Phosphorylated Sites in Substrates of Intracellular Protein Kinases: A Common Feature in Amino Acid Sequences

Abstract. *Examination of the primary amino acid sequences surrounding phosphorylated sites in many intracellular phosphoproteins indicated that the phosphorylated hydroxyamino acid (either serine or threonine) is, in general, surrounded by amino acids having a positively charged side chain and, more specifically, is frequently separated from a basic amino acid (either lysine or arginine) by only one amino acid. Possible reasons for this common feature are discussed.*

Intracellular protein kinases, which transfer phosphate from adenosine triphosphate (ATP) to proteins, have been isolated from many different mammalian tissues—for example, liver, muscle, testis, and thymus—and from many different species—for example, calf, rabbit, rat, swine, trout, and *Physarum polycephalum* (*1*, *2*). Their protein substrates have also been isolated from many different subcellular sources, such as membranes (sugar transport proteins), nuclei (basic and acidic proteins), ribosomes, and mitochondria (*3*). Recently the primary amino acid sequences surrounding many of the phosphorylated sites in several of the protein substrates have been elucidated, namely, the histones (*3–10*), phosphorylase b (*11*), glycogen synthetase (*12*), myelin basic protein (*13–15*) phosphorylase kinase (*16*), troponin-I (*17*), and pyruvate kinase (*18*).

Because one of the possible mechanisms by which the enzymes recognize the protein substrates could involve recognition of the primary sequence of amino acids surrounding the site to be modified it seemed reasonable to examine the primary amino acid sequence around the phosphorylated sites in these proteins for similarities. The amino acid sequences surrounding the phosphorylated sites in histones H1, H2a, H2b, and H4, phosphorylase b, glycogen synthetase, myelin basic protein, phosphorylase kinase, troponin-1, and pyruvate kinase are listed in Table 1. As will be seen, there seems to be a common feature present in

Table 1. The amino acid sequences surrounding sites phosphorylated by protein kinases (*27*). Phosphorylated amino acids are in boldface and, where known, their positions within the primary sequence are assigned. Proximate basic amino acid residues are in italics.

Protein modified	Amino acid sequence	Origin of enzyme	References
Histone H1*	-Arg-*Lys*-Ala-**Ser**^{38}-Gly-Pro-	Rat liver in vivo; calf liver in vitro; pig brain in vitro	(*5*, *6*)
	-Ser-Gly-**Ser**^{105}-Phe-*Lys*-Leu-	Calf liver in vitro; rat liver in vivo; calf thymus in vitro	(*7*)
	-{*Arg*/*Lys*}-Lys-**Ser**^{160}-Pro-*Lys*-†	Trout testis in vitro‡	(*4*, *8*)
Histone H2a	*N*-Acetyl-**Ser**^{1}-Gly-*Arg*-Gly-	Trout testis in vitro‡	(*8*)
	-Thr-*Arg*-Ser-**Ser**^{19}-Arg-Ala-	Pig brain in vitro	(*5*)
Histone H2b	-Arg-*Lys*-Glu-**Ser**^{36}-Thr-Ser-Val-	Pig brain in vitro; human lymphocytes in vitro	(*5*, *9*)
	-Lys-*Lys*-Gly-**Ser**^{14}-Lys-Ala-	Pig brain in vitro	(*5*)
	-Pro-Ala-Ac-Lys^{6}-**Ser**-Ala-Pro-*Lys*-Ac-Lys-	Trout testis in vitro	(*10*)
Histone H4	*N*-acetyl-**Ser**^{1}-Gly-*Arg*-Gly-Lys-	Trout testis in vivo	(*8*)
Phosphorylase b	-{Lys/Arg}-Gln-Ileu-**Ser**^{14}-Val/Ileu-*Arg*-Gly-	Rabbit muscle in vitro; human muscle in vitro	(*11*)
Glycogen synthetase	-{Arg/Lys}-Glu-Ileu-**Ser**-Val-*Arg*-	{Rabbit/Rat} muscle in vitro	(*12*)
Myelin basic protein	-Arg-*His*-Gly-**Ser**^{12}-Lys-Tyr-Leu-	Bovine cardiac muscle in vitro	(*13*)
	-Gly-*Arg*-Gly-Leu-**Ser**^{110}-Leu-Ser-*Arg*-	Rabbit muscle in vitro	(*14*)
	-Arg-Gly-**Ser**^{55}-Gly-*Lys*-Asp-Gly-	Rabbit muscle in vitro	(*15*)
Phosphorylase kinase			
α-subunit	-*Arg*-Leu-**Ser**-Ileu-Ser-Thr-Glu-	Rabbit muscle in vitro	(*16*)
β-subunit	-Gln-Ser-Gly-**Ser**-Val/Ileu-Tyr-Pro-	Rabbit muscle in vitro	(*16*)
Troponin-1	-Arg-Ala-Ileu-**Thr**-Ala-*Arg*-Arg-	Rabbit white skeletal muscle in vitro	(*17*)
	-Val-*Lys*-Ser-**Ser**-Lys-Glu-	Rabbit white skeletal muscle in vitro	(*17*)
	-Val-*Arg*-Met-**Ser**-Ala-Asx-	Rabbit white skeletal muscle in vitro	(*17*)
Pyruvate kinase	-Leu-Arg-*Arg*-Ala-**Ser**-Leu-	Pig liver in vitro	(*18*)
Protamines§	-(Arg)-Arg-**Ser-Ser-Ser**-Arg-(Pro)-	Trout sperm in vitro‡	(*22*)
	-(Arg)-Val-**Ser**-Arg-(Arg)-	Trout sperm in vitro	
	-(Arg)-Val-**Ser**-Arg-(Arg)-	Trout and salmon sperm in vitro‡	(*23*)
	-(Arg)-Ala-**Ser**-Arg-(Arg)-	Trout and salmon sperm in vitro	
	-(Arg)-Arg-**Ser-Ser**-Arg-(Arg)-	Trout and salmon sperm in vitro	

*Histone abbreviations are those agreed on during Ciba Foundation Symposium No. 28 (1975) (*4*). †The location of this sequence within the histone primary sequence can be inferred from the recently published sequence (*4*). ‡Experiments were done with a sperm cell suspension. §Parentheses around the Arg residues indicate probable positions, assigned in view of the substrate specificity of the trypsin used in sequencing.

many of the primary sequences surrounding the phosphorylated sites in these proteins. In general, all phosphorylated sites are separated from either lysine or arginine by no more than two amino acids. In many cases a very specific pattern can be seen. The hydroxyamino acid modified (either serine or threonine) is separated by only one amino acid from either of these amino acids (*19–21*).

Of the four exceptions to the specific pattern, the histone H2b sequence,

Ac
-Ala-Lys-Ser-Ala-Pro-Lys-

phosphorylase kinase (β subunit), and the two sequences in myelin basic protein, the one from myelin basic protein containing the histidine could be considered to fall into the specific pattern if the protein kinase were maximally active at any *p*H below 7 where the histidine would be in its protonated form. Such seems to be the case with the rabbit muscle protein kinase where maximum activity occurs at *p*H's below 7 (*1*). The sequences from myelin basic protein and histone H2b fall into the more general pattern since the modified serine is separated from the basic amino acid by no more than two amino acid residues.

Although the sequences around most of the phosphorylated serines in the nucleoprotamines seem to fit into the specific pattern (*22, 23*) (Table 1), the existence of the pattern is not certain because of the large stretches of arginines and serines within their sequences. The detection of a specific pattern would require that the sequence of events during serine modification by phosphate be known.

The requirement for the proximity of the cationic site in the amino acid sequence could result from two possible sources: one would originate from the requirements of the protein kinase; the other would originate from the requirements of the protein substrate.

In the first instance, the substrate specificity of all the protein kinases, regardless of origin, might be controlled by primary sequence dictates. This type of control would not be unique since the group transfer of carbohydrates to the immunoglobulins seems to be controlled, in part, by primary sequence dictates (*24*). Primary sequence dictates could play a major or minor role in the mechanism by which the protein kinases exhibit substrate specificity. If a major role were being played, the diverse functional nature of the protein substrates would suggest that all protein kinases, regardless of their origin, have the same requirements for substrate specificity. Overlapping substrate specificities have been noted, but the existing evidence is insufficient for direct comparisons to be made either between or even within sets of experiments (*1*). In addition, it has been pointed out that the phosphorylation of phosphorylase kinase must be carried out by an "extremely specific enzyme" (*16*), which would seem to rule out the possibility that all protein kinases isolated to date are in fact the same enzyme.

Primary sequence dictates might play only a minor role in determining the overall substrate specificity. The specific pattern of amino acid sequences might meet only part of the enzyme's requirements. The final determination of the specificity would rest either within the surrounding amino acid acids or within the three-dimensional structure of the protein surrounding the site being modified. The enzyme's dependency on cyclic nucleotides might also aid in determining the substrate specificity.

As an alternative to, or even in conjunction with, the enzyme's requirements, the protein substrate might require that the common amino acid sequence be present. The protein substrate's activity might be controlled by a conformational change brought on by the formation of an intramolecular salt link between the newly phosphorylated site and the neighboring basic (cationic) amino acid. Evidence that single-site phosphorylation changes the activity of the protein substrates has been presented (*11, 25, 26*). In addition, the presence of the hydrophilic cationic groups neighboring the phosphorylated site might be required to keep the site accessible to the protein kinase by drawing the site to the protein-water interface.

Before any or all of the reasons underlying the presence of this common structural feature in these protein kinase substrates can be determined more work on the intracellular protein kinases and their substrates will be required.

Ross E. Williams
Division of Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6

References and Notes

1. E. G. Krebs, in *Current Topics in Cellular Regulation*, B. L. Horecker and E. R. Stadtman, Eds. (Academic Press, New York, 1972), vol. 5, p. 99.
2. D. A. Walsh and E. G. Krebs, in *The Enzymes*, P. D. Boyer, Ed. (Academic Press, New York, 1973), vol. 8, p. 555; C. S. Rubin and O. M. Rosen, *Annu. Rev. Biochem.* **44**, 831 (1975); D. A. Walsh and C. D. Ashby, *Recent Prog. Horm. Res.* **29**, 329 (1973); T. A. Langan, *Adv. Cyclic Nucleotide Res.* **3**, 99 (1973); S. C. R. Elgin, S. C. Froehner, J. E. Smart, J. Bonner, *Adv. Cell Mol. Biol.* **1**, 1 (1971); L. S. Hnilica, *The Structure and Biological Function of Histones* (CRC Press, Cleveland, 1972).
3. G. Taborsky, in *Advances in Protein Chemistry*, C. B. Anfinsen *et al.*, Eds. (Academic Press, New York, 1974), vol. 28, p. 1.
4. S. C. R. Elgin and H. Weintraub, *Annu. Rev. Biochem.* **44**, 725 (1975).
5. S. V. Shlyapnikov, A. A. Arutyunyan, S. N. Kurochkin, L. V. Memelova, M. V. Nesterova, L. P. Saschenko, E. S. Severin, *FEBS Lett.* **53**, 316 (1975).
6. T. A. Langan, *Proc. Natl. Acad. Sci. U.S.A.* **64**, 1276 (1969).
7. ———, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **30**, 1089 (1971); *Ann. N.Y. Acad. Sci.* **185**, 160 (1971).
8. M. T. Sung and G. H. Dixon, *Proc. Natl. Acad. Sci. U.S.A.* **67**, 1616 (1970).
9. A. Farago, T. Romhanyi, F. Antoni, A. Takats, F. Fabian, *Nature (London)* **254**, 88 (1975); E. Hashimoto, M. Takeda, Y. Nishizuka, K. Hamana, K. Iwai, *Biochem. Biophys. Res. Commun.* **66**, 547 (1975).
10. A. J. Louie, E. P. M. Candido, G. H. Dixon, *Cold Spring Harbor Symp. Quant. Biol.* **38**, 803 (1974).
11. K. Titani, P. Cohen, K. A. Walsh, H. Neurath, *FEBS Lett.* **55**, 120 (1975); E. H. Fisher, D. G. Graves, E. R. Snyder-Crittenden, E. G. Krebs, *J. Biol. Chem.* **234**, 1698 (1959); D. P. Wolf, E. H. Fischer, E. G. Krebs, *Biochemistry* **9**, 1923 (1970).
12. J. Larner and F. Sanger, *J. Mol. Biol.* **11**, 491 (1965); A. M. Rosenkrans and J. Larner, *Biochim. Biophys. Acta* **315**, 317 (1973).
13. P. Daile and P. R. Carnegie, *Biochem. Biophys. Res. Commun.* **61**, 852 (1974).
14. P. R. Carnegie, B. E. Kemp, P. R. Dunkley, A. W. Murray, *Biochem. J.* **135**, 569 (1973); P. Daile, P. R. Carnegie, J. D. Young, *Nature (London)* **257**, 416 (1975).
15. P. R. Carnegie, P. R. Dunkley, B. E. Kemp, A. W. Murray, *Nature (London)* **249**, 147 (1974).
16. P. Cohen, D. C. Watson, G. H. Dixon, *Eur. J. Biochem.* **51**, 79 (1975).
17. T. S. Huang, D. B. Bylund, J. T. Stall, E. G. Krebs, *FEBS Lett.* **42**, 249 (1974); A. J. G. Moir, J. M. Wilkinson, S. V. Perry, *ibid.*, p. 253.
18. G. Hjelmquist, J. Andersson, B. Edlund, L. Engström, *Biochem. Biophys. Res. Commun.* **61**, 509 (1974).
19. If the probability of either lysine or arginine occurring in any particular position within the sequence is defined as P_L and P_A, respectively, then the probability of these two amino acids occurring in any particular position is $P_T = P_L + P_A$. The probability of the occurrence of these two amino acids in an average protein has been calculated (*20*). They are $P_L = 0.070$ and $P_A = 0.039$; thus $P_T = 0.109$. A more conservative estimate, from data related to the probability of nucleic acids occurring in triplets, gives $P_L = 0.055$ and $P_A = 0.107$; thus $P_T = 0.162$ (*21*). The conditional probability for the specific pattern, that is, that lysine or arginine occurs in either of the two positions separated from the phosphorylated serine by one amino acid, but not in both, is equal to $2P_T(1 - P_T)$. It is 0.194 if $P_T = 0.109$ and 0.271 if $P_T = 0.162$. Either of these is substantially less than the observed probability in the proteins listed in Table 1 ($P_T = 15/19 = 0.789$), excluding the nucleoprotamines.
20. M. O. Dayhoff and L. T. Hunt, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, p. D-355.
21. J. L. King and T. H. Jukes, *Science* **164**, 788 (1969).
22. M. M. Sanders and G. H. Dixon, *J. Biol. Chem.* **247**, 851 (1972).
23. C. J. Ingle and G. H. Dixon, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 1011 (1967).
24. L. T. Hunt and M. O. Dayhoff, *Biochem. Biophys. Res. Commun.* **39**, 757 (1970).
25. A. J. Adler, T. A. Langan, G. D. Fasman, *Arch. Biochem. Biophys.* **153**, 769 (1972).
26. G. Watson and T. A. Langan, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **32**, 588 (1973); P. Cohen and J. F. Antoniw, *FEBS Lett.* **34**, 43 (1973).
27. Abbreviations for the amino acids are those currently accepted and listed in *IUPAC Bull. No. 23* (June 1972).
28. Issued as NRC communication No. 15258. I thank Dr. Mike Zuker, Division of Biomathematics, National Research Council of Canada, for suggestions related to the calculation of the probabilities.

10 October 1975; revised 13 January 1976