

Review

Why nature chose phosphate to modify proteins

Tony Hunter*

Salk Institute for Biological Studies, La Jolla, CA 92037, USA

The advantageous chemical properties of the phosphate ester linkage were exploited early in evolution to generate the phosphate diester linkages that join neighbouring bases in RNA and DNA (Westheimer 1987 *Science* **235**, 1173–1178). Following the fixation of the genetic code, another use for phosphate ester modification was found, namely reversible phosphorylation of the three hydroxyamino acids, serine, threonine and tyrosine, in proteins. During the course of evolution, phosphorylation emerged as one of the most prominent types of post-translational modification, because of its versatility and ready reversibility. Phosphoamino acids generated by protein phosphorylation act as *new* chemical entities that do not resemble any natural amino acid, and thereby provide a means of diversifying the chemical nature of protein surfaces. A protein-linked phosphate group can form hydrogen bonds or salt bridges either intra- or intermolecularly, creating stronger hydrogen bonds with arginine than either aspartate or glutamate. The unique size of the ionic shell and charge properties of covalently attached phosphate allow specific and inducible recognition of phosphoproteins by phosphospecific-binding domains in other proteins, thus promoting inducible protein–protein interaction. In this manner, phosphorylation serves as a switch that allows signal transduction networks to transmit signals in response to extracellular stimuli.

Keywords: phosphate; ester; protein; phosphorylation; kinase; evolution

1. PHOSPHATE ESTERS HAVE ADVANTAGEOUS CHEMICAL PROPERTIES FOR THE EVOLUTION OF LIFE

Phosphate-containing molecules are essential constituents of all living cells. What then are the special properties of phosphate that led to its selection as a key building block during the evolution of self-replicating life forms? Phosphorus is a Group 15 element, and therefore has five electrons in its outer shell. By donating its electrons, phosphorus can form five covalent bonds; for example, by combining with four oxygen atoms, phosphorus forms orthophosphate, PO_4^{3-} . Phosphate has three pKas (2.2, 7.2 (5.8 as an ester) and 12.4) and is highly soluble in water, forming a large hydrated ionic shell. Phosphate is chemically versatile and can form mono-, di- and tri-esters with alkyl and aryl hydroxyl groups, as well as acid anhydrides. In addition, phosphorus can form P–N (phosphoramidate), P–S (phosphorothioate) and P–C (phosphonate) linkages.

Phosphate salts are very abundant on the Earth, and, owing to their water solubility, were readily available during the evolution of life. The ability of phosphate to form esters and anhydrides that are stable at ambient temperatures in water made it ideal for the generation of

biological molecules. Phosphate esters and anhydrides predominate in living organisms, but phosphoramidates, phosphorothioates and phosphonates are all found in nature. Phosphate esters are readily formed under physiological conditions using adenosine triphosphate (ATP), a phosphate anhydride, as a phosphate donor and an enzyme catalyst. Once formed, phosphate esters are chemically stable in aqueous solution at physiological pH, and yet they are easily hydrolysed by an appropriate enzyme catalyst, thus regenerating the original OH-group and free phosphate. In cells, important phosphate esters include nucleic acids and phosphoproteins. Examples of biological phosphate anhydrides are ATP and PAPS (3'-phosphoadenosine-5'-phosphosulphate). Although phosphate is rarely used as a leaving group in chemical syntheses, it is reactive at physiological temperatures in the presence of appropriate enzymatic catalysts. Because of its three pKas, phosphate is dianionic at pH 7, the physiological intracellular pH. Thus, phosphate monoester groups have one full negative charge and a second negative charge, which can be a partial or full charge depending on the chemical context; phosphate diesters retain a full negative charge. Importantly, when two nucleosides are linked by a phosphate diester, the phosphate is still fully ionized ($\text{pK}_a < 1$), providing a negative charge to retain nucleic acids within phospholipid membranes, and also protecting the ester linkage from hydrolysis through an attack by anionic nucleophiles, such as OH^- . Any

*hunter@salk.edu

One contribution of 13 to a Theme Issue 'The evolution of protein phosphorylation'.

molecule with a phosphate ester, including phosphorylated proteins, has these same attributes. The enormous stability of phosphate esters in water at pH 7 (phosphate monoesters are estimated to have a half-life of 10^{12} years at 25°C [1]) allows the formation of very long polynucleotides that are remarkably stable despite the large number of phosphate ester bonds, an attribute that is essential for long-term storage of genetic information.

2. THE ENERGETICS OF PHOSPHORYLATION AND DEPHOSPHORYLATION REACTIONS

Prebiotic generation of ATP and other nucleoside triphosphates was the key to the formation of polynucleotides, and the encoding of genetic information. The selection of ATP as the major energy storage compound also meant the ready availability of activated phosphate groups for transfer to other molecules. Because of its energy storage function, ATP is very abundant in cells with concentrations typically ranging from 2 to 4 mM. ATP is highly soluble and relatively stable in water at physiological pH and temperatures, and yet stores significant chemical energy in both its α - β and β - γ phosphate anhydride bonds, each having free energies of hydrolysis of approximately 8 – 12 kcal mol⁻¹, depending on the ionic conditions. Both the α - β and β - γ anhydride bonds can be used to drive important biological reactions. However, such reactions are in principle reversible in many situations, because alkyl and aryl hydroxy phosphate esters also have a relatively high free energy of hydrolysis (approx. 8 – 10 kcal mol⁻¹). Reactions involving cleavage of the α - β anhydride bond, such as polynucleotide synthesis, which generate pyrophosphate, are rendered irreversible through the abundant pyrophosphatase activity present in most cells. Reactions involving cleavage of the β - γ anhydride bond, such as protein phosphorylation, are rendered irreversible largely because cells maintain a very high ATP/adenosine diphosphate (ADP) ratio.

The energetics of phosphorylation are relatively well balanced, because the energy of the phosphate ester linkage is similar to that of the ATP β - γ anhydride bond. In fact, protein kinase equilibrium constants for the (protein + ATP \rightarrow P.protein + ADP) reaction range from 2 to 50, and many protein kinases will readily work in reverse to dephosphorylate a phosphoprotein in the presence of ADP, generating ATP [2–4]. Thus, in the cell, phosphorylation is dependent on the high ATP/ADP ratio, which prevents the back reaction. Although phosphate esters are rather stable chemically, they can be readily hydrolysed by an appropriate enzyme catalyst under physiological conditions. The relatively high energy of phosphate monoesters ensures that once they are hydrolysed, they cannot be re-formed through phosphorolysis by reaction with free phosphate, thus rendering dephosphorylation irreversible.

3. THE EVOLUTION OF PROTEIN PHOSPHORYLATION

The advantageous chemical properties of the phosphate ester linkage were exploited early in evolution, as exemplified by the phosphate diester linkage that joins neighbouring bases in RNA and DNA [5]. Once the

genetic code for the 20 common amino acids was fixed, a second possible use for phosphate ester modification emerged, namely phosphorylation of serine (Ser), threonine (Thr) and tyrosine (Tyr) residues in proteins as a regulatory mechanism. During evolution, phosphorylation became one of the most prominent types of post-translational modification (PTM) because of its versatility and ready reversibility. Protein phosphorylation, which can occur on nine out of the 20 amino acids in proteins, is readily catalysed by protein kinases under physiological conditions, using ATP as a phosphate donor (guanosine triphosphate (GTP) and phosphoenol pyruvate (PEP) can also serve as phosphate donors). Phosphorylation is easily reversed through enzyme-catalysed hydrolysis by protein phosphatases, which accelerate the rate of reaction many thousand-fold.

Phosphorylation of proteins is one of the most common PTMs in eukaryotes. Phosphate esters of Ser, Thr and Tyr predominate in eukaryotes, but phosphorylation of six other amino acids is chemically feasible (arginine (Arg), lysine (Lys), histidine (His), cysteine (Cys), aspartate (Asp) and glutamine (Glu)), and in many cases is known to occur. Phosphorylation of His, Lys and Arg to generate ‘high energy’ phosphoramidate bonds may have been important under prebiotic conditions. Indeed, Arg phosphate is used as an energy storage compound in plants. Phosphohistidine (P.His) is chemically unstable and has a relatively short half-life in aqueous solution at pH 7, but His phosphorylation is used as a regulatory mechanism in higher eukaryotes [6]. In addition, P.His is used as an intermediate in phosphate transfer to proteins by a large number of prokaryotic two-component signalling systems in which a catalytic biosensor protein phosphorylates a response regulator protein on an Asp, thus transmitting a signal, usually transducing an extracellular input at the membrane into activation of a transcription factor. In these systems, the high energy of the P.His phosphoramidate linkage generated on the biosensor in response to signal input drives the coupling of the phosphate to the β -COOH group of Asp, forming a mixed anhydride linkage.

4. WHY IS PROTEIN PHOSPHORYLATION SO IMPORTANT?

The critical feature of phosphoamino acids in proteins is that they act as *new* chemical entities that do not resemble any natural amino acid, and thereby provide a means of diversifying the chemical nature of protein surfaces. In particular, the phosphate group, with its large hydrated shell and its negative charge greater than 1, is chemically quite distinct from the only negatively charged amino acids, Asp and Glu, whose carboxyl side chains only have a single negative charge and a smaller hydrated shell than phosphate. For this reason, it has been suggested that a vicinal pair of Asp or Glu residues might serve as a better phosphomimetic mutation than a single Asp or Glu, since this would generate a local double negative charge [7].

A protein-linked phosphate group can form hydrogen bonds or salt bridges either intra- or intermolecularly. In particular, the phosphate group is well suited for interacting with the guanidino group of Arg, which has

a rigid, planar structure that can make directed hydrogen bonds to the doubly charged phosphate group at physiological pH. Because of the higher density of negative charge and the larger hydrated shell, phosphoamino acids form stronger and more stable hydrogen bonds and salt bridges than do Asp or Glu with Arg [8]. In this manner, a single phosphate is able to exert either intra- or intermolecular effects. Protein phosphates can act sterically or ionically to regulate function or the interaction of another protein or small molecule, or more commonly to elicit a conformational change within a protein monomer or an allosteric transition within a protein multimer. The unique size and charge properties of covalently attached phosphate also allow specific and inducible recognition of phosphoproteins by phosphospecific-binding domains in other proteins, thus promoting inducible protein-protein interaction. Indeed, the ability of a phosphate within a specific sequence context to generate a phospho-dependent-binding site for another protein is arguably the most important function of protein phosphorylation. Phosphorylation-dependent protein interactions are crucial for transducing signals intracellularly, but phosphorylation can also cause a change in the subcellular location of a protein or create a phosphodegron, leading to ubiquitin-dependent protein degradation.

5. WHY WAS PHOSPHATE PREFERRED OVER OTHER MOLECULES FOR PROTEIN MODIFICATION?

Several elements in the vicinity of phosphorus in the periodic table form polyoxyanions. Sulphate (sulphur is in Group 16 next to phosphorus) can also form aryl and aromatic hydroxy esters, which have a high free energy of hydrolysis, and there are examples of sulphate esters in biology, including sulphated Tyr in secreted proteins. However, the energetics of sulphation are not so favourable, and sulphate (SO_4^{2-}) has only two ionizable oxygens, both of which have pKas below pH 2, meaning that sulphate monoesters always retain a single negative charge. Moreover, alkyl sulphate diesters not only lack a negative charge, but also are unstable and have the undesirable property of alkylating biological molecules (e.g. methyl methanesulphonate). Orthosilicate (silicon is in Group 14 next to phosphorus) is also very abundant on Earth, but its lowest pKa is 9.5, and its esters are extremely unstable in aqueous solution.

From a chemical perspective, arsenate is the most plausible alternative to phosphate. Arsenic like phosphorus is in Group 15, and arsenate (V) (AsO_4^{3-}) has ester-forming properties and pKas (2.1, 6.9 and 11.5) similar to those of phosphate. Like phosphate, arsenate can form mono-, di- and tri-esters. However, arsenate triesters and diesters are very unstable in aqueous solution (the half-life of an arsenate triester is less than 0.02 s in water at pH 7 and room temperature, and diesters are even less stable) [9]. Likewise, one would expect nucleic acid precursors, such as adenosine triarsenate, to be very labile. In addition, arsenate has undesirable chemical properties, including reaction with cysteine thiol groups; As(V) is readily reduced by free thiols to generate As(III) trithiolates and oxidized dithiols, which is the basis for arsenical toxicity. Finally,

As–O bond lengths are approximately 10 per cent longer than P–O bond lengths, and as a consequence the radius of AsO_4^{3-} is approximately 10 per cent larger than that of PO_4^{3-} . In addition, the partial negative charge on the P–O oxygen atoms is approximately 6 per cent greater than for As–O. These chemical differences would alter the structure and properties of macromolecules using arsenate diesters as linkages.

For these reasons, a recent report claiming that a halobacterium isolated from Mono Lake in California, USA is able to use arsenate instead of phosphate was surprising [10]. It is claimed that arsenate is incorporated into DNA, RNA, lipid and protein (80% of total) in this organism based on labelling with radioactive arsenate and various types of physical analysis. However, the DNA analysis showed that there was still phosphate in the DNA from cells isolated from cells grown in arsenate medium without added phosphate, presumably derived from contaminant phosphate in chemicals used in the growth medium. No direct evidence was presented that the DNA has arsenate diester bonds, or that proteins are arsenylated on Ser, Thr and Tyr. Further analysis of this organism is required to determine the extent to which arsenate can substitute for phosphate to sustain life, and modify proteins through esterification [9].

6. CODA

The phosphate group has special properties that can be exploited to regulate critical biological functions when attached to proteins. The phosphate group serves as a switch to promote inducible protein-protein interactions, which allows signal transduction networks to transmit transient signals in response to extracellular stimuli. Phosphate has almost ideal chemical properties for the formation of biological polymers and, in particular, for the modification of proteins. Life as we know it could not have evolved without phosphate.

I am grateful to many colleagues in the protein phosphorylation field for stimulating discussions on this topic, and to Frank Westheimer's thought-provoking article on why nature chose phosphates published in *Science* 25 years ago.

REFERENCES

- 1 Lad, C., Williams, N. H. & Wolfenden, R. 2003 The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proc. Natl Acad. Sci. USA* **100**, 5607–5610. (doi:10.1073/pnas.0631607100)
- 2 Rosen, O. M. & Erlichman, J. 1975 Reversible autophosphorylation of a cyclic 3':5'-AMP-dependent protein kinase from bovine cardiac muscle. *J. Biol. Chem.* **250**, 7788–7794.
- 3 Fukami, Y. & Lipmann, F. 1983 Reversal of Rous sarcoma-specific immunoglobulin phosphorylation on tyrosine (ADP as phosphate acceptor) catalyzed by the src gene kinase. *Proc. Natl Acad. Sci. USA* **80**, 1872–1876. (doi:10.1073/pnas.80.7.1872)
- 4 Kole, H. K., Abdel-Ghany, M. & Racker, E. 1988 Specific dephosphorylation of phosphoproteins by protein-serine and -tyrosine kinases. *Proc. Natl*

- Acad. Sci. USA* **85**, 5849–5853. (doi:10.1073/pnas.85.16.5849)
- 5 Westheimer, F. H. 1987 Why nature chose phosphates. *Science* **235**, 1173–1178. (doi:10.1126/science.2434996)
- 6 Kee, J. M. & Muir, T. W. 2012 Chasing phosphohistidine, an elusive sibling in the phosphoamino acid family. *ACS Chem. Biol.* **7**, 44–51. (doi:10.1021/cb200445w)
- 7 Pearlman, S. M., Serber, Z. & Ferrell Jr, J. E. 2011 A mechanism for the evolution of phosphorylation sites. *Cell* **147**, 934–946. (doi:10.1016/j.cell.2011.08.052)
- 8 Mandell, D. J., Chorny, I., Groban, E. S., Wong, S. E., Levine, E., Rapp, C. S. & Jacobson, M. P. 2007 Strengths of hydrogen bonds involving phosphorylated amino acid side chains. *J. Am. Chem. Soc.* **129**, 820–827. (doi:10.1021/ja063019w)
- 9 Fekry, M. I., Tipton, P. A. & Gates, K. S. 2011 Kinetic consequences of replacing the internucleotide phosphorus atoms in DNA with arsenic. *ACS Chem. Biol.* **6**, 127–130. (doi:10.1021/cb2000023)
- 10 Wolfe-Simon, F. *et al.* 2011 A bacterium that can grow by using arsenic instead of phosphorus. *Science* **332**, 1163–1166. (doi:10.1126/science.1197258)