Protein Science (1997), 6:469–472. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Identification of a novel phosphatase sequence motif

JOSEPH STUKEY[1] AND GEORGE M. CARMAN[2]

[1]Department of Biology, Hope College, Holland, Michigan 49422
[2]Department of Food Science, Rutgers University, Cook College, New Brunswick, New Jersey 08903

(RECEIVED September 20, 1996; ACCEPTED November 8, 1996)

Abstract: We have identified a novel, conserved phosphatase sequence motif, KXXXXXXRP-(X_{12-54})-PSGH-(X_{31-54})-SRXXXXX HXXXD, that is shared among several lipid phosphatases, the mammalian glucose-6-phosphatases, and a collection of bacterial nonspecific acid phosphatases. This sequence was also found in the vanadium-containing chloroperoxidase of *Curvularia inaequalis*. Several lines of evidence support this phosphatase motif identification. Crystal structure data on chloroperoxidase revealed that all three domains are in close proximity and several of the conserved residues are involved in the binding of the cofactor, vanadate, a compound structurally similar to phosphate. Structure–function analysis of the human glucose-6-phosphatase has shown that two of the conserved residues (the first domain arginine and the central domain histidine) are essential for enzyme activity. This conserved sequence motif was used to identify nine additional putative phosphatases from sequence databases, one of which has been determined to be a lipid phosphatase in yeast.

Keywords: chloroperoxidase; homology; glucose-6-phosphatase; lipid phosphatase; motif; nonspecific acid phosphatase; phosphatase

Phosphatases (phosphohydrolases) catalyze the hydrolysis of phosphoester and phosphoanhydride bonds of a diverse set of substrates including phosphorylated sugars, proteins, nucleic acids, and lipids (Boyer et al., 1961). Various criteria have been used to classify phosphatases into families. For example, phosphatases are classified on the basis of characteristics such as the molecular nature of the phosphate-containing substrate (phosphomonoester or phosphodiester), substrate type (proteinaceous or non-proteinaceous), substrate specificity (glucose-6-phosphatases, protein-tyrosine phosphatases), pH optimum (acid and alkaline phosphatases), size (high and low molecular weight phosphatases), and on the identity of the enzyme residue that is transiently phosphorylated during catalysis (histidine, serine, and cysteine phosphatases) (Boyer et al., 1961; Walsh, 1979; Vincent et al., 1992; Ostanin et al., 1994; Stone & Dixon, 1994).

Reprint requests to: J. Stukey, Department of Biology, Hope College, P.O. Box 9000, Holland, Michigan 49422-9000; e-mail: stukey@hope.edu.

Another useful means of classification is one based on amino acid sequence homology. Conserved amino acid sequence data provides the opportunity to search for structurally and functionally related proteins by computer and to tentatively identify the function of proteins encoded by newly sequenced genes. Several motifs (signature sequences) identifying phosphatase function have been described. For example, the (I/V)HCXAGXGR(S/T)G sequence is found in protein-tyrosine phosphatases (Stone & Dixon, 1994), the DXH-($X_{\approx 25}$)-GDXXD-($X_{\approx 25}$)-GNH(D/E) sequence is found in the serine/threonine protein phosphatases (Zhuo et al., 1994), and the RHG sequence is found in 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase, phosphoglycerate mutases, and several acid phosphatases (Bazan et al., 1989). In this report we describe the identification of a novel phosphatase signature motif KXXXXXXRP-(X_{12-54})-PSGH-(X_{31-54})-SRXXXXXHXXXD. This motif was found in a family of glucose-6-phosphatases, nonspecific acid phosphatases, and several lipid phosphatases. Moreover, we have used this motif to identify additional putative phosphatases currently reported as hypothetical gene products in sequence databases.

Discussion: In our effort to identify possible lipid phosphatase genes, we used the *Escherichia coli pgpB* gene product (Icho, 1988) in a BLAST search of sequence databases at the National Center for Biotechnology Information (NCBI) (Altschul et al., 1990). The protein showing the greatest similarity was the *pgpB* homologue from *Haemophilus influenzae* Rd (Fleischmann et al., 1995). The alignment of *E. coli* and *H. influenza pgpB* proteins showed two ungapped segments covering nearly the entire length of each protein and two short regions of concentrated amino acid identity. Additional proteins containing similar sequences included two nonspecific acid phosphatases, one neutral phosphatase, a single ATP diphosphohydrolase, and three hypothetical gene products. On examining these sequences we found alignments defining two short conserved domains (PSGH and SRXXXXXHXXD, where X represents any residue) (Table 1). Using the conserved residues found in these two domains, we performed additional BLAST database searches and identified five additional phosphatases, seven "hypothetical" gene products, and a chloroperoxidase that also contained these sequences (Table 1). A third conserved domain (KXXXXXXRP), N-terminal to the PSGH sequence, was subsequently identified following our analysis of the chloroperoxidase

Table 1. *Phosphatase motif consensus sequence and protein alignments*[a]

Protein (source, accession #)	Domain 1[b,c]	Domain 2	Domain 3[d]	Reference[e]
PGP[f] phosphatase (*E. coli*, P18201)	96-KDKVQEPRP-54-	PSGH	-36-SRLLLGMHWPRD-254	Icho, 1988
PGP phosphatase (*H. influenzae*, P44570)	93-KALFEEPRP-54-	PSGH	-41-SRVRLGMHYPID-241	Fleischmann et al., 1995
PA[g] phosphatase (*M. musculus*, D84376)	119-KYTIGSLRP-39-	YSGH	-44-SRVSDYKHHWSD-283	Kai et al., 1996
Glucose-6-phosphatase (*H. sapiens*, P35575)	75-KWILFGQRP-31-	PSGH	-49-SRIYLAAHFPHQ-357	Lei et al., 1993
Glucose-6-phosphatase (*R. norvegicus*, L37333)	71-KWILFGQRP-31-	PSGH	-49-SRIYLAAHFPHQ-353	Haber et al., 1995
Glucose-6-phosphatase (*M. musculus*, P35576)	75-KWILFGQRP-31-	PSGH	-49-SRIYLAAHFPHQ-357	Shelly et al., 1993
Phosphatase (*T. denticola*, L25421)	82-KRILKIPRP-17-	PSGH	-54-SRVYLGVHYPTD-341	Ishihara & Kuramitsu, 1995
Apyrase (*S. flexneri*, U04539)	123-KEYYKRVRP-23-	PSGH	-31-SRVICGAHWQSD-246	
NS-phosphatase[h] (*S. typhimurium*, S14515)	122-KKYYMRTRP-23-	PSGH	-31-SRVICGAHWQSD-232	Kasahara et al., 1991
NS-phosphatase (*M. morganii*, P28581)	132-KEHYMRIRP-23-	PSGH	-31-SRVICGYHWQSD-249	Thaller et al., 1994
NS-phosphatase (*P. stuartii*, P26975)	132-KEKYMRIRP-23-	PSGH	-31-SRVICGYHWQSD-248	
Acid phosphatase (*Z. mobilis*, P14924)	131-KNNWNRKRP-23-	PSGH	-31-SRIVCGAHWFSD-264	Pond et al., 1989
Hypothetical (*S. cerevisiae*, U51031)[i]	117-KNWIGRLRP-39-	PSGH	-46-SRTQDYRHHFVD-289	
Hypothetical (*S. cerevisiae*, U33057)	135-KLIIGNLRP-41-	PSGH	-38-SRVIDHRHHWYD-275	
Hypothetical (*S. cerevisiae*, X87371)	127-KDYWCLPRP-20-	PSSH	-42-GRIYCGMHGILD-409	
Hypothetical (*S. cerevisiae*, Z28278)	128-KDYWCLPRP-20-	PSSH	-42-GRVYCGMHGMLD-404	
Hypothetical (*C. elegans*, Z68105)	160-KCYVGRLRP-44-	PSGH	-47-TRVTDNWHFPTD-318	
Hypothetical (*C. elegans*, U28738)	143-KHVVGRLRP-41-	YSGH	-45-SRITDNKHHWSD-341	
Hypothetical (*C. elegans*, U39648)	82-KFYFHRERP-17-	PSGH	-31-SRVALGRHYITD-345	
Hypothetical (*Synechocystis sp.*, D64003)	133-KPFFNRTRP-12-	PSGH	-39-ASMYCRVHWATD-240	
Hypothetical (*D. mobilis*, S01073)	77-KHLFNTPRP-12-	PSGH	-33-SRLYLRAHYPID-225	Kjems & Garrett, 1987
Hypothetical (*B. megaterium*, S32217)	106-KLFFQRARP-13-	PSGH	-41-SRIYLGVHYPSD-216	Rauschenbach et al., 1993
Chloroperoxidase (*C. inaequalis*, X85369)	352-KWEFEFWRP-39-	PSGH	-84-SRIFLGVHWRFD-609	Simons et al., 1995
CONSENSUS SEQUENCE	KXXXXXXRP	PSGH	SRXXXXXHXXXD	

[a]Protein sequences were aligned relative to their phosphatase motifs. The three phosphatase motif domains are indicated at top.
[b]Numbers preceding domain 1 indicate length in amino acids of N-terminus.
[c]Numbers between domains indicate amino acid spacer lengths.
[d]Numbers following domain 3 indicate total number of amino acids in each protein.
[e]All sequences listed without reference were deposited directly into databases.
[f]PGP: phosphatidylglycerol phosphate.
[g]PAL: phosphatidic acid.
[h]NS: nonspecific.
[i]Identified as diacylglycerol pyrophosphate (DGPP) phosphatase.

crystal structure (Messerschmidt & Wever, 1996) and protein sequence alignments. For clarity, we will refer to the sequence nearest the N-terminus, KXXXXXXRP, as domain 1, the centrally located PSGH sequence as domain 2, and the sequence nearest the C-terminus, SRXXXXXHXXD, as domain 3.

All three domains are between four and eleven amino acids in length and include three or four conserved residues. Separating each of the three domains are protein segments ranging from 12 to 54 amino acids in length. Furthermore, the order of all three domains in each protein is invariant. We propose that the three identified conserved domains collectively define a new phosphatase signature motif. A consensus sequence and list of 23 established or hypothetical proteins possessing this phosphatase signature motif is presented in Table 1. The largest group of proteins of known function are phosphomonoesterases. They are found in a diverse range of organisms from bacteria to humans. The known and predicted phosphatases vary in size from 216 to 357 residues in length and include both membrane-bound and cytosolic predicted forms (Icho, 1988; Pond et al., 1989; Lei et al., 1993; Thaller et al., 1994; Ishihara & Kuramitsu, 1995; Kai et al., 1996). The single known nonphosphatase is the vanadium chloroperoxidase of *Curvularia inaequalis* and it is the largest at a predicted 609 amino acids.

We can make some predictions that test the significance of this alignment data: (1) the conserved residues are essential for enzyme function, and as such, may be found at the catalytic site, and (2) the listed hypothetical gene products are phosphatases. In addressing these predictions, we found evidence supporting the identification of the phosphatase signature motif in the recent publications on the crystal structure of the vanadium chloroperoxidase of *C. inaequalis* and a structure–function analysis of the membrane-bound human glucose-6-phosphatase.

The enzyme chloroperoxidase catalyzes the oxidation of halides in the presence of hydrogen peroxide and requires the cofactor, vanadium. Orthovanadate (vanadate) is the biologically active form of vanadium (Willsky, 1990) and is structurally similar to orthophosphate (Clarkson, 1993). As such, vanadate is also known to compete in enzymatic reaction involving orthophosphate, apparently by molecular and ionic mimicry (Clarkson, 1993). Furthermore, many enzymes that function via a phosphoenzyme intermediate, including some phosphatases (Singh et al., 1981), or that bind ATP or other high-energy phosphate compounds (Cantley et al., 1978; Karlish et al., 1979), are sensitive to vanadate. Thus, information on the binding of vanadate by chloroperoxidase could be useful in modeling a "phosphatase active site" relevant to our list of known and putative phosphatases.

The crystal structure of an azide-chloroperoxidase complex (Messerschmidt & Wever, 1996) shows the conserved motif residues participating in the binding of the enzyme cofactor, vanadate.

The bound vanadate assumes a trigonal bipyramidal coordination within the enzyme pocket with the central metal and three peripheral nonprotein oxygens equatorial to the azide ligand at one apical position and the $\epsilon 2$ nitrogen of His 496 at the other apical position. This trigonal pyramidal coordination is remarkably similar to that described for the nucleophilic attack of orthophosphate (Walsh, 1979; Vincent et al., 1992) and suggests a structural and catalytic model for this conserved phosphatase motif. The features of this model include: (1) nucleophilic attack of the substrate's phosphoryl group by the histidine of domain 3 and production of a phosphoenzyme catalytic intermediate, (2) conserved arginine residues of domains 1 and 3 hydrogen binding to the equatorial phosphoryl oxygens, and (3) the histidine of domain 2 participating in the protonation of the substrate leaving group.

Additional data supporting the significance of these conserved residues are available from studies of the human glucose-6-phosphatase. Humans deficient for glucose-6-phosphatase are afflicted with glycogen storage disease (GSD) type 1a. Previous investigations of the glucose-6-phosphatase encoding gene of GSD type 1a patients have identified several residues essential for normal enzyme function including Arg 83 (Lei et al., 1993). This residue is the conserved arginine in domain 1 of the phosphatase motif. In addition, earlier experiments on glucose-6-phosphatase indicated that a phosphoenzyme intermediate is formed during catalysis, and that a histidine residue is the phosphoryl acceptor (Feldman & Butler, 1972; Countaway et al., 1988). Using site-directed mutagenesis to study the structure–function relationship of the human glucose-6-phosphatase, Lei et al. (1995) separately introduced amino acid substitutions for Arg 83 and His 119 (the conserved histidine present in domain 2) and tested the mutated enzymes for glucose-6-phosphatase activity. All tested amino acid substitutions for those two residues (including the conservative Lys substitution for Arg 83) resulted in loss of glucose-6-phosphatase activity. Finally, although most of the conserved amino acids of domain 3 are present in the human glucose-6-phosphatase, those residues were not investigated in that study (Lei et al., 1995). The proposed glucose-6-phosphatase topology model predicts domain 3 on the opposite and noncatalytic side of the membrane from domains 1 and 2. In light of the evidence presented here for a three-domain phosphatase motif, the glucose-6-phosphatase topology model may need reconsideration.

In conclusion, we have identified a conserved phosphatase sequence motif that can serve as a predictor of phosphatase enzyme function and that indicates homology among known bacterial acid phosphatases, glucose-6-phosphatases, several lipid phosphatases, and the vanadium-containing chloroperoxidase of *C. inaequalis*. The following existing experimental evidence supports the identification of this phosphatase signature motif: (1) these conserved sequences are conceptually present in 13 known phosphatases, (2) the crystal structure of chloroperoxidase shows the conserved residues forming a vanadate-binding enzyme pocket and suggests a model for the binding of phosphorylated substrates by this family of phosphatases, (3) structure–function analysis of the human glucose-6-phosphatase indicates two of the absolutely conserved residues (arginine of domain 1 and histidine of domain 2) are essential for enzyme function, and (4) a current database "hypothetical" gene product (accession #U51031) having this phosphatase signature motif, has been identified as a diacylglycerol pyrophosphate (DGPP) phosphatase of *Saccharomyces cerevisiae* (Zeimetz & Carman, in prep.).

Acknowledgments: We wish to thank Drs. Virginia M. McDonough and Joanne Stewart for helpful discussions related to this work, and Drs. Virginia M. McDonough and Maria Burnatowska-Hledin for critical reading of this manuscript. This work was supported by United States Public Health Service Grant GM-28140 from the National Institutes of Health, National Science Foundation Grant MCB-9204588, and by the Hope College Department of Biology. This is New Jersey Agricultural Experiment Station Publication D-10108-1-96.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol 215*:403–410.

Bazan JF, Fletterick RJ, Pilkis SJ. 1989. Evolution of a bifunctional enzyme: 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. *Proc Natl Acad Sci USA 86*:9642–9646.

Boyer PD, Lardy H, Mayback K, eds. 1961. *The enzymes, Vol V*. New York, Academic Press.

Cantley LC, Resh MD, Guidotti G. 1978. Vanadate inhibits the red cell (Na+, K+) ATPase from the cytoplasmic side. *Nature 272*:552–554.

Clarkson TW. 1993. Molecular and ionic mimicry of toxic metals. *Annu Rev Pharmacol Toxicol 33*:545–571.

Countaway JL, Waddell ID, Burchell A, Arion WJ. 1988. The phosphohydrolase component of the hepatic microsomal glucose-6-phosphatase system is a 36.5-kilodalton polypeptide. *J Biol Chem 263*:2673–2678.

Feldman F, Butler LG. 1972. Protein-bound phosphoryl histidine: A probable intermediate in the microsomal glucose-6-phosphatase-inorganic pyrophosphatase reaction. *Biochim Biophys Acta 268*:698–710.

Fleischmann RD et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science 269*:496–512.

Haber BA, Chin S, Chuang E, Buikhuisen W, Naji A, Taub R. 1995. High levels of glucose-6-phosphatase gene and protein expression reflect an adaptive response in proliferating liver and diabetes. *J Clin Invest 95*:832–841.

Icho T. 1988. Membrane-bound phosphatases in *Escherichia coli*: Sequence of the *pgpB* gene and dual subcellular localization of the *pgpB* product. *J Bacteriol 170*:5117–5124.

Ishihara K, Kuramitsu HK. 1995. Cloning and expression of a neutral phosphatase from *Treponema denticola*. *Infect Immun 63*:1147–1152.

Kai M, Wada I, Imai S-i, Sakane F, Kanoh H. 1996. Identification and cDNA cloning of 35-kDa phosphatidic acid phosphatase (type 2) bound to plasma membranes. Polymerase chain reaction amplification of mouse H_2O_2-inducible *hic53* clone yielded the cDNA encoding phosphatidic acid phosphatase. *J Biol Chem 271*:18931–18938.

Karlish SJD, Beauge LA, Glynn IM. 1979. Vanadate inhibits (Na^+, K^+) ATPase by blocking conformational change of the unphosphorylated from. *Nature 282*:333–335.

Kasahara M, Nakata A, Shinagawa H. 1991. Molecular analysis of the *Salmonella typhimurium phoN* gene, which encodes nonspecific acid phosphatase. *J Bacteriol 173*:6760–6765.

Kjems J, Garrett RA. 1987. Novel expression of the ribosomal RNA genes in the extreme thermophile and archaebacterium *Desulfurococcus mobilis*. *EMBO J 6*:3521–3530.

Lei K-J, Shelly LL, Pan C-J, Sidbury JB, Chou JY. 1993. Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1A. *Science 262*:580–583.

Lei K-J, Pan C-J, Liu J-L, Shelly LL, Chou JY. 1995. Structure–function analysis of human glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1A. *J Biol Chem 270*:11882–11886.

Messerschmidt A, Wever R. 1996. X-ray structure of a vanadium-containing enzyme: Chloroperoxidase from the fungus *Curvularia inaequalis*. *Proc Natl Acad Sci USA 93*:392–396.

Ostanin K, Saeed A, Van Etten RL. 1994. Heterologous expression of human prostatic acid phosphatase and site-directed mutagenesis of the enzyme active site. *J Biol Chem 269*:8971–8978.

Pond JL, Eddy CK, Mackenzie KF, Conway T, Borecky DJ, Ingram LO. 1989. Cloning, sequencing, and characterization of the principal acid phosphatase, the *phoC*$^+$ product, from *Zymomonas mobilis*. *J Bacteriol 171*:767–774.

Rauschenbach R, Isernhagen M, Noeske-Jungblut C, Boidol W, Siewert G. 1993. Cloning, sequencing, and expression of the gene for cytochrome P450meg, the steroid-15B-monooxygenase from *Bacillus megaterium* ATCC 13368. *Mol Gen Genet 241*:170–176.

Shelly LL, Lei K-J, Pan C-J, Sakata SF, Ruppert S, Schutz G, Chou JY. 1993. Isolation of the gene for murine glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1A. *J Biol Chem 268*:21482–21485.

Simons BH, Barnett P, Vollenbroek EG, Dekker HL, Muijsers AO, Messerschmidt A, Wever R. 1995. Primary structure and characterization of the

vanadium chloroperoxidase from the fungus *Curvularia inaequalis*. *Eur J Biochem* 229:566–574.

Singh J, Nordlie RC, Jorgenson RA. 1981. Vanadate: A potent inhibitor of multifunctional glucose-6-phosphatase. *Biochim Biophys Acta* 678:477–482.

Stone RL, Dixon JE. 1994. Protein-tyrosine phosphatases. *J Biol Chem* 269: 31323–31326.

Thaller MC, Berlutti F, Schippa S, Lombardi G, Rossolini GM. 1994. Characterization and sequence of *PhoC*, the principal phosphate-irrepressible acid phosphatase of *Morganella morganii*. *Microbiology* 140:1341–1350.

Vincent JB, Crowder MW, Averill BA. 1992. Hydrolysis of phosphate monoesters: a biological problem with multiple solutions. *Trends Biochem Sci* 17:105–110.

Walsh C. 1979. *Enzymatic reaction mechanisms*. New York: W. H. Freeman and Co.

Willsky G. 1990. Vanadium in the biosphere. In: Chasteen ND, ed. *Vanadium in biological systems*. Amsterdam, The Netherlands: Kluwer Academic Publishers. pp 1–24.

Zhuo S, Clemens JC, Stone RL, Dixon JE. 1994. Mutational analysis of a Ser/Thr phosphatase. Identification of residues important in phosphoesterase substrate binding and catalysis. *J Biol Chem* 269:26234–26238.