the way from pyruvate and D-glyceraldehyde-3-phosphate to the isoprenoid precursors, isopentenyl diphosphate and dimethylallyl diphosphate, these enzymes process small mono- and diphosphates. The phosphate moieties contribute a large part of the molecular mass of these rather hydrophilic substrates. Only phosphate- and phosphonate-based ligands with modest to moderate affinities have been reported so far for some of the enzymes (DXS, IspC, IspF) in the non-mevalonate pathway.[6] This started our interest in learning more about biological phosphate recognition and provided the incentive for this research review. An enhanced understanding of phosphate binding sites should benefit the development of new phosphate analogues in drug development.

In fact, phosphate groups are ubiquitous in biology and nearly half of all known proteins interact with partners containing such a residue. They are an integral part of recognition events involving proteins, nucleic acids, cofactors, and antibodies. Binding of a phosphoryl group is essential to a myriad of biological processes ranging from metabolism and biosynthesis, to gene regulation, signal transduction, muscle contraction, and antibiotic resistance. Phosphate binding confers extra stability to enzymes such as Asp aminotransferase[7] and induces iron deposition in apoferritin.[8]

Protein phosphorylation is central to regulating transmembrane and intracellular signal transduction pathways.[9,10] Although protein kinases (PKs)[11,12] catalyze the transfer of the terminal phosphate group from adenosine triphosphate (ATP) to specific amino acid residues such as serine, threonine, and tyrosine, the process is reversed with the help of protein phosphatases (PPs),[13] which cleave the phosphate residue from the amino acid. Both classes of enzymes deliver some of the most important targets in the fight against diseases such as cancer and obesity. While the development of small-molecule PK inhibitors in most cases avoids occupation of the highly polar triphosphate binding site of ATP,[14,15] many PP inhibitors occupy the phosphate site with a phosphate-mimicking moiety.[16,17] Phosphates themselves would confer unfavorable pharmacokinetic properties to a ligand, such as low membrane permeability and hydrolytic instability. Although a variety of phosphate surrogates

[*] A. K. H. Hirsch,[+] F. R. Fischer,[+] Prof. Dr. F. Diederich
Laboratorium für Organische Chemie
ETH Zürich
Hönggerberg, HCI, 8093 Zürich (Switzerland)
Fax: (+41) 1-632-1109
E-mail: diederich@org.chem.ethz.ch
[+] AKHH and FRF made equal contributions to this review.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

have been introduced,[17] including vanadate-based phosphate analogues, a diversity of conjugated anions of acids such as carboxylic, tetronic, oxamic, difluoromethylenesulfonic and difluoromethylenephosphonic acids, and other acidic residues, the search for efficient phosphate analogues with desired binding and physicochemical properties is still very much ongoing. We hope that this review will contribute to facilitating this search.

To the best of our knowledge, there is no comprehensive evaluation of phosphate binding in proteins that uses the full array of X-ray crystal structures deposited in the PDB.[18] This review analyzes the molecular recognition of phosphate groups that use all protein–ligand complexes available in the data bank. Given that the focus is on the interactions between proteins and drug-like molecules, structures containing DNA/RNA were excluded.

This review starts with a discussion of the known phosphate binding modes, such as the P loop.[19–21] We shall note on a few occasions the close analogy of some of the biological binding motifs to those seen in the phosphate complexation by synthetic receptors. The field of anion recognition by synthetic binders has been well reviewed.[22] In the following, we provide a statistical evaluation of the X-ray crystal structures of proteins with bound phosphate ligands. Several issues are addressed, such as 1) the types of amino acids involved in the recognition process, 2) the role of amino acid side chains compared with protein backbone binding, 3) the involvement of metal ions, 4) the role of basic amino acids, 5) the role of specific binding motifs such as the P loop, and 6) possible binding characteristics for selected classes of enzymes. The final section is devoted to phosphate binding by PKs and PPs in view of the eminent role of these enzymes as targets in medicinal chemistry.

2. Known Phosphate Binding Modes

In 1974, Rossmann et al. identified a common protein fold of dinucleotide binding proteins, known as the "Rossmann fold", which is also seen in mononucleotide binding proteins.[23] Its key features are a parallel β sheet with α helices connecting the strands in a right-handed manner. The ever-increasing number of protein–ligand X-ray crystal structures led to the identification of "sequence fingerprints" that became useful in the identification of the function of new proteins.

2.1. Glycine-Rich Sequence

Originating with the discovery of the Rossmann fold, two consensus sequences have been identified, which can be treated as fingerprints for mono- and dinucleotide binding, respectively.[21,24] They are referred to as Gly-rich sequences with X referring to any amino acid and alternative residues at one position (such as S, T) shown in brackets:[21,25]

- GXGXXG for dinucleotide binding
- GXXGXGK(S,T) or GXXX for mononucleotide binding

2.2. Dinucleotide Binding Proteins

These proteins bind nicotinamide adenine dinucleotide (NAD) and the corresponding phosphate (NADP) or flavin adenine dinucleotide (FAD). The Gly-rich element is located at a tight turn between a β strand and an α helix of a Rossmann fold. Invariably, the phosphate groups are stabilized by the positively charged N-terminal domain of the helix dipole.[26,27] The conserved Gly resides are important for several reasons: they provide space for the complexation of the bulky diphosphate ion and make a tight turn possible. One of the exceptions to this general binding motif is exemplified by aldole reductase, which employs an alternative NADP binding motif.[28]

2.3. Mononucleotide Binding Proteins

A comparison of 491 mononucleotide binding sites found in the PDB by Kinoshita et al. in 1999 led to the identification of a conserved four-residue sequence: GXXX, called a "structural P loop".[29] This sequence includes a number of previously described sequence motifs such as the P loop or the GXGXXG consensus sequence in PKs. This motif is shared by 13 superfamilies of proteins. Other motifs were identified that are shared by merely two superfamilies and some do not have a consensus sequence at all.

Anna K. H. Hirsch was born in 1982 in Trier, Germany. She studied Natural Sciences and Chemistry at the University of Cambridge where she did her Master's thesis in the group of Prof. S. V. Ley. During her undergraduate degree, she spent a year studying at the Massachusetts Institute of Technology and worked in the group of Prof. T. Jamison. She joined the group of Prof. François Diederich in 2004 and her PhD thesis is concerned with the design and synthesis of inhibitors for IspE.

Felix R. Fischer was born in 1980 in Germany. He studied chemistry at the Ruprecht-Karls-University in Heidelberg and received his diploma in 2004 under the supervision of Prof. R. Gleiter. In 2005 he joined the group of Prof. François Diederich at the ETH Zürich for his PhD thesis. Currently he is working on the design and the synthesis of model systems for the measurement of biologically relevant multipolar interactions.

2.4. P loop

Originally called motif A by Walker et al., the P loop is commonly found in ATP and GTP binding proteins.[30] Furthermore, a less well-conserved second site, called motif B, can be present.

The consensus sequence for the P loop is GXXXXGK-(S,T). Within protein families, it is possible to refine the consensus sequences as common features are often shared within a family. Wittinghofer and co-workers compared seven selected ATP- or GTP-binding protein families in the Swissprot database. An example of a refined consensus sequence for adenylate kinases is GXPGXGKGT, which features a Gly inserted between the conserved Lys and Thr residues.[20] The conserved Lys residue is present in all cases and is postulated to be both important for the conformation of the P loop as well as for the stabilization of the β- and γ-phosphates. In addition, it is believed to accompany the transferred terminal phosphoryl group. Superpositition of a number of P-loop-containing proteins showed the positions of the α- and β-phosphates to be identical.[28] The two conserved Gly residues adopt conformations that would not be tolerated by any amino acid with a side chain.

As opposed to dinucleotide binding motifs, the P loop is rather long and connects a β sheet with an α helix. Just as for dinucleotide binding proteins, the loop is usually found at the N terminus of an α helix. The P loop is sometimes referred to as "giant anion hole".[19] In a wider sense, it was recently described as a "nest". This is defined as a three to six amino acid motif in which successive backbone amide groups bind anions such as phosphates or iron sulfur centers.[31]

An example of a protein containing a P loop is p21, the product of the *H-ras* oncogene.[32] Pai et al. solved the X-ray crystal structure of p21 in complex with the slowly hydrolyzing GTP analogue 5′-guanyl-β,γ-amidotriphosphate (GppNp) and a Mg^{2+} cation (Figure 1, PDB code: 5P21).[33] The conserved P loop in the phosphate binding site stretches from residues 10–18 with the consensus sequence GXXXXGKS. Thr 35 has an additional side-chain interaction with the γ-phosphate. As already pointed out, the Ramachandran diagram of the refined structure shows the highly conserved Gly 10 and Gly 15 in conformations that are only allowed for Gly residues. The phosphate groups are surrounded by a positively polarized electrostatic field set up by the backbone NH groups of residues 13–18, which all point towards the phosphate groups and undergo ionic H-bonding.

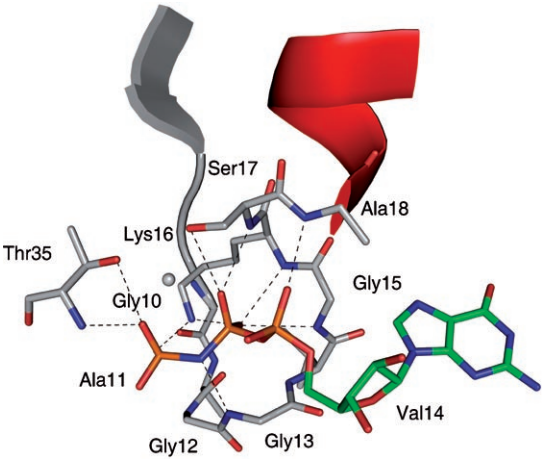


Figure 1. Section of the X-ray crystal structure of p21 binding the GTP analogue GppNp by using a P loop (PDB code: 5P21, 1.35-Å resolution).[33] Dashed lines are shown for H-bonding contacts below 3.2 Å (distance between heavy atoms). Color code: ligand skeleton: green; C: gray, O: red; N: blue; P: orange. This distance selection for H-bonding and the color code are maintained throughout the review if not otherwise stated.

François Diederich, born in Luxemburg (1952), studied chemistry at the University of Heidelberg (1971–1977) and completed his PhD with Prof. H. A. Staab (1979). After postdoctoral studies with Prof. O. L. Chapman at UCLA (1979–1981), he returned to Heidelberg for his habilitation at the Max-Planck-Institut für Medizinische Forschung (1981–1985). 1989 he became Full Professor of Organic and Bioorganic Chemistry at UCLA. In 1992, he joined the ETH Zürich. His research interests include dendritic mimics of globular proteins, synthetic and biological receptors, and non-peptidic enzyme inhibitors.

2.5. Novel P loop

A novel nucleotide binding fold has been identified for the galacto kinase, homoserine kinase, mevalonate kinase, phosphomevalonate kinase (GHMP) superfamily. It was named the "novel P loop" and is distinct from the classical P loop as it binds ADP/ATP in the unusual *syn* conformation.[34] Nucleotides are usually bound in the *anti* conformation. A highly conserved motif, originally called motif 2, was identified as PXXXGLGSSAA in a loop between strand F and the α-helix B. This loop forms an enormous anion hole, which is also located at the N terminus of an α helix.

The novel and the classical P loop share some similarities: both are located between a β strand and an α helix and use the stabilizing effects of the helix dipole and ionic H-bonds to backbone amides for phosphate binding. The structure and sequence, however, are different: the novel P loop is two amino acids longer and the conserved Lys/Arg is absent. It can be postulated that the longer loop provides more ionic P–O^-···H–N H-bonds, which could compensate for the lack of the positively charged Lys/Arg side chain.

2.6. Protein Kinases

An alignment of catalytic domain amino acid sequences of 65 PKs led to the identification of a conserved sequence GXGXXG.[28,35] This is the same as that for dinucleotide binding proteins. Structurally, however, PKs have phosphate binding domains, which are more similar to those of

mononucleotide binding proteins.[21] It seems clear that both the Gly-rich anion hole as well as the vicinal Lys residue are essential for phosphoryl transfer as both seem to have evolved independently for two distinct chain folds: the classical P loop and the PK fold.

2.7. The $C^{\alpha}NN$ Structural Motif

Denesyuk et al. identified a novel anion binding motif, starting from their original "phosphate-group binding cup", which was identified in pyridoxal-5′-phosphate (PLP) binding proteins.[36] By performing a structural analysis of all fold-representative protein complexes of the FSSP (families of structurally similar proteins) database,[37] they identified a motif that is common to 62 different folds. It recognizes both free phosphate and sulfate ions as well as phosphate groups in nucleotides and cofactors. The motif includes one C^{α} and two backbone N atoms and is usually found in functionally important regions of the protein.

The complex of pig cytosolic Asp aminotransferase and PLP shows a clear example of such a binding element (Figure 2, PDB code: 1AJS).[38] The phosphate moiety of PLP

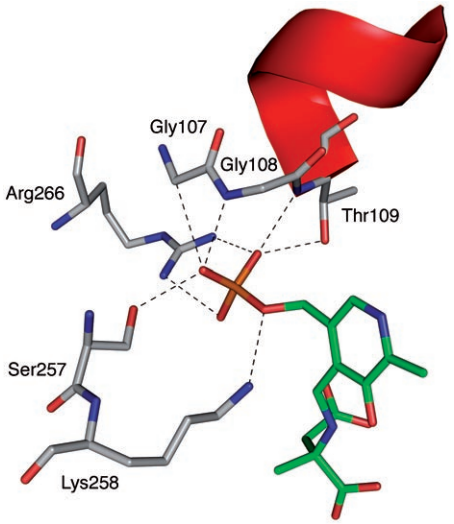


Figure 2. Section of the X-ray crystal structure of pig cytosolic Asp aminotransferase complexed with pyridoxal-5′-phosphate by using the $C^{\alpha}NN$ structural motif (EC 2.6.1.1, PDB code: 1AJS, 1.60-Å resolution).[38]

is held in place by interactions from the $C^{\alpha}NN$ element: a (very weak) C–H···O H-bond with Gly107 (heavy atom distance 3.39 Å) and two strong (ionic) H-bonds with the backbone N atoms of Gly108 and Thr109. In addition to this, the phosphate moiety is stabilized by H-bonds with the side chains of Thr109 and Ser257, as well as ion pairing (accompanied by ionic H-bonds) with Lys258 and Arg266.

In roughly 90% of the cases, the motif was shown to be phosphate specific. Even though this motif is clearly distinct from other anion binding sites, it also takes advantage of amino acids with small or no side chain, Gly in particular, and uses main-chain H-bond interactions for phosphate binding.[39]

Taken together, the results in Sections 2.1–2.7 clearly show that nucleotide binding proteins often feature specific chain folds and a number of characteristic structural features: Gly residues as part of a loop, an adjacent Lys residue participating in phosphoryl transport, and the proximity of the positively polarized N terminus of an α helix. Nevertheless, the presented chain folds and sequence fingerprints are not the only phosphate binding motifs in proteins. Actin, HSP70 (heat shock protein 70), and sugar kinases such as hexokinases, for instance, bind phosphate groups with residues from two β hairpins.[34]

2.8. Some Comparisons with Synthetic Phosphate Receptors[22]

Phosphate binding by synthetic receptors in aqueous solution requires multiple charge interactions accompanied by ionic H-bonding. Stable complexes form with a variety of fully protonated macrocyclic polyamines such as $\mathbf{1}\cdot 6H^{+}$ or $\mathbf{2}\cdot 6H^{+}$ (Figure 3) introduced by Lehn.[22a] Binding strength increases with the number of charge–charge interactions, in the case of $\mathbf{1}\cdot 6H^{+}$ from adenosine monophosphate (AMP;

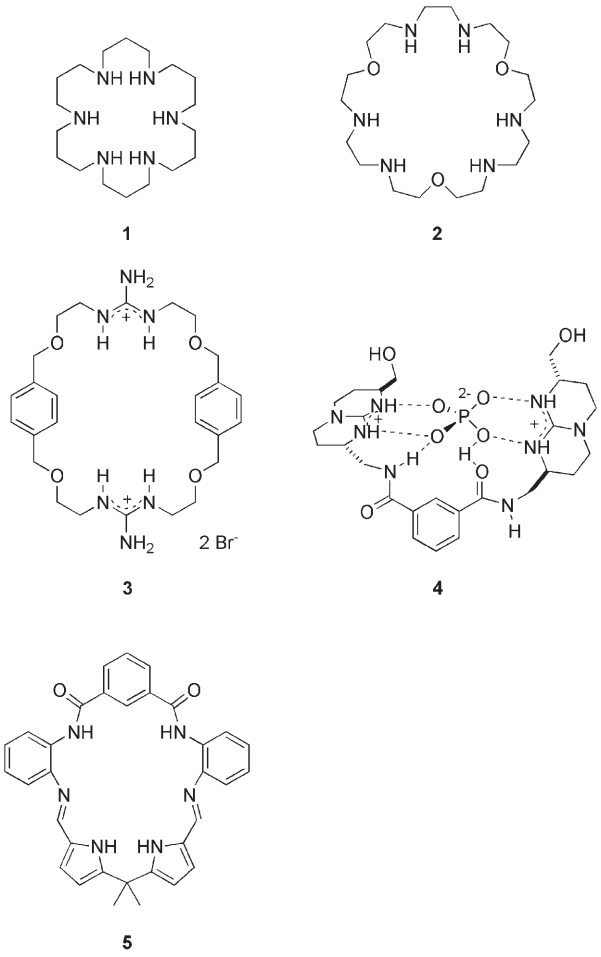


Figure 3. Synthetic receptors for phosphate anions.

$\log K_{ass} = 3.4$ in 0.1M aqueous Me_4NCl; ass = association), to adenosine diphosphate (ADP; 6.5), and ATP (8.9). The selectivity, however, is rather low and other anions such as oxalate (3.8), sulfate (4.0), and citrate (4.7) are also bound.[40a] In the host–guest complexes, the macrocycles wrap around the bulky phosphate anions forming nesting-type complexes with optimized H-bond ion pairing.[40b,c]

Receptors based on guanidinium ions have been intensively investigated. Thus, macrocycle **3** complexes PO_4^{3-} in water with $\log K_{ass} = 1.7$ and in $MeOH/H_2O$ (9:1) with $\log K_{ass} = 3.1$.[41] A variety of cleft-type mono- and bis-guanidinium receptors have been shown to efficiently complex oxoanions in polar solvents; in complex **4**, which was formed in Me_2SO, two guanidinium residues are proposed to coordinate in a tetrahedral fashion to HPO_4^{2-}.[22c,42] A general comparison[22] suggests that the performance of one or two primary ammonium ions (RNH_3^+) interacting with phosphate anions is much weaker than the binding of a phosphate by a single guanidinium residue. Phosphate recognition by guanidinium ions benefits from the "chelate effect" and the formation of convergent ionic H-bonds with favorable secondary electrostatic interaction patterns. By analogy, we believe that the contribution of an interacting Arg side chain to the binding free energy in protein–phosphate complexes is most probably much larger than that of a Lys side chain.

Furthermore, in weak or noncompetitive (in terms of the H-bonding capacity) solvents, complexation can be realized by means of amide and heterocyclic NH residues converging towards the bound phosphate ion. This has been nicely shown by Sessler et al. for macrocycle **5**, which is proposed to wrap around a nesting $H_2PO_4^-$ anion in MeCN ($K_{ass} = 340\,000\ L\,mol^{-1}$), thereby benefiting from interactions with the two amide and pyrrole NH moieties.[43]

Both the wrapping of neutral NH residues around the phosphate ion bound to **5** as well as the encircling of the oxoanion by the protonated $-NH_2^+-$ residues of the macrocyclic polyamines **1**·6 H^+ or **2**·6 H^+ create bonding geometries that closely resemble those of P loop sites such as shown in Figure 1. In fact, the 3D visualization of phosphate binding by P loops reveals, for quite a number of protein complexes, an astonishing, near-macrocyclic organization of the converging NH residues around the bound anion. The loop wraps around the phosphate residue to optimize its ability to form ionic H-bonds or, in other words, the anion optimizes the geometry of its receptor site (see the complex of a triphosphate analogue with the enzyme IspE (PDB code: 1OJ4) in the non-mevalonate pathway of isoprenoid biosynthesis, Figure 6, Section 3.3).[44]

Nature has optimized the spatial arrangement of anion binding sites to an extent that allows even for differentiation between such small differences in size and charge as in sulfate and phosphate.[45]

3. Statistical Evaluation

Even though anion binding sites in proteins have been examined before, making use of the information from X-ray crystal structures of protein complexes with phosphate- and sulfate-containing ligands, the statistical evaluation was limited to rather small datasets (< 70 structures).[28,44] In addition to this, a number of evaluations have been performed for individual enzyme classes or specific nucleotides such as ATP.[46] To our knowledge, there is, however, no comprehensive evaluation of phosphate binding in proteins by using the ensemble of X-ray crystal structures in the PDB.

To enhance the understanding of the way phosphate groups are bound within proteins, we conducted a comprehensive PDB search.[18] Our analysis is based on atomic coordinates available in this data bank. If more than one X-ray crystal structure was available, the one with the better resolution was considered. In the case of oligomeric proteins, the analysis was restricted to one subunit as homologous subunits generally have an identical mode of binding.

The program Relibase was used to study the short-contact interactions of phosphate groups in proteins.[47,48] We defined the following search parameters (Figure 4): First, the search

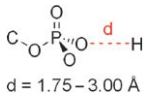


Figure 4. Parameters used for the Relibase search.

was limited to α-phosphates bound to a C atom to exclude X-ray crystal structures containing free phosphates, which are often cocrystallized in locations that do not correspond to the active site as a result of the crystallization conditions used. As the free valencies on the O atom of the α-phosphate are not defined, this also takes into account structures featuring di-, tri-, and pentaphosphates. Second, we imposed a distance constraint between the phosphate O atom and a protein H atom of 1.75–3.00 Å. We chose this rather unconventional description of a H-bond to include all types of H-bond donors (HO, HN, HS) in one search. Finally, for the evaluation, we examined all possible H-bonds by using a cut off of 3.20 Å between heavy atoms.

3.1. The Entire Set of Structures ("All")

As of February 14, 2006, the total number of structures in the PDB amounted to 35 144. The amino acid propensities for the entire data set are shown in 1SI in the Supporting Information.[49] A total of 14 590 entries showed the structural features searched for. Out of these, 3003 matched the imposed distance constraint; they form the entire set named "**All**". This group featured a total of 19 713 H-bonds to 5520 phosphate groups, corresponding to an average of 3.6 H-bonds per phosphate group. All 3003 structures were individually visualized and inspected. The bar graph in 2SI in the Supporting Information shows the percentage of the various amino acids that participate in phosphate recognition. Note that amino acids with H-bonding side chains are listed twice, with the first percentile indicating participation of the backbone NH group and the second indicating the participation of the side chain. The comparison between the amino acid

propensities in the entire PDB and those involved in phosphate binding revealed, as expected, a large difference.

3.2. A First Subset: Omitting Identified Metal Ions ("All–M")

Given that anionic phosphate groups can undergo ion pairing with metal ions and as this Coulombic interaction would almost certainly override any H-bonding effects, we decided to exclude structures involving binding of metal ions to the phosphate substrate. This deletion left 2456 entries (out of 3003): metal ions are far less involved in phosphate recognition than is commonly expected. This first subset was named "**All–M**". The comparison of the amino acids involved in H-bonding to phosphate residues in all structures ("**All**") with the first subset "**All–M**" (2SI in the Supporting Information) shows nearly identical amino acid patterns. Noteworthy are the high proportions of Gly residues, the polar residues Ser and Thr, and expectedly the basic amino acids Lys and Arg. The high occurrence of Gly is in agreement with the observation that Gly-rich loops are important phosphate binding motifs (see Section 2): Gly residues conformationally allow the folding and wrapping of the loop around the bound phosphate. More than half of the entries in the subset have a Lys or Arg side chain involved in phosphate binding. The overall proportion of amino acids with apolar (total 25%) and basic (total 28%) side chains are rather similar. The rather low number of Tyr residues (compared with Ser and Thr) involved in H-bonding comes as a surprise as the OH group of Tyr is more acidic and thus should be the better H-bond donor.[50] Steric factors presumably lead to the preference for Ser and Thr over Tyr.

Given the rather small number of phosphate groups complexed by metal ions and the highly similar amino acid propensities in the presence and absence of such ions, we decided to use the subset "**All–M**" for all future comparisons.

3.3. A Second Subset: Phosphate Binding Sites Without Ion Pairing ("All–M–Lys/Arg")

From a medicinal chemistry viewpoint, it was of particular interest to explore to what extent phosphates are also bound at "neutral" recognition sites without the assistance of ion pairing with metal ions and/or protonated basic amino acid side chains. In such phosphate binding sites, H-bonding would be the major interaction and the recognition sites could be occupied by parts of lead compounds with pronounced multiple H-bonding acceptor capabilities.

Exclusion of X-ray crystal structures featuring Arg/Lys side chains involved in phosphate binding from the first subset "**All–M**" led to the unexpectedly high number of 1070 structures featuring a total of 5303 H-bonds to 1668 phosphate groups, an average of 3.2 H-bonds per phosphate moiety. Nearly a third of all phosphate binding sites do not employ ion pairing interactions!

A comparison of the two subsets "**All–M**" and "**All–M–Lys/Arg**" (Figure 5) shows that the lack of basic residues is compensated for by an increase in polar residues

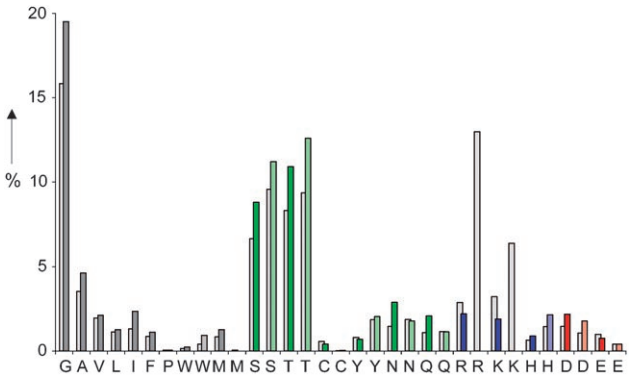


Figure 5. Bar graph showing a comparison of the amino acid residues (one-letter code) involved in H-bonding to phosphate groups of the first subset "**All–M**" (light-gray bars at the back) with the second subset "**All–M–Lys/Arg**" (colored bars at the front). Amino acids with side chains capable of forming H-bonds are shown twice with the first bar referring to backbone NH groups and the second shaded bar referring to side-chain interactions. The amino acids are grouped into classes and color coded accordingly: gray: apolar; green: polar; blue: basic; red: acidic. This color code is maintained throughout the article if not otherwise stated.

such as Ser, Thr, His (which could be protonated), and Asn in particular, as well as apolar residues such as Gly but also Ala or Ile. This trend seems to agree with the description of the novel P loop (Section 2.5) in which the absence of a conserved Lys residue is postulated to be compensated for by a longer loop that contains more conserved residues. Besides the amino acid side chains, backbone NH groups contribute to phosphate binding through H-bonding and by setting up a positive electrostatic environment.

A representative example of a protein from the "**All–M–Lys/Arg**" subset is presented by the ternary complex of 4-diphosphocytidyl-2*C*-methyl-D-erythritol kinase (IspE), its substrate, and the non-hydrolyzable ATP analogue 5′-adenyl-β,γ-amidotriphosphate (AppNp) (Figure 6, PDB

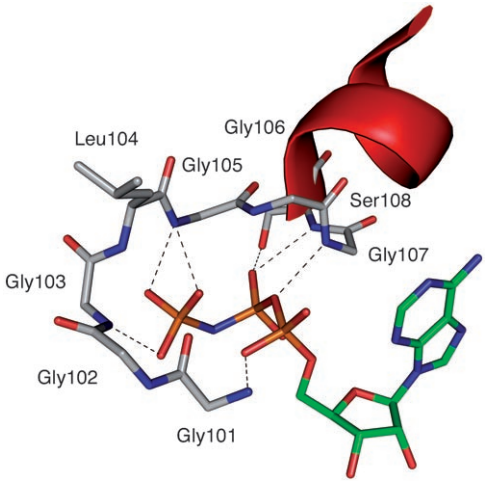


Figure 6. Section of the X-ray crystal structure of the enzyme IspE binding to the hydrolysis-resistant ATP analogue AppNp without the use of a metal ion or a positively charged side chain (EC 2.7.1.148, PDB code: 1OJ4, 2.01-Å resolution).[51]

code: 1OJ4).[51] IspE is one of the seven enzymes in the non-mevalonate pathway for the synthesis of the isoprenoid precursors isopentenyl diphosphate and dimethylallyl diphosphate, which are used in most bacteria and some parasites but not in humans. IspE displays a two-domain fold consisting of a substrate and an ATP binding domain, which are highly characteristic of kinases. In this complex, the ATP binding site is marked by the presence of a long Gly-rich loop that features only one polar Ser residue involved in phosphate binding, but a large number of Gly backbone NH moieties (residues 101–107) pointing towards the bound anion. Another example of a structure from the second subset can be found in 3SI in the Supporting Information.[52]

3.4. Subset "All–M" in Different Classes of Enzymes

The "**All–M**" subset was grouped into different classes of enzymes by using the description given in the PDB (Table 1). Only the four most populated classes as well as statistically significant changes in the amino acid propensities are discussed.

Table 1: Division of the first subset "**All–M**" into enzyme classes.

Enzyme class	Number	Proportion [%]
oxidoreductase	787	32
others	520	21
transferase	430	18
lyase	245	10
hydrolase	149	6
isomerase	119	5
signaling protein	77	3
electron transport	73	3

The bar graphs comparing the two most populated groups (oxidoreductases and transferases) within subset "**All–M**" clearly show that each enzyme class has a highly characteristic distribution of amino acids (Figure 7 and 4SI in the Supporting Information). Focusing on the most prominent group, the oxidoreductases, a rather low proportion of Gly residues is striking. On the other hand, the relative proportion of apolar residues such as Ala, Val, Leu, Ile, and Met is increased. This increase in apolar residues might be essential to tune the environmental polarity for an efficient electron-transfer processes. A reduction in Thr residues is compensated by an increased number of Ser residues. Similarly, an increase in Arg compensates for a lower number of Lys residues.

The second most common class, the transferases, does not show any significant changes in the apolar residues when compared with the entire subset. On the other hand, the polar subgroup is reversed with the number of Thr residues increased and the number of Ser residues reduced. Transferases and oxidoreductases feature nearly the same distribution of basic residues.

The next two most common groups, the lyases and isomerases, are compared with the entire subset "**All–M**" in Figure 8 and 5SI in the Supporting Information. The former shows an increase in the number of Gly residues that seems to be matched by a decline in the proportion of remaining apolar residues. Once more, the polar residues Ser and Thr residues show opposing trends, and the same applies to Arg and Lys residues.

In the class of isomerases, the apolar subgroup of amino acid residues shows a somewhat different behavior. The rise in the number of Leu and Ile residues makes up for a decreased contingent of Ala and Val residues. Interestingly, both Ser and Thr residues, which are usually very prominent phosphate binding residues, show a decimated number, whereas the increased number of Tyr and Asn residues as well as Gln side chains is highly unusual. Once more, this could imply a substituting role. The basic residues again exhibit typical behavior with a rise in the number of Lys residues mirrored by a decrease in Arg residues.

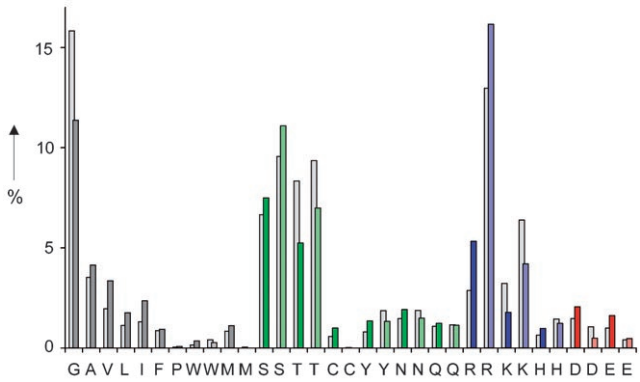


Figure 7. Bar graph showing a comparison of the amino acid residues (one-letter code) involved in H-bonding to phosphate groups of the first subset "**All–M**" (light-gray bars at the back) with its oxidoreductases (colored bars at the front). Amino acids with side chains capable of forming H-bonds are shown twice, with the first bar referring to backbone NH groups and the second shaded bar referring to side-chain interactions.

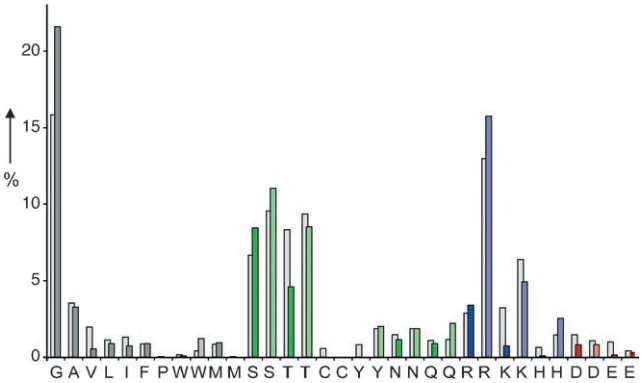


Figure 8. Bar graph showing a comparison of the amino acid residues (one-letter code) involved in H-bonding to phosphate groups of the first subset "**All–M**" (light-gray bars at the back) with its lyases (colored bars at the front). Amino acids with side chains capable of forming H-bonds are shown twice, with the first bar referring to backbone NH groups and the second shaded bar referring to side-chain interactions.

A general trend seems to emerge from this analysis. The proportion of the different subgroups of amino acid residues (apolar, polar, basic) involved in phosphate binding seems to be more or less constant. It seems, however, that each class of enzymes has a clear preference for the types of residues from each subgroup of amino acids that it uses for phosphate binding. For example, a decreased number of Thr residues is frequently mirrored by an increased number of Ser residues. Similar trends are obvious for the apolar and basic residues. For the subgroup of acidic amino acid residues, this does not seem to apply. The number of acidic residues involved in phosphate binding is, however, rather small, making it difficult to identify statistically significant changes. The result of this behavior is the emergence of highly characteristic distributions of amino acids used for phosphate binding, which could in theory be used as "fingerprints" to identify an enzyme class.

3.5. A Third Subset: Absence of Metal Ions and Loop ("All–M–loop")

Given that a loop is a common structural element for phosphate binding, the next logical step was to examine how phosphates are being bound in the absence of both a metal ion and a loop. For this purpose, we defined a loop to be a series of at least three consecutive amino acids that are involved in phosphate H-bonding. Also included are cases where the first and third, but not the second residue, form a H-bond. Subtraction of the concerned entries led to a third subset "**All–M–loop**". It contains a total of 1675 entries with 8428 H-bonds to 2740 phosphate groups. This is equivalent to an average of 3.1 H-bonds per phosphate group.

Figure 9 shows a comparison of the first ("**All–M**") with the third subset ("**All–M–loop**"). A decrease in the number of Gly residues is apparent and is mirrored by an increase in the number of Ala and Ile residues. The increase in Ser side chains and Tyr residues could be compensating for a decline in the number of Thr residues. Finally, in the basic subgroup, a rise in the involvement of His side chains and Arg residues might counteract a reduced number of Lys residues.

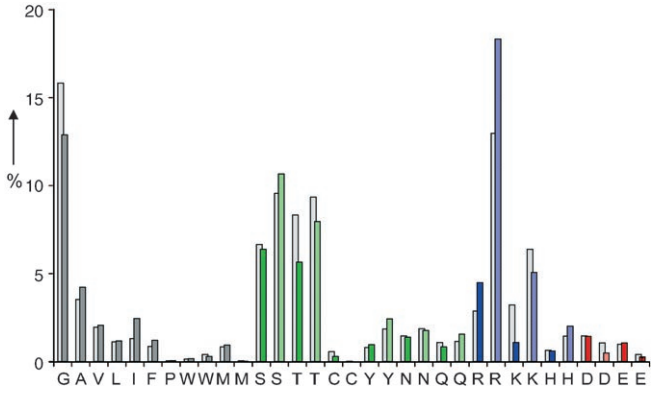


Figure 9. Bar graph showing a comparison of the amino acid residues (one-letter code) involved in H-bonding to phosphate groups of the first subset "**All–M**" (light-gray bars at the back) with the third subset "**All–M–loop**" (colored bars at the front). Amino acids with side chains capable of forming H-bonds are shown twice with the first bar referring to backbone NH groups and the second shaded bar referring to side-chain interactions.

An illustrative example of this third subset is the complex of the riboflavin kinase from *S. pombe* with one of its products, ADP (Figure 10, PDB code: 1N07).[53] The phosphate moieties of ADP are bound by three isolated residues, a Tyr side chain, a Gly backbone, and a Ser residue that interacts through an amide group in its backbone and also through the side chain. Even though the Gly and Ser residues are part of a structural loop, it does not correspond to our definition of a phosphate binding loop as there are too many residues located in between that do not contribute to phosphate binding.

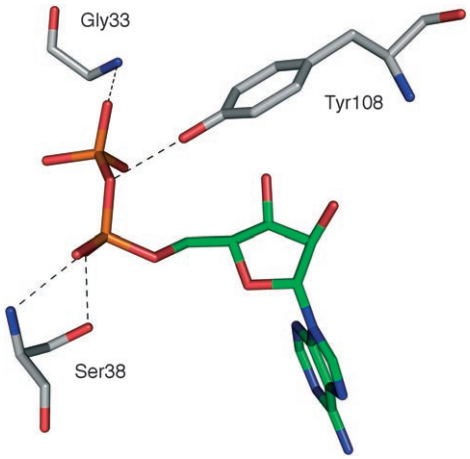


Figure 10. Section of the X-ray crystal structure of riboflavin kinase binding ADP without the assistance of a metal ion or a loop (EC 2.7.1.26, PDB code: 1N07, 2.45-Å resolution).[53]

The results of this statistical evaluation are summarized in Table 2. There are hardly any differences between the set "**All**" and the first subset ("**All–M**"), confirming once more our decision to exclusively use this subset. The second subset shows a clear drop in basic residues as the only remaining cases are His residues and the backbone amides of Lys and Arg. This is paralleled by a sharp rise in polar residues and to some extent also apolar residues. Expectedly, the acidic residues remain more or less unchanged.

The third subset shows a marginal increase in basic residues, which makes sense in that the basic residues could well compensate for the lack of a loop. Hence, it seems that the third subset follows a similar behavior to that observed for the different enzyme classes, that is, a compensation of residues of one type of amino acids for another (apolar, polar, basic).

4. Phosphate Binding by Protein Kinases and Phosphatases

It was already discovered in the early 1950s that the activity of enzymes can be regulated through phosphorylation and dephosphorylation.[10,54] PKs and PPs do not just have

Table 2: Summary of the statistical evaluation of the different subsets.

Subset	H-bonds per phosphate	H-bonds [%][a]				Entries [%][b]	
		Apolar	Polar	Basic	Acidic	Loop	Arg, Lys side chain
"All"	3.6	5023 (25)	8418 (43)	5533 (28)	717 (4)	1198 (40)	1733 (58)
"All−M"	3.6	4107 (26)	6734 (43)	4344 (27)	614 (4)	943 (38)	1434 (58)
"All−M−Arg/Lys"	3.2	1751 (33)	2858 (54)	372 (7)	267 (5)	344 (32)	–
"All−M−loop"	3.1	2141 (25)	3356 (40)	2659 (32)	272 (3)	–	908 (54)

[a] Number of H-bonds formed by each class of amino acid residues; the percentage that this subgroup represents is given in brackets. [b] Number of entries; the percentage that this represents is given in brackets.

opposing actions, rather they work together to regulate cell growth and differentiation.[55] A disruption of this equilibrium almost inevitably leads to disease. Hence, both PKs and PPs are important targets in medicinal chemistry. In the last part of this review, we therefore focus on the mechanisms of phosphate binding by PKs and PPs, aiming at enhancing the understanding of these recognition sites, which could potentially benefit lead developments in medicinal chemistry.

4.1. Protein Kinases

PKs, a subgroup of kinases,[56] phosphorylate OH groups of protein substrates. Protein serine–threonine kinases (PSTKs) are the most common, followed by protein tyrosine kinases (PTKs), and finally the so-called dual specificity kinases (DSKs), which can phosphorylate all three amino acids. In other classifications, the structural information on the specific protein–ligand interactions at the ATP binding site or sequence alignments were used to divide the PKs into subfamilies.[11,57]

Thanks to a wealth of both structural and biochemical studies, kinases are probably some of the best-studied enzymes. Their diversity and substrate specificity are remarkable considering that they all catalyze essentially the same reaction, the transfer of the γ-phosphate group of ATP to a substrate. A number of structural features that are a recurring theme amongst kinases have been identified and extensively reviewed.[58] They include the presence of one and sometimes two divalent metal cations (Mg^{2+} or Mn^{2+}), a nucleotide binding motif such as a P loop or a Gly-rich sequence, a positively charged residue, usually Lys, and the positively charged N terminus of an α helix. The mechanism of kinase-catalyzed phosphoryl transfer has been the subject of intensive study.[59] About 1.7% of the human genome encode for PKs.[12]

The search for efficient and selective PK inhibitors as possibly the "major drug targets of the 21st century"[60] is one of the largest ongoing efforts in the pharmaceutical industry.[10,13] An example of a successfully developed small-molecule drug is gleevec, which inhibits the Abl (Abelson) PTK of the fusion protein Bcr-Abl and is applied against myeloid leukemia[61] and gastrointestinal stromal tumors.[62] Furthermore, sorafenib has been developed for the treatment of metastatic renal cell carcinomas.[63] A final example of a successful development is the humanized monoclonal antibody herceptin, which is used for the treatment of metastatic breast cancers and binds to the Her2/neu receptor PTK.[64]

4.2. Examples of Phosphate Binding Sites in Protein Kinases

PKs use a great diversity of phosphate binding motifs as illustrated in the following examples. Homoserine kinase (HSK) is a member of the GHMP kinase superfamily and catalyzes the first committed step in the biosynthesis of Thr, the formation of *O*-phospho-L-homoserine from L-homoserine and ATP. The X-ray crystal structures of a number of ternary complexes of HSK with homoserine or Thr (a feedback inhibitor) and the ATP analogue AppNp have been solved at resolutions between 1.8 and 2.0 Å (Figure 11, PDB code: 1H72).[34,65] Both homoserine and AppNp are bound in a 13-Å-deep pocket formed by an α helix and three rather flexible loop structures. The homoserine is located at the lower end of the cavity and forms, among others, a strong salt bridge to Arg235 (2.71 Å and 2.65 Å). In the upper part of the pocket, AppNp is bound at the N terminus of the α helix. The triphosphate adopts an *sc*,*ac*,*sc*,*sc*,*sp* conformation (starting from the ribose moiety; *sc* = *synclinal*, *ac* = *anticlinal*, *sp* = *synperiplanar*) and is bound by five direct and three water-mediated H-bonds. The α-phosphate forms H-bonds in a chelating manner to the backbone NH group of Ser98 (2.47 Å) and to its side-chain OH group (3.02 Å). Gly92 forms a H-bond with one of the α-phosphate O atoms (2.57 Å), whereas a strongly coordinated water molecule

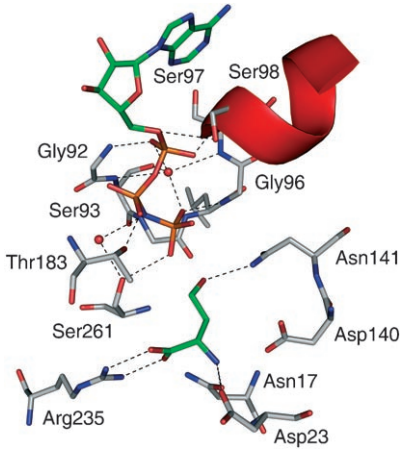


Figure 11. Section of the X-ray crystal structure of homoserine kinase in complex with the ATP analogue AppNp and homoserine (EC 2.7.1.39, PDB code: 1H72, 1.80-Å resolution).[65] Crystallographically localized water: red spheres.

binds both an α-phosphate (2.91 Å) and a γ-phosphate (2.43 Å) O atom. The latter phosphate forms a H-bond with Gly96 (2.55 Å), whereas only the β-phosphate O atoms form one strong H-bond with the side-chain OH group of Thr183 (2.02 Å). A remarkable feature of this binding pocket is the absence of cationic amino acid residues in the triphosphate binding pocket. This example also illustrates how the visual inspection of all structures helped to identify major contributions of (crystallographically localized) water molecules to phosphate binding. In addition, it seems rather common that the β-phosphate undergoes the weakest interactions in a triphosphate complex.

Another example illustrating some of these features, such as the absence of basic amino acids in direct proximity to the bound phosphate, is the X-ray crystal structure of the complex of pyruvate dehydrogenase kinase (PDK3) with L2 (inner lipoyl domains), Mg^{2+}, and ATP (see 6SI in the Supporting Information).[66]

The Src family of PTKs, as well as a number of other proteins involved in intracellular signaling, share the highly conserved SH2 domain. It is responsible for the specific recognition of phosphotyrosine-containing motifs in activated cell-surface receptors. Thus, the SH2 domain is key for signal transduction. Figure 12 shows the X-ray crystal structure of the complex of the Src homology domain of Src kinase and an isostere of an O-phosphorylated YEEI tetrapeptide (PDB code: 1IS0).[67] To probe the topography of the binding pocket, a cyclopropane residue has been introduced to stiffen the backbone of the peptide chain. The ligand binds on the surface of the protein and only the *O*-phosphotyrosine fragment can reach into a shallow cavity formed by the N terminus of an α helix and a loop connecting two β strands. The phosphate is strongly bound through salt bridges with the side chains of Arg175 (2.69 Å and 2.78 Å) and Arg155 (2.90 Å). Further H-bonds are formed with the backbone NH group of Glu178 (2.56 Å) and the side-chain OH group of Thr179 (2.69 Å). A sixth rather weak interaction can be observed between the SH group of Cys185 (3.21 Å) and the phosphorylated Tyr O atom. The prevalence of cationic or polar amino acid residues as well as the shallow pocket on the surface of the present protein stand in sharp contrast to the example of the homoserine kinase complex discussed previously.

Bruton's tyrosine kinase is an important enzyme for the maturation of B cells. In humans, a single point mutation in the enzyme leads to X-linked agammaglobulinemia, a severe immunodeficiency disease. The protein contains a domain that specifically binds phosphatidylinositol-3,4,5-triphosphate. In the X-ray crystal structure of the dimeric complex solved at a resolution of 2.4 Å (Figure 13, PDB code: 1B55),[68] the phosphatidylinositol-3,4,5-triphosphate ligand is bound in a shallow pocket on the surface of the enzyme. Interactions are observed with amino acids of a loop stretching from Arg28–Tyr39, which connects two β strands. The free OH groups (C2 and C6) of the ligand form H-bonds with the side chains of Ser21 (2.81 Å) and Asn24 (2.32 Å).

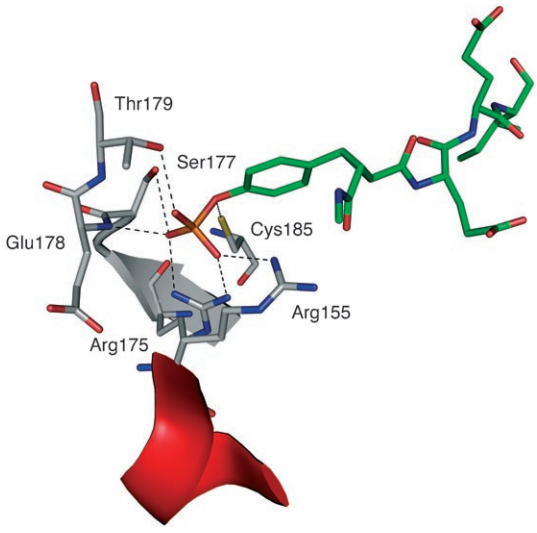
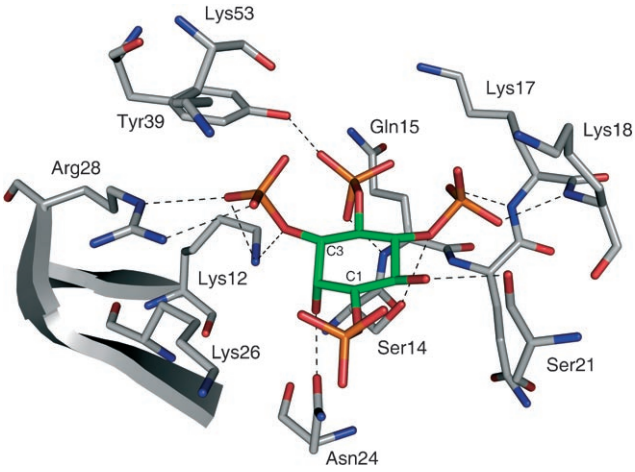


Figure 12. Section of the X-ray crystal structure of the Src kinase in complex with an O-phosphorylated YEEI tetrapeptide (EC 2.7.1.112, PDB code: 1IS0, 1.90-Å resolution).[67]



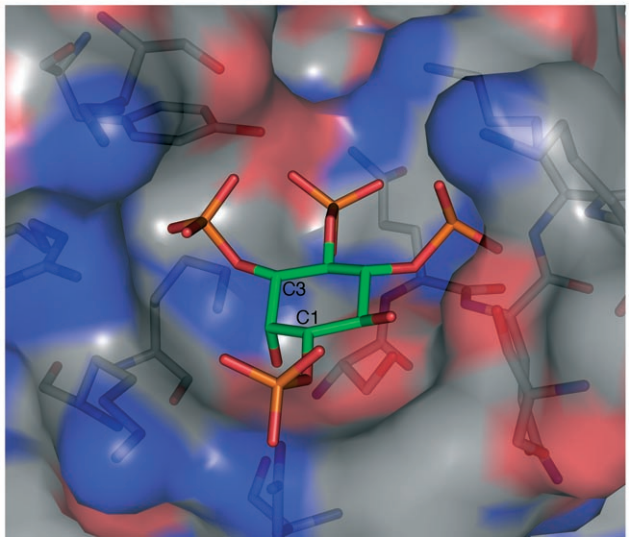


Figure 13. Top: Section of the X-ray crystal structure of Bruton's tyrosine kinase in complex with phosphatidylinositol-3,4,5-triphosphate (EC 2.7.1.112, PDB code: 1B55, 2.40-Å resolution).[68] Bottom: Electrostatic potential showing the electropositive region around C1 to C5 of the ligand at the entrance to the bowl-type binding site.

The C1 phosphoryl group has no direct H-bonding contact to the protein but benefits from the positively polarized environment set up by the proximity of the Lys26 side chain. The phosphate group on C3, on the other hand, is strongly bound by salt bridges to Arg28 (3.08 Å and 3.18 Å) and Lys12 (2.67 Å and 2.96 Å). A rare H-bond between the OH group of Tyr39 and the phosphate group on C4 (2.38 Å) can be found, as well as an interaction with the backbone NH group of Gln15 (2.39 Å). The backbone NH groups of Lys17 (2.73 Å) and Lys18 (2.81 Å) form H-bonds to the C5 phosphate along with an interaction with the side-chain OH group of Ser14 (3.04 Å). To conclude, one can observe that the protein surface at the bottom of the bowl-like binding domain has a rather negative electrostatic potential (Ser14, Ser21, Asn24), whereas the rim of the bowl is formed by positively charged amino acid residues (Lys12, Lys17, Lys18, Lys26, Arg28, Lys53). The rim, however, is not completely positively charged with a gap at the unphosphorylated C6 position. Thus, the protein surface potential perfectly matches the charge density distribution on the phosphatidylinositol-3,4,5-triphosphate ligand, giving rise to selectivity over differentially phosphorylated derivatives.

4.3. Other Phosphate Binding Sites in Kinases

A few additional examples of phosphate recognition by kinases are discussed in the Supporting Information to complete the illustration of the rich structural diversity involved:

a) The X-ray crystal structure of a ternary complex of yeast adenylate kinase, bis(adenyl)-5′-pentaphosphate (Ap_5A), and a Mg^{2+} ion at a resolution of 1.63 Å (7SI in the Supporting Information; PDB code: 2AKY)[69] shows the pentaphosphate bound in a "giant anion hole" featuring the consensus sequence GXXGXGK.[21] Adenylate kinases are ubiquitous enzymes that catalyze the transfer of a phosphoryl group from an ATP molecule to an AMP molecule, producing two molecules of ADP. This process is Mg^{2+} dependent.

b) In the complex of glycerol kinase with ADP, glycerol-3-phosphate, and a Mn^{2+} ion (8SI in the Supporting Information, PDB code: 1GLD),[70] both phosphates are bound to the Mn^{2+} ion in a 15-Å-deep pocket at the interface of the N- and the C-terminal domains.

c) Also in the complex of riboflavin kinase with ADP, flavinmononucleotide (FMN), and Mg^{2+} (9SI in the Supporting Information, PDB code: 1P4M),[71] the two phosphates reach into a deep cavity to coordinate to the metal ion. Riboflavin kinase is an enzyme that catalyzes the phosphorylation of riboflavin (vitamin B_2) to yield FMN.

d) A nice example of ADP in a pocket formed by a loop connecting a β strand to an α helix is seen in the X-ray crystal structure of human deoxycytidine kinase in complex with ADP, Mg^{2+}, and the prodrug gemcitabine (10SI in the Supporting Information, PDB code: 1P62).[72] Deoxycytidine kinase catalyzes the phosphorylation of natural deoxyribonucleosides, such as deoxycitidine, deoxyguanosine, deoxyadenosine, as well as numerous synthetic nucleoside analogues used as prodrugs in antiviral and cancer chemotherapy.

4.4. Protein Phosphatases

PPs are classified by substrate and structure specificity into protein serine–threonine phosphatases (PSTPs), protein tyrosine phosphatases (PTPs), and dual-specificity phosphatases (DSPs). The latter two classes are related and show high sequence homology. The mechanism of dephosphorylation, involving catalytic Asp, Arg, and His residues, is largely understood.[73,74] PPs have only more recently become hot targets in drug-discovery research.[15,17,64,75] Although other phosphatases are under investigation,[76] the PPs are attracting the most attention as potential drug targets.

4.5. Examples of Phosphate Binding Sites in Protein Phosphatases

The protein tyrosine phosphatase PTP1B is responsible for dephosphorylating the phosphotyrosine residues of the insulin receptor kinase IRK, thus negatively regulating the insulin signaling pathway. The structure of PTP1B in complex with a diphosphorylated model peptide mimicking the substrate has been reported at a resolution of 2.4 Å (Figure 14, PDB code: 1G1H).[77] One of the O-phosphorylated Tyr residues is bound in a 7-Å-deep pocket where it is stabilized by the dipole moment at the N terminus of an α helix. The loop connecting this helix to a β sheet forms a strong, pseudo-polyazamacrocycle-type binding site for the phosphate group, forming six H-bonds to the backbone NH group of Ser216 (3.00 Å), Ala217 (3.26 Å), Ile219 (2.96 Å), Gly220 (2.89 Å), Arg221 (2.99 Å), and a salt bridge to the side-chain guanidinum group of Arg221 (2.92 Å and 3.04 Å). The pocket is

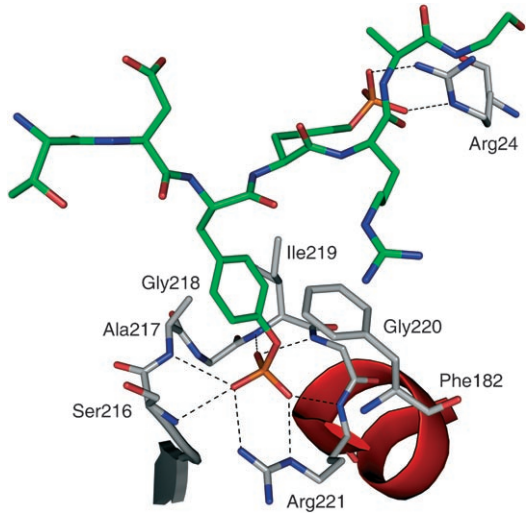


Figure 14. Section of the X-ray crystal structure of protein tyrosine phosphatase PTP1B in complex with a diphosphorylated model peptide (EC 3.1.3.48, PDB code: 1G1H, 2.40-Å resolution).[77]

shielded from solvent interactions by the aromatic ring of Phe 182, which is involved in a π-stacking interaction with the aromatic side chain of the phosphorylated Tyr. The second O-phosphorylated Tyr in the short peptide is only bound on the surface of the protein by the side chain of Arg 24 (3.20 Å and 3.23 Å).

Another X-ray crystal structure of a protein tyrosine phosphatase in complex with *p*-nitrophenyl phosphate is shown in the table-of-contents picture (EC 3.1.3.48, PDB code: 1D1Q, 1.70-Å resolution).[78]

Phosphoserine phosphatase (PSP) belongs to a large class of enzymes that catalyze the phosphoester hydrolysis by utilizing a phosphoaspartate intermediate. PSP is likely to be involved in the regulation of the steady-state concentration of the D-serine level in the brain. The X-ray crystal structure of the binary complex of PSP and O-phosphorylated L-serine has been solved at a resolution of 1.9 Å (Figure 15, PDB code: 1L7P).[79] Upon binding, the protein completely folds around the substrate, covering it efficiently from the solvent. The phosphate group forms H-bonds with three backbone NH groups of Phe 12 (2.81 Å), Asp 13 (3.09 Å), and Gly 100 (2.84 Å). In addition, several side chains from Asp 13 (2.82 Å), Asn 11 (3.25 Å), Ser 99 (2.59 Å), Lys 144 (2.68 Å), and Asn 170 (3.09 Å) form H-bonds with the phosphate O atoms. The carboxylate group of L-serine is bound through a salt bridge with Arg 56 (2.92 Å and 3.15 Å), whereas the ammonium group is stabilized by the side chain of Glu 20 (2.50 Å).

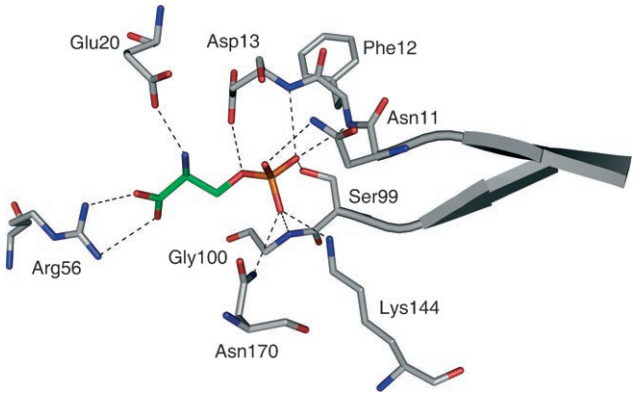


Figure 15. Section of the X-ray crystal structure of phosphoserine phosphatase in complex with O-phosphorylated L-serine (EC 3.1.3.3, PDB code: 1L7P, 1.90-Å resolution).[79]

5. Summary and Conclusions

Despite the large interest in the development of drugs that target phosphate-containing substrates, in particular PKs and PPs, molecular recognition of phosphates at biological active sites had not been comprehensively reviewed. Taking advantage of the abundant structural information contained in the PDB, we now present such a survey, with a focus on the most important interaction between the phosphate and the receptor, namely H-bonding. We first reviewed the known phosphate binding motifs such as the Gly-rich loop and the P loop and established the close analogy to complexation by oligoazamacrocycles: facilitated by the conformational flexibility imparted by the Gly residues in these loops, the phosphate guests organize the receptor site with the loop wrapping around the anion and forming H-bonds with the converging backbone NH residues. This is reminiscent of the principle of ion complexation by flexible receptors: also in this case, the guest organizes its host.

The subsequent statistical analysis yielded quite a number of unexpected results: Among the 3003 considered structures that have each been analyzed individually, the remarkably high number of 2456 entries report phosphate ion binding without the assistance of metal ions. Even more remarkably, a third of the entire data set ("**All**"), 1070 structures, showed phosphate ion complexation without involvement of a metal ion or the presence of basic (protonated) side chains of Arg or Lys residues within the defined H-bonding distance (< 3.2 Å). An analysis of the propensities of amino acids in various classes of phosphate binding enzymes (oxidoreductases, transferases, lyases, and isomerases) led to the emergence of highly characteristic distributions of amino acids used for phosphate binding, which can be viewed as "fingerprints" of the various classes.

The review ends with examples of phosphate binding by PKs and PPs, some of the hottest targets in current drug-discovery research. Although many PK inhibitors, which bind at the ATP site, avoid the triphosphate site, most PP inhibitors bind to the monophosphate site usually through an acidic residue, which mimics the anionic phosphate upon deprotonation. With more than one third of all phosphate binding sites lacking metal ions or basic (protonated) Arg or Lys residues within H-bonding distance, it seems that neutral substituents of small-molecule drugs should also be considered to fill at least these "neutral" binding sites, which often are located rather deeply within the protein. In particular, small heteroalicyclic and heteroaromatic, "drug-like" residues featuring extended H-bond acceptor functionalities in their periphery should be well suited to interact with the convergent H-bond donor groups at the phosphate recognition site. It can be assumed that the Gly-rich loops that frequently shape the binding sites possess sufficient flexibility to wrap around such residues. In a structure-based design approach, we are currently testing this proposal of the binding to the Gly-rich loop of the ATP binding site of IspE with small H-bond-accepting heterocycles (Figure 6). To illustrate this approach, examples of such heterocycles modeled into the phosphate binding site of IspE are shown in 11SI and 12SI in the Supporting Information. Even if a Lys side chain is involved in phosphate binding, neutral ligand moieties could be suitable as Lys side chains can often swing into another position. Only in the case of Arg side chains converging into the phosphate binding site with their charge uncompensated by Asp or Glu located in the vicinity might it be difficult to fill the phosphate recognition site with a neutral residue. In general, it can be said that the nature of phosphate binding sites shows a strong dependence on its location: towards the surface, cationic residues tend to dominate, whereas deep inside, neutral amino acids are key.

Clearly, this analysis shows that structure-based lead development and optimization will benefit from an in-depth, atom-by-atom inspection of phosphate binding sites, such as is illustrated in this review, if all options for innovative phosphate replacement are to be exploited.

Bar graphs showing the distributions of the amino acids involved in phosphate recognition in the entire set of structures considered, in comparison to various subsets, within different classes of enzymes, and selected examples for phosphate binding sites from the RCSB Protein Data Bank for this article can be found in the Supporting Information.

This work was supported by the ETH Research Council, Hoffmann-La Roche Ltd (Basel), and Chugai Pharmaceuticals. We thank Jörg Klein, Christian Kramer, and Fabian Weibel for their valuable contributions to the PDB searches. Much stimulation for this review has come from numerous discussions at Roche and Chugai, which are gratefully acknowledged. We thank Dr. W. Bernd Schweizer (ETH Zurich) and the Cambridge-based Relibase team for assistance in the Relibase searches.

Received: August 21, 2006

[1] E. A. Meyer, R. K. Castellano, F. Diederich, *Angew. Chem.* 2003, 115, 1244 – 1287; *Angew. Chem. Int. Ed.* **2003**, *42*, 1210 – 1250.
[2] R. Paulini, K. Müller, F. Diederich, *Angew. Chem.* **2005**, *117*, 1820 – 1839; *Angew. Chem. Int. Ed.* **2005**, *44*, 1788 – 1805.
[3] a) K. Schärer, M. Morgenthaler, R. Paulini, U. Obst-Sander, D. W. Banner, D. Schlatter, J. Benz, M. Stihle, F. Diederich, *Angew. Chem.* **2005**, *117*, 4474 – 4479; *Angew. Chem. Int. Ed.* **2005**, *44*, 4400 – 4404; b) J. C. Ma, D. A. Dougherty, *Chem. Rev.* **1997**, *97*, 1303 – 1324.
[4] C. M. Crane, J. Kaiser, N. L. Ramsden, S. Lauw, F. Rohdich, W. Eisenreich, W. N. Hunter, A. Bacher, F. Diederich, *Angew. Chem.* **2006**, *118*, 1082 – 1087; *Angew. Chem. Int. Ed.* **2006**, *45*, 1069 – 1074.
[5] a) F. Rohdich, S. Hecht, A. Bacher, W. Eisenreich, *Pure Appl. Chem.* **2003**, *75*, 393 – 405; b) M. Rohmer, M. Knani, P. Simonin, B. Sutter, H. Sahm, *Biochem. J.* **1993**, *295*, 517 – 524; c) M. K. Schwarz, PhD Dissertation, ETH Zürich, No. 10951, **1994**; d) S. T. J. Broers, PhD Dissertation, ETH Zürich, No. 10978, **1994**.
[6] J. Wiesner, R. Ortmann, H. Jomaa, M. Schlitzer, *Angew. Chem.* **2003**, *115*, 5432 – 5451; *Angew. Chem. Int. Ed.* **2003**, *42*, 5274 – 5293, zit. Lit.
[7] J. H. Martinez-Liarte, A. Iriarte, M. Martinez-Carrion, *Biochemistry* **1992**, *31*, 2712 – 2719.
[8] Y. G. Cheng, N. D. Chasteen, *Biochemistry* **1991**, *30*, 2947 – 2953.
[9] T. Hunter, *Cell* **1995**, *80*, 225 – 236.
[10] T. Hunter, *Cell* **2000**, *100*, 113 – 127.
[11] S. K. Hanks, T. Hunter, *FASEB J.* **1995**, *9*, 576 – 596.
[12] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, *Science* **2002**, *298*, 1912 – 1934.
[13] Z.-Y. Zhang, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 209 – 234.
[14] a) A. J. Bridges, *Chem. Rev.* **2001**, *101*, 2541 – 2571; b) M. E. M. Noble, J. A. Endicott, L. N. Johnson, *Science* **2004**, *303*, 1800 – 1805.
[15] K. Grosios, P. Traxler, *Drugs Future* **2003**, *28*, 679 – 697.
[16] a) R. H. van Huijsduijnen, A. Bombrun, D. Swinnen, *Drug Discovery Today* **2002**, *7*, 1013 – 1019; b) R. E. Honkanen, T. Golden, *Curr. Med. Chem.* **2002**, *9*, 2055 – 2075; c) G. Liu, *Curr. Med. Chem.* **2003**, *10*, 1407 – 1421.
[17] L. Bialy, H. Waldmann, *Angew. Chem.* **2005**, *117*, 3880 – 3906; *Angew. Chem. Int. Ed.* **2005**, *44*, 3814 – 3839.
[18] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235 – 242.
[19] D. Dreusicke, G. E. Schulz, *FEBS Lett.* **1986**, *208*, 301 – 304.
[20] M. Saraste, P. R. Sibbald, A. Wittinghofer, *Trends Biochem. Sci.* **1990**, *15*, 430 – 434.
[21] G. E. Schulz, *Curr. Opin. Struct. Biol.* **1992**, *2*, 61 – 67.
[22] For reviews on the complexation of anions, and in particular phosphates, by synthetic receptors, see: a) J.-M. Lehn, *Angew. Chem.* **1988**, *100*, 91 – 116; *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 89 – 112; b) M. P. Mertes, K. Bowman Mertes, *Acc. Chem. Res.* **1990**, *23*, 413 – 418; c) F. P. Schmidtchen, M. Berger, *Chem. Rev.* **1997**, *97*, 1609 – 1646; d) P. D. Beer, P. A. Gale, *Angew. Chem.* **2001**, *113*, 502 – 532; *Angew. Chem. Int. Ed.* **2001**, *40*, 486 – 516; e) C. A. Iloudis, J. W. Steed, *J. Supramol. Chem.* **2001**, *1*, 165 – 187; f) J. M. Llinares, D. Powell, K. Bowman-James, *Coord. Chem. Rev.* **2003**, *240*, 57 – 75; g) K. Bowman-James, *Acc. Chem. Res.* **2005**, *38*, 671 – 678; h) J. L. Sessler, P. A. Gale, W.-S. Cho, *Anion Receptor Chemistry*, Royal Society of Chemistry, Cambridge, **2006**.
[23] M. G. Rossmann, D. Moras, K. W. Olsen, *Nature* **1974**, *250*, 194 – 199.
[24] W. Möller, R. Amons, *FEBS Lett.* **1985**, *186*, 1 – 7.
[25] T. W. Traut, *Eur. J. Biochem.* **1994**, *222*, 9 – 19.
[26] a) W. G. J. Hol, P. T. Vanduijnen, H. J. C. Berendsen, *Nature* **1978**, *273*, 443 – 446; b) R. K. Wierenga, M. C. H. De Maeyer, W. G. J. Hol, *Biochemistry* **1985**, *24*, 1346 – 1357; c) B. E. Bernstein, P. A. M. Michels, W. G. J. Hol, *Nature* **1997**, *385*, 275 – 278.
[27] R. R. Copley, G. J. Barton, *J. Mol. Biol.* **1994**, *242*, 321 – 329.
[28] D. Bossemeyer, *Trends Biochem. Sci.* **1994**, *19*, 201 – 205.
[29] K. Kinoshita, K. Sadanami, A. Kidera, N. Go, *Protein Eng.* **1999**, *12*, 11 – 14.
[30] J. E. Walker, M. Saraste, M. J. Runswick, N. J. Gay, *EMBO J.* **1982**, *1*, 945 – 951.
[31] a) E. J. Milner-White, M. J. Russell, *Origins Life Evol. Biosphere* **2005**, *35*, 19 – 27; b) J. D. Watson, E. J. Milner-White, *J. Mol. Biol.* **2002**, *315*, 171 – 182.
[32] J. Feuerstein, R. S. Goody, M. R. Webb, *J. Biol. Chem.* **1989**, *264*, 6188 – 6190.
[33] E. F. Pai, U. Krengel, G. A. Petsko, R. S. Goody, W. Kabsch, A. Wittinghofer, *EMBO J.* **1990**, *9*, 2351 – 2359.
[34] T. Zhou, M. Daugherty, N. V. Grishin, A. L. Ostermann, H. Zhang, *Structure* **2000**, *8*, 1247 – 1257.
[35] S. K. Hanks, A. M. Quinn, T. Hunter, *Science* **1988**, *241*, 42 – 52.
[36] A. I. Denesyuk, K. A. Denessiouk, T. Korpela, M. S. Johnson, *J. Mol. Biol.* **2002**, *316*, 155 – 172.
[37] L. Holm, C. Sander, *Science* **1996**, *273*, 595 – 602.
[38] S. Rhee, M. M. Silva, C. C. Hyde, P. H Rogers, C. M. Metzler, D. E. Metzler, A. Arnone, *J. Biol. Chem.* **1997**, *272*, 17293 – 17302.
[39] K. A. Denessiouk, M. S. Johnson, A. I. Denesyuk, *J. Mol. Biol.* **2005**, *345*, 611 – 629.
[40] a) M. W. Hosseini, J.-M. Lehn, *Helv. Chim. Acta* **1987**, *70*, 1312 – 1319; b) Q. Lu, R. J. Motekaitis, J. J. Reibenspies, A. E. Martell, *Inorg. Chem.* **1995**, *34*, 4958 – 4964; c) A. C. Warden, M. Warren, M. T. W. Hearn, L. Spiccia, *Inorg. Chem.* **2004**, *43*, 6936 – 6943.
[41] B. Dietrich, T. M. Fyles, J.-M. Lehn, L. G. Pease, D. L. Fyles, *J. Chem. Soc. Chem. Commun.* **1978**, 934 – 936.
[42] F. P. Schmidtchen, *Tetrahedron Lett.* **1989**, *30*, 4493 – 4496.
[43] J. L. Sessler, E. Katayev, G. D. Pantos, Y. A. Ustynyuk, *Chem. Commun.* **2004**, 1276 – 1277.
[44] P. Chakrabarti, *J. Mol. Biol.* **1993**, *234*, 463 – 482.

[45] a) J. W. Pflugrath, F. A. Quiocho, *Nature* **1985**, *314*, 257 – 260; b) H. Luecke, F. A Quiocho, *Nature* **1990**, *347*, 402 – 406; c) J. J. He, F. A. Quiocho, *Science* ***1991***, *251*, 1479 – 1481.
[46] a) K. L. Longenecker, P. J. Roach, T. D. Hurley, *J. Mol. Biol.* **1996**, *257*, 618 – 631; b) N. Kobayashi, N. Go, *Eur. Biophys. J.* **1997**, *26*, 135 – 144; c) K. A. Denessiouk, J. V. Lehtonen, M. S. Johnson, *Protein Sci.* **1998**, *7*, 1768 – 1771.
[47] PDB searches were performed by using Relibase V. 1.3.2. (August 2005)[48] and the PDB[18] update of February 14, 2006; copyright M. Hendlich 1994 – 1999 and Cambridge Crystallographic Data Centre 1999 – 2005, Union Road, Cambridge CB2 1EZ, United Kingdom.
[48] a) M. Hendlich, A. Bergner, J. Günther, G. Klebe, *J. Mol. Biol.* **2003**, *326*, 607 – 620; b) J. Günther, A. Bergner, M. Hendlich, G. Klebe, *J. Mol. Biol.* **2003**, *326*, 621 – 636.
[49] We thank S. Robinson from Relibase for assistance in determining the amino acid propensities of the entire PDB.
[50] P. R. Rablen, J. W. Lockman, W. L. Jorgensen, *J. Phys. Chem. A* **1998**, *102*, 3782 – 3797.
[51] L. Miallau, M. S. Alphey, L. E. Kemp, G. A. Leonard, S. M. McSweeney, S. Hecht, A. Bacher, W. Eisenreich, F. Rohdich, W. N. Hunter, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9173 – 9178.
[52] W. Shi, N. R. Munagala, C. C. Wang, C. M. Li, P. C. Tyler, R. H. Furneaux, C. Grubmeyer, V. L. Schramm, S. C. Almo, *Biochemistry* **2000**, *39*, 6781 – 6790.
[53] S. Bauer, K. Kemter, A. Bacher, R. Huber, M. Fischer, S. Steinbacher, *J. Mol. Biol.* **2003**, *326*, 1463 – 1473.
[54] E. G. Krebs, J. A. Beavo, *Annu. Rev. Biochem.* **1979**, *48*, 923 – 959.
[55] W. Vogel, R. Lammers, J. Hunag, A. Ullrich, *Science* **1993**, *259*, 1611 – 1614.
[56] a) S. Cheek, H. Zhang, K. Ginalski, N. V. Grishin, *BMC Struct. Biol.* **2005**, *5*, 6; b) S. Cheek, H. Zhang, N. V. Grishin, *J. Mol. Biol.* **2002**, *320*, 855 – 881.
[57] T. Naumann, H. Matter, *J. Med. Chem.* **2002**, *45*, 2366 – 2378.
[58] a) J. R. Knowles, *Annu. Rev. Biochem.* **1980**, *49*, 877 – 919; b) W. W. Cleland, A. C. Hengge, *FASEB J.* **1995**, *9*, 1585 – 1594.
[59] a) A. S. Mildvan, *Proteins Struct. Funct. Genet.* **1997**, *29*, 401 – 416; b) Y.-W. Xu, S. Morera, J. Janin, J. Cherfils, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 3579 – 3583; c) I. Schlichting, J. Reinstein, *Biochemistry* **1997**, *36*, 9290 – 9296; d) S. D. Lahiri, G. Zhang, D. Dunaway-Mariano, K. N. Allen, *Science* **2003**, *299*, 2067 – 2071; e) J. Knowles, *Science* **2003**, *299*, 2002 – 2003.
[60] P. Cohen, *Nat. Rev. Drug Discovery* **2002**, *1*, 309 – 315.
[61] a) B. J. Druker, S. Tamura, E. Buchdunger, S. Ohno, G. M. Segal, S. Fanning, J. Zimmermann, N. B. Lydon, *Nat. Med.* **1996**, *2*, 561 – 566; b) B. J. Druker, N. B. Lydon, *J. Clin. Invest.* **2000**, *105*, 3 – 7; c) T. Schindler, W. Bornmann, P. Pellicena, W. T. Miller, B. Clarkson, J. Kuriyan, *Science* **2000**, *289*, 1938 – 1942; d) B. Okram, A. Nagle, F. J. Adrián, C. Lee, P. Ren, X. Wang, T. Sim, Y. Xie, X. Wang, G. Xia, G. Spraggon, M. Warmuth, Y. Liu, N. S. Gray, *Chem. Biol.* **2006**, *13*, 779 – 786; e) S. W. Cowman-Jacob, V. Guez, G. Fendrich, J. D. Griffin, D. Fabbro, P. Furet, J. Liebetanz, J. Mestan, P. W. Manley, *Mini-Rev. Med. Chem.* **2004**, *4*, 285 – 299.
[62] A. T. van Oosterom, I. Judson, J. Verweij, S. Stroobants, E. D. di Paola, S. Dimitrijevic, M. Martens, A. Webb, R. Sciot, M. Van Glabbeke, S. Silberman, O. S. Nielsen, E. O. R. T. C. So, *Lancet* **2001**, *358*, 1421 – 1423.
[63] a) P. T. C. Wan, M. J. Garnett, S. M. Roe, S. Lee, D. Niculescu-Duvaz, V. M. Good, C. M. Jones, C. J. Marshall, C. J. Springer, B. Barford, R. Marais, *Cell* **2004**, *116*, 855 – 867; b) T. Ahmad, T. Eisen, *Clin. Cancer Res.* **2004**, *10*, 6388s – 6392s.
[64] a) H.-S. Cho, K. Mason, K. X. Ramyar, A. M. Stanley, S. B. Gabelli, D. W. Denney, Jr., D. J. Leahy, *Nature* **2003**, *421*, 756 – 760; b) L. K. Shawver, D. Slamon, A. Ullrich, *Cancer Cell* **2002**, *1*, 117 – 123; c) D. J. Slamon, B. Leyland-Jones, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, T. Fleming, W. Eiermann, J. Wolter, M. Pegram, J. Baselga, L. Norton, *N. Engl. J. Med.* **2001**, *344*, 783 – 792; d) D. J. Slamon, G. M. Clark, S. G. Wong, W. J. Levin, A. Ulrich, W. L. McGuire, *Science* **1987**, *235*, 177 – 182.
[65] S. S. Krishna, T. Zhou, M. Daugherty, A. Osterman, H. Zhang, *Biochemistry* **2001**, *40*, 10810 – 10818.
[66] a) M. Kato, J. L. Chuang, S. C. Tso, R. M. Wynn, D. T. Chuang, *EMBO J.* **2005**, *24*, 1763 – 1774; b) R. Dutta, M. Inouye, *Trends Biochem. Sci.* **2000**, *25*, 24 – 28.
[67] J. P. Davidson, O. Lubman, T. Rose, G. Waksman, S. F. Martin, *J. Am. Chem. Soc.* **2002**, *124*, 205 – 215.
[68] E. Baraldi, K. Djinovic Carugo, M. Hyvönen, P. Lo Surdo, A. M. Riley, B. V. L. Potter, R. O'Brien, J. E. Ladbury, M. Saraste, *Structure* **1999**, *7*, 449 – 460.
[69] U. Abele, G. E. Schulz, *Protein Sci.* **1995**, *4*, 1262 – 1271.
[70] J. H. Hurley, H. R. Faber, D. Worthylake, N. D. Meadow, S. Roseman, D. W. Pettigrew, S. J. Remington, *Science* **1993**, *259*, 673 – 677.
[71] S. Karthikeyan, Q. Zhou, F. Mseeh, N. V. Grishin, A. L. Osterman, H. Zhang, *Structure* **2003**, *11*, 265 – 273.
[72] E. Sabini, S. Ort, C. Monnerjahn, M. Konrad, A. Lavie, *Nat. Struct. Biol.* **2003**, *10*, 513 – 519.
[73] For reviews, see: a) K. Hinterding, D. Alonso-Diaz, H. Waldmann, *Angew. Chem.* **1998**, *110*, 716 – 780; *Angew. Chem. Int. Ed.* **1998**, *37*, 688 – 749; b) D. Barford, *Trends Biochem. Sci.* **1996**, *21*, 407 – 412; c) E. B. Fauman, M. A. Saper, *Trends Biochem. Sci.* **1996**, *21*, 413 – 417.
[74] a) H. L. Schubert, E. B. Fauman, J. A. Stuckey, J. E. Dixon, M. A. Saper, *Protein Sci.* **1995**, *4*, 1904 – 1913; b) K. L. Guan, J. E. Dixon, *J. Biol. Chem.* **1991**, *266*, 17026 – 17030; c) J. M. Denu, J. E. Dixon, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5910 – 5914.
[75] M. Elchebly, P. Payette, E. Michaliszyn, W. Cromlish, S. Collins, A. L. Loy, D. Normandin, A. Cheng, J. Himms-Hagen, C.-C. Chan, C. Ramachandran, M. J. Gresser, M. L. Tremblay, B. P. Kennedy, *Science* **1999**, *283*, 1544 – 1548.
[76] For a recent study, see: M. J. Begley, G. S. Taylor, M. A. Brock, P. Ghosh, V. L. Woods, J. E. Dixon, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 927 – 932.
[77] A. Salmeen, J. N. Andersen, M. P. Myers, N. K. Tonks, D. Barford, *Mol. Cell* **2000**, *6*, 1401 – 1412.
[78] S. Wang, L. Tabernero, M. Zhang, E. Harms, R. L. Van Etten, C. V. Stauffacher, *Biochemistry* **2000**, *39*, 1903 – 1914.
[79] W. Wang, H. S. Cho, R. Kim, J. Jancarik, H. Yokota, H. H. Nguyen, I. V. Grigoriev, D. E. Wemmer, S. H. Kim, *J. Mol. Biol.* **2002**, *319*, 421 – 431.

