Covalent Structure, Disulfide Bonding, and Identification of Reactive Surface and Active Site Residues of Human Prostatic Acid Phosphatase*

(Received for publication, August 27, 1990)

Robert L. Van Etten†, Ross Davidson, Panayiotis E. Stevis, Heather MacArthur, and D. Lane Moore
From the Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

The pairing of the half-cysteine residues of human prostatic acid phosphatase was established by proteolytic digestion and analysis of the resulting peptide mixtures by fast atom bombardment mass spectrometry (FAB-MS). An independently derived, full length cDNA clone was used as the basis for the interpretation of the FAB-MS data. The sequence of the native protein is that predicted from the present cDNA sequence, except for the carboxyl-terminal end and some possible post-translational deamidations. Isolated human prostatic acid phosphatase was found to have multiple carboxyl-terminal ends, terminating in Thr, Glu, and Asp, corresponding to residues 349-351 of the 354-residue protein that is predicted from the cDNA sequence after removal of a leader peptide. The protein contains no free sulfhydryl groups. The identical monomer chains of the dimeric native enzyme are found to contain three disulfide bonds, specifically Cys-129 to Cys-340, Cys-183 to Cys-281, and Cys-315 to Cys-319. In view of the conserved positions of cysteines in the homologous human and rat liver lysosomal acid phosphatases, an identical disulfide bonding pattern may be predicted for those proteins. The location of a potential antigenic site was established by selective labeling of proximate tyrosines residues predicted to be on the surface. A conserved RHGXXRXP sequence is present in the prostatic, lysosomal, Escherichia coli, and yeast acid phosphatases and is predicted to be of mechanistic significance. In addition, residue Arg-54 is shown to be an active site residue by reaction of the enzyme with phenylglyoxal. Interestingly, this residue is present in a sequence RXRY (R,H) that is also present in lysosomal acid phosphatases and in recently described protein tyrosine phosphatases.

As a group, the acid phosphatases are quite heterogeneous, with monomeric 18-kDa proteins, dimeric 90-kDa proteins, larger glycoprotein and membrane-bound forms, and even one well documented metalloenzyme example (1, 2). Because it was one of the first identified examples of a tumor marker, human prostatic acid phosphatase has been regarded as the most important and is the most extensively studied of the acid phosphatases (3, 4). Little is known about the three-dimensional structure of the protein, and, until last year, only fragmentary peptide sequence data had been described (5, 6). However, two groups have recently described the cloning and cDNA sequencing of human prostatic acid phosphatase (7, 8). Unfortunately, significant discrepancies exist between these two sequences. Moreover, the sequences also differed from results obtained in this laboratory, both at the cDNA level and in direct studies of the protein. Accordingly, we present independent results for the cDNA sequence. In these studies, the protein itself has been used to provide direct confirmation of the cDNA sequence, either by means of peptide sequencing, FAB-MS, or direct measurement of protein properties such as sulfhydryl content. With this sequence as a basis, we were able to obtain information related to the three-dimensional structure of the protein, including the disulfide bonding pattern and the location of reactive surface residues.

Despite a long history of study of human prostatic acid phosphatase, the number of sulfhydryl and disulfide bonds in the protein has remained uncertain. An early study concluded that each molecule (of the dimeric enzyme) "contains two unreactive sulfhydryl groups, the blocking of which . . . causes only a partial (30-50%) loss of activity," and that the molecule "contains a relatively large number of disulfide bridges (10/molecule), which probably play an essential role in shaping the internal structure of the enzyme molecule" (9). However, the purification of this enzyme was quite difficult until the advent of tetraromatic acid affinity chromatography methods (10, 11), so that these early results must be regarded as tentative. Another area of concern in structural studies of HPAP is the presence of multiple forms of the enzyme that are separable on isoelectric focusing (12). The majority of these forms remain even after removal of the carbohydrate (13). Finally, a comparison of homologous sequences permits a tentative identification of active site residues. One such active site residue was experimentally confirmed by covalent modification studies, and a potential surface epitope was also identified.

MATERIALS AND METHODS
Cloning and cDNA Sequence Determination—Oligonucleotide probes were initially synthesized by Dr. Gerald Zon (Federal Drug Administration, Bethesda, MD) and subsequently by the Laboratory of Biomedical Imaging, Inc. The abbreviations used are: FAB-MS, fast atom bombardment mass spectrometry; HPAP, human prostatic acid phosphatase; HLAP, human lysosomal acid phosphatase; TFA, trifluoroacetic acid; dsHPAP, desialylated human prostatic acid phosphatase; ds-CMHPAP, desialylated, carboxymethylated human prostatic acid phosphatase; HPLC, high performance liquid chromatography.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence and reprint requests should be addressed.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M34830.

1 The sequence reported in this paper and the additional sequence of bases 1200 to 3061 of the noncoding region including the poly(A) tail have been deposited in the GenBank data base (accession number M34830).

2 The abbreviations used are: FAB-MS, fast atom bombardment mass spectrometry; HPAP, human prostatic acid phosphatase; HLAP, human lysosomal acid phosphatase; TFA, trifluoroacetic acid; dsHPAP, desialylated human prostatic acid phosphatase; ds-CMHPAP, desialylated, carboxymethylated human prostatic acid phosphatase; HPLC, high performance liquid chromatography.
for Macromolecular Structure, Purdue University. We had previously (6) sequenced seven cyanogen bromide peptides (Fig. 1) of HPAP isolated from human seminal fluid (11, 14). The first probe was a 16-fold degenerate 17-mer (RVE-1) based on the sequence of cyanogen bromide peptide 1 (Fig. 1). Publication (15) of the sequence of the human lysosomal acid phosphatase (HLPAP) made possible the recognition that HPAP and HLPAP are homologous and allowed us to align our seven HPAP peptides. From two peptides which appeared to be contiguous, a nondegenerate oligonucleotide (78-mer) was designed using the recommendations of Lathe (16) and the codon frequency of HLPAP (15).

A human prostate agt11 cDNA library (Clonetech) was successfully screened with the 78-mer. The probe was end-labeled using T4 polynucleotide kinase (New England Biolabs) and [γ-32P]ATP (Du Pont-New England Nuclear, 6,000 Ci/mmole) to a specific activity of 1-2 × 1010 cpm/µg. The prostate cDNA library was plated using Y1090 host cells onto 90-mm plates (10,000 plaque/plate). Duplicate filters from 4 plates were prepared and hybridized with the 78-mer following Maniatis et al. (17). Filters were prehybridized at 42°C for 2-4 h and then hybridized for 16 h at varying temperatures with 8 × 106 cpm/blot in a solution consisting of 5 × Denhardt's solution, 5 × SSPE, 0.1% sodium dodecyl sulfate, 100 µg/ml salmon sperm DNA and containing 0% formamide (65°C) or 10, 20, and 40% formamide (42°C). All filters were washed in 0.2 × SSC, 0.1% sodium dodecyl sulfate at 42°C (18, 19). λ-DNA from these clones was produced using Y1090 host cells (20).

Plaque-purified clones were identified by DNA sequencing of purified double-stranded λ DNA (21) using Sequenase (22) (US. Biochemical Corp.), Xgtll primers (New England Biolabs), and DNA. Plasmid DNA for sequencing was prepared with the omission of 0.1% sodium dodecyl sulfate, 100 µg/ml salmon sperm DNA and containing 0% formamide (65°C) or 10, 20, and 40% formamide (42°C). All filters were washed in 0.2 × SSC, 0.1% sodium dodecyl sulfate at 42°C (18, 19). λ-DNA from these clones was produced using Y1090 host cells (20).

Observed Sequences of Cyanogen Bromide Peptides

1. MEQY IEGY IEBK KFLKX ESYKH KERVY RSDTV D
2. MTXLR ESEL SLQSL YGIEK QCKKS
3. MTNLN ALFPF EGSXI NVPI LXXKI
4. MLPGX SPSSX LEXF
5. KELKF VTLVF RHGDR SPIDT FPTM I
6. MTTYX ETQNE FY
7. MSAH DTVS GLQM

Oligonucleotide Probes and Related Peptide Sequences

RVE-1 3'-TACATGATGGCCTTGCTCTGGGTCTTGCTCGGGATGGGC-5'

Peptide 1 MetGluGlnHisTyrGluLeuGlyGluTyrIleArgLysArg

Peptides N-MetTyrTyrArgAsnGluThrGlnHisGluProTyrPro LeuLeuLeuProGlyCysSerProSerCysProLeuGlu-C

78-mer 3'-TAGATGGGCCCTTCCTGCTCCCTGCTCCCGGAGGCC CAGTACGAGCGAGCGAGGACCGGTGCGGACGTCCTC-5'

Fig. 1. Sequences of HPAP peptides and derived oligonucleotides used in screening the cDNA library.

Other oligonucleotides corresponding to appropriate regions of the developing cDNA sequence, including the 16-fold degenerate RVE-1 probe, were used as primers in the sequencing reactions as in Ref. 25. Use of these primers minimized subcloning and deletion generation.

Protein Sequencing and Other Protein Chemistry—The number of disulfide bonds present in native HPAP was determined by titration of a 1 µM sample of HPAP with 1% dithiothreitol (DTT) in the absence and presence of 500 µM or 5 mM phenylmethylsulfonyl fluoride (PMSF). The cystine content of HPAP in the presence of PMSF remained constant, indicating that the disulfide bonds are not labile in the presence of PMSF.

Due to the large size of the HPAP monomer and the mass limitations of the spectrometer used (practical range 500-2500 mass units), it was necessary to use multiple cleavages to create fragments of a proper size for disulfide bond identification. The same limitations made it necessary to use overlap techniques in order to verify the full length amino acid sequence. Some of the cleavage methods were chosen to avoid the occurrence of fragments that would otherwise be too short for detection.

Purified HPAP (11, 14) was desialylated (5) (dsHPAP) and subjected to cleavage under a number of different conditions. dsHPAP was carboxymethylated (dsCMHPAP) by a variation of the method of McWherter et al. (27). In the experiment described as Series 1, dsCMHPAP (2 mg) was dissolved in 400 µl of 0.1 M NH4HCO3, pH 8.5, containing 1 mM tosylphenylalanyl chloromethyl ketone. Proteolysis was initiated by the addition of tryptic cleavage with respect to dsCMHPAP. The reaction was allowed to proceed for 6 h at 37°C and was terminated by storage at −80°C until ready for separation. The reaction mixture was thawed, and then solvent was added to yield a final volume of 1 ml and a final concentration of 0.1% TFA and 20% acetonitrile. Aliquots were analyzed using an IBM LC 9533 HPLC instrument with a Synchropak RP-P-C18 reverse phase column. The eluting gradient consisted of acetonitrile/0.1% TFA versus water/0.1% TFA, with an increase of the organic solvent at a rate of 1%/min and a constant flow rate of 1 ml/min. Dual UV detectors measured absorbance at 210 and 280 nm. The HPLC fractions were combined into pools corresponding to regions of the chromatogram where baseline separation of groups of peaks occurred. All subsequent experiments requiring HPLC separation were performed in this manner. The pooled fraction volumes were reduced to approximately 5–10 µl, formic acid was added to a concentration of approximately 20%, and the mixtures were analyzed on a Kratos 50 mass spectrometer in a matrix of 1:1 glycerol/thioglycerol. This is in order to observe both the intact disulfide fragment and its component sulfhydryl fragments. Series 2 and 3 experiments were identical except that the tryptic cleavage was performed in 6 M guanidine HCl and 4 M urea, respectively, in addition to the other buffer components.

Series 4–7 experiments were each performed on 2 mg of dsHPAP. In series 4, dsHPAP was subjected to cleavage for 24 h using a 100-fold (over methionine) excess of CNBr in 70% TFA, then sequentially subjected to tryptic cleavage as per series 1, followed by cleavage by a 1:20 ratio of V8 protease in 0.1 M NH4HCO3 plus 4 M urea, pH 7.8, at room temperature for 24 h. When CNBr cleavage takes place in a stronger acid such as TFA, some cleavage at tryptophan residues has been noted, as well as a reduced resistance to cleavage of Met-X bonds where X = Ser, Thr, or Cys (28). Of the 10 CNBr cleavage sites in HPAP, 3 were expected to demonstrate such a resistance to cleavage. Series 5 experiments utilized 70% formic acid in place of the TFA for the methionyl cleavage portion. Series 6 experiments utilized only the V8 protease cleavage. Series 7 experiments utilized V8 protease cleavage, followed by tryptic cleavage (i.e. no CNBr step).

FAB-MS data was sorted using the program AMONG, wrote by Ted Smith and graciously supplied by Prof. David Smith. The program lists all of the potential HPAP peptides within a specified mass range. The program was set up to use ±1.0 mass unit as the range for these experiments. Peptides that are not likely to be products of a given cleavage are eliminated. If more than one possible peptide remained, the sample was subjected to one cycle of Edman degradations, followed by another FAB-MS run. The new peaks should correspond to the original mass minus the mass of the amino-terminal amino acid.

There were discrepancies in cysteine content between the present FAB-MS and cDNA data with the sequence data in an earlier report (8). [35S]-labeled dsCMHPAP was subjected to tryptic cleavage according to series 1 conditions. Since the predicted HPAP sequence indicated...
that the disputed residues should be located within 14C-labeled fragments, the identity of the residues could be established by gas phase sequencing of those fragments. To resolve a discrepancy involving a sequence of 8 residues near the amino-terminal end as predicted from the present cDNA data and that in (7) native HPAP was submitted to gas phase sequencing without any prior proteolysis. The gas phase sequencing reactions were performed on an Applied Biosystems Model 470A instrument.

For carboxy-terminal determinations, HPAP was cleaved using the CNBr procedure mentioned previously. The cleavage mixture was dried in a Speed Vac and redissolved in 25% formic acid. The entire sample was analyzed, without separation, by plasma desorption mass spectrometry in a 252-Cf fission instrument.

HPAP was iodinated by the Pierce IODO-BEAD procedure using 0.1 mCi of 125I (14.3 mCi/μg 125I, Amersham) and following the spectrometry in a 252-Cf fission instrument. For labeling with [14C]phenylglyoxal (27.6 mCi/mmol, Amersham), 500 μl of HPAP (1 mg/ml in 0.1 M NaHCO3, at either pH 8.0 or pH 7.0, using a 2- to 1000-fold molar excess of phenylglyoxal. Incubation was done at 37°C and at ambient temperature. Aliquots were removed at selected intervals and assayed for phosphatase activity (11). Percent activity was determined versus a HPAP control incubated without phenylglyoxal.

For labeling with [14C]phenylglyoxal (27.6 μCi/mmol, Amersham), 500 μl of HPAP (1 mg/ml in 0.1 M NaHCO3, pH 7.0) solution was incubated with 15 μl of 3.6 mM [14C]phenylglyoxal (5 × molar excess) for 40 min. The reaction was terminated by adding sufficient glacial acetic acid to bring the final acid concentration to 10%. To permit quantitation of label incorporation, dual samples were run. One sample was assayed for HPAP activity. This sample was then subjected to repeated acetone precipitation and centrifugation to remove unbound label. The precipitate was resuspended, and the amount of bound 14C was determined by scintillation counting. The other sample was immediately subjected to pepsin cleavage. Pepsin (1 mg/ml) was added at a ratio of 1:10, and the sample was incubated at 37°C for 6 h. Pepsin cleavage was chosen, since significant loss of label during proteolysis at alkaline or neutral pH, as well as loss during subsequent peptide separation and sequencing steps, had been previously described for other proteins (29). The mixture was separated by reverse phase HPLC, and the peaks containing radioactivity were submitted for gas phase sequencing.

The kinetics of inactivation of HPAP by phenylglyoxal was studied by incubating 500-μl samples of 1 mg/ml HPAP in 0.1 M NH4HCO3 at either pH 8.0 or pH 7.0, using a 2- to 1000-fold molar excess of phenylglyoxal. Incubation was done at 37°C and at ambient temperature. Aliquots were removed at selected intervals and assayed for phosphatase activity (11). Percent activity was determined versus a HPAP control incubated without phenylglyoxal.

Restriction mapping of the clones revealed the presence of two internal EcoRI sites (Fig. 2), which probably account for the low frequency of full length HPAP cDNAs and the difficulty of isolating clones with intact 5' ends (7, 8). Additional restriction mapping with HindIII, PstI, BgII, and SacI yielded the map shown in Fig. 2. For sequencing purposes, the inserts from the positive clones were subcloned into the EcoRI site of pUC18. Sequencing utilized universal (U) and reverse (R) primers to access both ends of the insert. Additional regions of the DNA inserts were accessed by utilizing the HindIII, BgII, and SacI sites to generate deletions which were then sequenced with U or R primers. Remaining gaps were completed by utilizing HPAP-specific primers (Fig. 2). The present sequence showed that 13 of the 78 nucleotide assignments in the 78-mer were incorrect. The isolation of HPAP cDNAs with this probe using high stringency washes despite these mismatches indicates the value of such long, nondegenerate oligonucleotide probes.

Several differences existed between the present sequence (Fig. 3) and those described earlier (7, 8). First, a clone was isolated that contained the entire protein coding region, with no deletions or insertions. Second, the leader peptides differ at 10 consecutive amino acid residues due to a deletion/insertion at position 60/92 of the nucleotide sequence of Ref. 7. Third, a sequence of 8 amino acids starting at amino acid 34 in the coding region of the protein would differ as a consequence of an insertion/deletion in the cDNA sequence in Ref. 7. Fourth, the present nucleotide sequence data clearly indicated the presence of a Pro and Cys at positions 180 and 340 of the protein sequence, respectively. Substitutions involving Pro are frequently important for structural reasons, while an error involving Cys would be critical with respect to disulfide bond formation. Fifth, several minor differences were noted in the noncoding 3' region. Finally, some other isolated inconsistencies (discussed below) were resolved by the present cDNA sequence results.

Both pr4 and pr13 hybridized to the 78-mer and also to RVE-1, indicating that they contained the majority of the HPAP coding region. Sequencing showed that they had different 5' and 3' end regions, but otherwise the regions coding for the mature HPAP were identical. The leader region of pr4 contained an inversion, relative to pr13, with one end point within the leader sequence and the other one at the 5' end of pr4. The inverted fragment contains the initiation codon and therefore results in a disruption of the reading frame. Reversal of the inversion would, on paper, generate a sequence that would be consistent with pr13. Presumably the inversion event occurred during the cDNA cloning process and was facilitated by a potential stem and loop structure that could

Note also that Ref. 8 uses the nucleotide sequence coding for the first 60 amino acids from Ref. 7.
sequence (data not shown). Furthermore, the location of the irregular regions corresponds to the positions of introns in the homologous HLP gene (31).

Calculations according to von Heijne (32) (using the program PSIGNAL of PC-Gene software) predict that the leader region of the present sequence possesses a single leader peptide cleavage site that would in fact yield the observed (5) amino-terminal sequence of the protein. In contrast, the sequence resulting from the deletion/insertion in Ref. 7 predicts four weaker cleavage sites, and cleavage at the weakest of these would be needed to give the mature protein. A recent analysis has pointed out the rarity of signal peptide cleavage at alternative or multiple sites (33).

**Protein Sulfhydryl Group Determination and Sequence Assignment**—A spectrophotometric titration of disulfide bonds showed the presence of three disulfides (per monomer) in HPA. The number of disulfide bonds determined by this procedure was the same whether the sample had been lyophilized or not. An earlier study established that the two subunits are not covalently linked (34). Nonreduced HPA was not labeled by [14C]iodoacetate, even in the presence of 6 M guanidine HCl. Reduced HPA was found to be labeled at 6 residues by the iodoacetate. It was therefore established that HPA monomer contains three disulfide bonds and no free sulfhydryl groups. The conclusion (8) that HPA contains 5 cysteines per monomer is clearly in error.

For completeness in protein sequencing, it was necessary to utilize different sets of proteolytic conditions in order to create sufficient overlaps to describe the entire amino acid sequence. Some fragments were not detected either because they were too long or too short. This problem is exacerbated in disulfide bond detection, since fragments A and B must be small enough that their combined (disulfide-bonded) fragment mass is sufficiently low to be detected, yet large enough so that the A and B fragments can be individually identified.

Compared to the entire amino acid sequence of the protein predicted from the present cDNA sequence, residues 1 to 351 were positively identified. Gas-phase peptide sequence data show that the sequence for the first 45 residues is that which is predicted by our cDNA sequence data, in agreement with reported amino acid sequence data (5, 8). We confirm the speculation (8) that deletion/insertion errors occurred in Ref. 7 which resulted in an 8-amino acid translation error in the region of the mature protein. The additional sequence differences between our work and Sharief et al. (8) appear to be due to reading and/or translation errors in Ref. 8. 4 Our cDNA sequence, together with our protein data demonstrating the presence of three disulfide bonds and direct sulfhydryl titration data which show no free sulfhydryls, establishes that the substitution of a Val for a Cys residue at position number 340 of the expressed protein does not occur. HLP (15), a homologous protein, also has a Cys in this position, further strengthening this point. Finally, gas phase sequencing of the [14C]cysteine labeled fragment showed conclusively that this residue is a Cys and not a Val, since the residue corresponding to amino acid 340 co-eluted with carboxymethylcysteine and was radiolabeled. Sequencing another fragment from the same mixture demonstrated that residue 180 is Pro and not Ala. The only remaining discrepancies seem to involve whether or not some Gln or Asn residues are present as the amide form or have been deamidated to Glu or Asp.

**Presence of Post-translational Deamidation**—The results of the FAB-MS FAB-MS sequence data show in some cases that Gln or Asn has been converted to Glu or Asp. Presence of Post-translational Deamidation—The results of the FAB-MS sequence data show in some cases that Gln or Asn have been deamidated to Glu or Asp.

---

*In all but two of these instances, Ref. 8 lacks agreement between its nucleotide sequence and amino acid translation.*

---

**FIG. 3. Nucleotide sequence of HPA cDNA and the amino acid sequence of the protein.** The entire sequence was derived from six overlapping clones (pr4, pr6, pr8, pr9, pr12, and pr13). The sequence of the coding region was independently derived from pr4 and pr13. All regions were sequenced in both directions. The majority of the coding region was sequenced at least twice (in clones pr4, pr6, pr8, and pr13). The numbering of the amino acids begins with the first amino acid of the mature protein (5). The form between the inverted repeats which subend the end points of the inversion (TGGCCCAG....CTGGCCA). It is not known whether the latter, structure (or another one, located within the boundaries of the inversion) may be involved in modulating the expression of HPA in vivo.

Two other clones (pr5 and pr12) were truncated at the 5' end, with no obvious explanation (e.g. position of an EcoRI site or presence of poly(A) tract). Comparison of the terminal sequence with that of pr13 revealed a region of irregularity. Inspection of this region suggested that it was a remnant of an intron-exon junction. In fact, the consensus sequence for a 3' splice site (30) was found at the end of the irregular
Disulfide-bonded and reduced peptide fragments of human prostatic acid phosphatase determined by FAB-MS

Observed peptide masses are accumulated by the calculated values (shown in parentheses). Note that the values in this table are MH⁺ values.

<table>
<thead>
<tr>
<th>Disulfide-containing peptides</th>
<th>Disulfide MH⁺</th>
<th>Fragment A MH⁺</th>
<th>Fragment B MH⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>128–131, 340–350</td>
<td>1694.7 (1694.7)</td>
<td>488.9 (489.2)</td>
<td>1208.3 (1208.5)</td>
</tr>
<tr>
<td>180–184, 281–285</td>
<td>1223.0 (1223.6)</td>
<td>624.4 (624.3)</td>
<td>602.7 (602.3)</td>
</tr>
<tr>
<td>312–323</td>
<td>1257.1 (1256.6)</td>
<td>1259.0 (1258.6)</td>
<td>1259.0 (1258.6)</td>
</tr>
</tbody>
</table>

* Numbers refer to amino acids of the mature protein (Fig. 3).

**POTENTIAL ACTIVE SITE ANALOGOUS PEPTIDES**

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Analogous Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAPHUMAN</td>
<td>KEKFTYFLFRGQGDNLQDTFP</td>
</tr>
<tr>
<td>LAPHUMAN</td>
<td>KEKFTYLYFRHGDSSLKQYP</td>
</tr>
<tr>
<td>LAPLVRAT</td>
<td>KEKFTYLYFRHGDSSLKQYP</td>
</tr>
<tr>
<td>PPASCOUL</td>
<td>LEKESV1ISVRSDVATKATQ</td>
</tr>
<tr>
<td>PPAISOCHO</td>
<td>CKKQITFEQGSSNTQGNA</td>
</tr>
<tr>
<td>PHOSYREAST</td>
<td>CEKIQGMEMRHGEYTVSKG</td>
</tr>
<tr>
<td>PPASYEAST</td>
<td>CEKIQGMEGRHGEYTVSLA</td>
</tr>
<tr>
<td>CMERAT</td>
<td>SRRSEIVPFRHEIRSNVSG</td>
</tr>
<tr>
<td>FPAPASCOUL</td>
<td>VERNDDPYDGMHGGRANVWL</td>
</tr>
</tbody>
</table>

**Fig. 4.** Highly conserved peptide sequence in acid phosphatases and two other proteins. The proteins are (in descending order): human prostatic, human lysosomal, rat lysosomal, E. coli, and three yeast acid phosphatases, as well as the rat sodium channel protein and E. coli penicillin binding protein.

to Glu or Asp. A few of these may represent true post-translational modifications. In some instances the FAB-MS data were ambiguous since the mass accuracy of the instrument was ± 1 mass unit. This difference coincidentally corresponds to the change in mass of a peptide in which a residue has been deamidated. A deamidated residue is detected by the presence of a mass peak that is 1 mass unit above the value calculated assuming the presence of the amide. Peptides that contain this residue will be consistently detected in the FAB-MS spectra at 1 mass unit higher than expected. It seems likely that some of the differing isolectric forms are due to deamidations.

**Nonspecificity of Cleavage Reactions**—Our data show the presence of nonspecific cleavages, particularly those involving V8 protease. These cleavages occur regularly at Asp-X, Gln-X, Asn-X, and occasionally Ser-X residues. It has been mentioned that Asp-X cleavage occurs only at pH 7.8 (35) and in the presence of phosphate, and it should be noted that trace levels of phosphate remain from the protein preparation (shown by plasma desorption results). Several examples of these nonspecific cleavages have been noted by others (27, 36).

**Elucidation of Disulfide Bonding Pattern**—Mass spectrometry is proving to be extremely useful in establishing disulfide bonding patterns (37). In order to positively identify disulfide bond linkages by the procedures used here, three conditions must be satisfied. 1) The S-S fragment molecular ion mass must be found; 2) the individual sulfhydryl fragments must be found; and 3) all three masses must be identified in the same sample, since the reductions take place in the sample matrix. The disulfides that were identified are shown in Table 1.

All samples were searched for potential peaks. Many samples contained peaks satisfying one or two of the aforementioned conditions, but not all three. All peaks corresponding to S-S fragment ion masses (without component sulfhydryl peaks) were found to match one of the disulfides. No evidence of any other disulfide linkages was found. The performance of "in sample" reduction in the matrix after HPLC separation has the advantage of reducing the likelihood of reoxidation occurring to form new disulfide linkages not present in the native enzyme.

**Identification of Carboxyl-terminal Residues**—The spectra from both FAB-MS and plasma desorption MS of CNBr-treated HPAP demonstrate the presence of multiple carboxyl residues. HPAP fragments were found ending in residues 349, 350, and 351. Two of the fragments could result from post-translational processing by an aspartyl/glutamyl protease, as has been noted elsewhere for HLAP (38). Morris et al. (13) have suggested that variable carboxyl-terminal residues may be responsible for some of the differing isoelectric forms of HPAP.

**Identification of Surface Tyrosyl Residues**—Tryptic and V8 cleavage of 125I-labeled HPAP yielded four major radioactive peaks in the HPLC chromatogram of each mixture, representing approximately 90% of the total radioactivity. No other fraction contained a significant peak of radioactivity. Gas phase sequencing of the labeled fragments found label present at Tyr-52 and Tyr-57. An antigenic site predicted by the method of Hopp and Woods (39) (using the program ANTI-GEN of PC-Gene software) indicated that residues 54–59 constitutes the most likely epitope region for HPAP. However, this region is adjacent to Asn-62, one of the three sites of N-glycosylation that have been experimentally confirmed in human prostatic acid phosphatase (40). It is conceivable that the carbohydrate may interfere with antibody formation or binding.

**Identification of an Active Site Arginine Residue**—Inhibition studies with phenylglyoxal indicated that phenylglyoxal is extremely reactive toward HPAP. Remarkably, even a 5-fold molar excess of phenylglyoxal at pH 8.0 completely inactivated phosphatase activity within 5 min at 37 °C. (A control sample without phenylglyoxal remained fully active.) At this temperature and pH, the adduct that was formed was unstable with regard to label retention. Lowering the temperature to room temperature and lowering the pH to 7.0 resulted in better retention of label. The inactivation was also much slower under these conditions: after 40 min, the phosphatase activity was reduced to approximately 50% of that of the control. Gas phase sequencing identified the labeled residue as Arg-54. The presence of 20 mM L-tartrate, a strong competitive inhibitor (Ki = 20 ,um), prevented incorporation of the label. Dialysis to remove L-tartrate and phenylglyoxal restored the activity of the enzyme. In the absence of L-tartrate, the enzyme became covalently labeled and dialysis at pH 4.5 did not lead to a recovery of activity. Incorporation of approximately 0.55 mol of labeled phenylglyoxal per mol of HPAP was coincident with a loss of 49% of the enzyme
activity. We therefore conclude that Arg-54 is at or near the active site of HPAP.

Comparison of this region of the HPAP sequence with other phosphatases allows us to propose a potentially critical sequence in some of these enzymes. The subsequence RKKRYR (residues 54–58) is homologous to a RQRYH subsequence found in HAP (15) and the corresponding rat liver enzyme (41). The substitution of a His for Arg in the latter appears to be conservative, since at acidic pH, His will retain a positive charge. It is perhaps not surprising that a similarity exists between HPAP and the HAP sequences, given that these enzymes are otherwise homologous. A more interesting homology is that demonstrated with protein tyrosine phosphatases. Protein tyrosine phosphatases from rat brain (42) and human placenta (43) contain a RNYR subsequence. The striking aspect of this observation is that the remainder of the protein tyrosine phosphatase sequences show no homology with HPAP. Perhaps these positively charged regions may interact with the negatively charged phosphate group in the phosphatase reaction.

Other Potential Active Site Residues—Phosphoenzyme trapping experiments have indicated the presence of an essential histidine residue that acts as a nucleophile (44). Searching the PC-Gene SWISS-PROT database for peptides related to the histidine-containing regions in HPAP revealed that an identical RHGXRP sequence was present in human lysosomal, rat liver lysosomal, yeast, and even Escherichia coli acid phosphatases (Fig. 4). It is striking that this includes not only residues (Arg and His) that have been identified earlier as active site residues (44), but also ones (Gly and Pro) that are frequently conserved for structural reasons. Finally, it has recently been concluded that acid phosphatases are distantly related to phosphoglycerate mutase and phosphofructokinase/fructose bisphosphatase (45). The latter study, particularly in combination with the present disulfide bonding pattern, may open the way to structural modeling predictions.

Acknowledgments—We thank Prof. K. von Figura for sharing with us results on the sequence of human lysosomal acid phosphatase prior to their publication (15). We thank Dr. Peter Huber for his kindness and his help in obtaining the enzyme, and Dr. Karl Wood and Prof. David Smith for their help with the mass spectrometry.

REFERENCES

3. Huggins, C., and Hodges, C. V. (1941) Cancer Res. 1, 293–297

### Table: Covalent Structure of Prostatic Phosphatase

The calculated mass values are those achieved using the protonic sequence and the corresponding Q-values, and are V-values (other than P-values) for mass of comparison with the calculated M-value. The observed values have all been corrected by subtraction of one 106 unit.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence</th>
<th>Fragment_ID</th>
<th>M (Experimental)</th>
<th>M (Theoretical)</th>
<th>V (Experimental)</th>
<th>V (Theoretical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>FSH</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-3</td>
<td>LEU</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-4</td>
<td>ALA</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-5</td>
<td>LYS</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-6</td>
<td>GLU</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-7</td>
<td>GLN</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-8</td>
<td>PRO</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-9</td>
<td>GLN</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-10</td>
<td>ARG</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-11</td>
<td>LYS</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-12</td>
<td>LEU</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-13</td>
<td>LYS</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-14</td>
<td>VAL</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td>1-15</td>
<td>ASP</td>
<td>1</td>
<td>1234</td>
<td>1234</td>
<td>1.15</td>
<td>1.15</td>
</tr>
</tbody>
</table>

**References:**
1. Covalent Structure of Prostatic Phosphatase
2. Supplementary Material: Precipitated and Observed Masses of Peptide Fragments
3. Covalent Structure, Dissolved Protein, and Identification of cellular Surface and Antigenic Sites of Human Prostatic Acid Phosphatase
4. Robert L. van der Kleef, Rhee D.W., Jorgensen E., Marcheau R., and G. Lamont
5. Supplementary Material: Precipitated and Observed Masses of Peptide Fragments
6. The calculated mass values are those achieved using the protonic sequence and the corresponding Q-values, and are V-values (other than P-values) for mass of comparison with the calculated M-value. The observed values have all been corrected by subtraction of one 106 unit.