Vol. 186, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1992 Pages 122-128

COMPUTER MODELING OF TWO INORGANIC PYROPHOSPHATASES

Mauno Vihinen, Maria Lundin* and Herrick Baltscheffsky#

Department of Biochemistry and Centre for Biotechnology, University of Turku, SF-20500 Turku, Finland

*Ludwig Institute for Cancer Research, Uppsala Branch, Biomedical Center, Box 595, S-751 24 Uppsala, Sweden

#Department of Biochemistry, Arrhenius Laboratories for Natural Sciences, University of Stockholm, S-106 91 Stockholm, Sweden

Received April 6, 1992

The yeast *Saccharomyces cerevisiae* has two inorganic pyrophosphatases that are structurally related. One, PPA1, is a cytoplasmic enzyme. The other, PPA2, is located in the mitochondria and appears to be energy-linked. The sequence similarity of PPA1 and PPA2 is about 66 % and the identity is about 50 %. All amino acids known to be important for catalysis are conserved, except one glutamate which is substituted by an aspartate in PPA2. The structures of PPA2 and the cytoplasmic PPase from *Schizosaccharomyces pombe* were modeled based on the three dimensional structure of PPA1. Two cysteines in PPA2 and one in the *S. pombe* enzyme are located at the catalytic cleft. Four residues form an unique insertion near the entrance of the catalytic cleft in the mitochondrial enzyme.

Inorganic pyrophosphate (PP_i) is produced as a by-product in a number of metabolic reactions *e.g.* RNA- and DNA-synthesis, amino acid activation, and fatty acid metabolism [1]. Inorganic pyrophosphatases (E.C.3.6.1.1.) hydrolyze PP_i in the presence of Mg^{2+}. There exist both membrane-bound and cytoplasmic PPases. The photosynthetic bacterium *Rhodospirillum rubrum* contains an energy-linked, intrinsic membrane PPase [2]. Membrane bound PPases also exist in organelles of eukaryotes. Mitochondria contain an energy-linked, membrane-bound PPase-complex whose catalytic part appears to be active also in a soluble form [3-5]. The primary structure of the soluble, cytoplasmic *Saccharomyces cerevisiae* PPase has been determined by both protein-sequencing [6] and deduced from the DNA sequence [7]. The three dimensional structure of yeast cytoplasmic PPase, PPA1, has been determined by X-ray

Abbreviations used: PP_i, inorganic pyrophosphate; PPase, inorganic pyrophosphatase; PPA1, cytoplasmic PPase of *S. cerevisiae*; PPA2, mitochondrial PPase of *S. cerevisiae*.

0006-291X/92 $4.00

crystallography at 3.0 Å resolution [8, 9]. Some other known sequences of cytoplasmic PPases are from *Kluyveromyces lactis* [10], *Schizosaccharomyces pombe* [11], *Escherichia coli* [12] and the thermophilic bacterium PS3 [13]. Mutational and chemical modification studies on yeast cytoplasmic and *E. coli* PPases, have ascertained residues important for function [14, 15]. We have recently cloned and sequenced a gene, *PPA2*, coding for the catalytic subunit of a mitochondrial PPase from *S. cerevisiae* [16].

The similarity between amino acid sequences of PPA2 and PPA1 is 65.8 %. The mitochondrial enzyme appears to have an energy-linked function and would thus be associated with a membrane complex [5, 16]. The most striking differences between the amino acid sequences are one unique cysteine adjacent to an insertion of four residues, and two adjacent substitutions of cysteines in PPA2. These differences are on either side of a stretch of conserved residues important for catalysis [16]. One of the two adjacent cysteines was also found in the *S. pombe* sequence. In order to locate these differences within the three dimensional structure, we modeled the structures of PPA2 and the *S. pombe* cytoplasmic PPase based on the coordinates of PPA1.

MATERIALS AND METHODS

Database searches were according to [17] as implemented in the GCG program package [18]. Sequence alignment was according to [19] and multiple sequence alignment according to Eurants and Vihinen (in preparation). Multiple sequence analysis and comparison of secondary structural predictions were according to [20]. Secondary structure predictions were according to [21] and hydropathy according to [22].

The coordinates of PPA1 [10] were from Brookhaven database. The secondary structural regions of PPA1 were determined with the program DSSP [23]. The molecular modeling was performed on Evans & Sutherland ESV 30 workstation.The modeling software package Sybyl was from Tripos Associates, Inc. (St. Louis Missouri). The models were tested with the program Poldiag [24].

Mitochondria were isolated from a PPA2 overproducing strain described [7]. The matrix fraction was the supernatant obtained after centrifugation of sonicated mitochondria in 150 000 x g for 45 min. The PPase activity was measured according to [25].

RESULTS AND DISCUSSION

Several PPase sequences have recently been determined. Database searches did not show any other proteins with significant sequence similarities to the PPases. The sequence for PPA1 derived from the DNA sequence differs somewhat from that determined by protein sequencing, mainly in that one of two adjacent Lys residues is absent in the latter. The similarities and identities of the known *S. cerevisiae* and *S. pombe* sequences are shown in Fig.1. The two bacterial enzymes are about 100 residues shorter than the four yeast sequences but those residues found to be important for activity and metal binding are conserved in all the sequences [16]. The main difference in functional residues between PPA1 and PPA2 is the replacement of a glutamate 185 by aspartate. The major differences between all PPases occur in the C-terminal part. This region is a loosely packed loop on the surface of PPA1 and is absent in the bacterial proteins.

```
                                                                        7
SOL2  ..............................TYTTRQIGAKNTLEYKVYIE.KDGKPVSAFHDIPLYADKENN
SPO   ..............................MSEYTTREVGALNTLDYQVYVE.KNGTPISSWHDIPLYANAEKT
MIT   MNLLRMNALTSKARSIERLKQTLNILSIRNHRQFSTIQQGSKYTLGFKKYLTLLNGEVGSFFHDVPLDLNEHEK
SOL1  ..............................TYTTRQIGAKNTLEYKVYIE.KDGKPVSAFHDIPLYADKEDN

            *       * *                   *           *   *                  15
SOL2  FNMVVEIPRWTNAKLEITKEETLNPIIQDTKKGKLRFVRNCFPHHGYIHNYGAFPQTWEDPNVSHPETKA....
SPO   LNMVVEIPRWTQAKLEITKEATLNPIKQDTKKGKLRFVRNCFPHHGYIWNYGAFPQTYEDPNVVHPETKA....
MIT   VNMIVEVPRWTTGKFEISKELRFNPIVQDTKNGKLRFVNNIFPYHGYIHNYGAIPQTWEDPTIEHKLGKCDVAL
SOL1  FNMVVEIPRWTNAKLEITKEETLNPIIQNT.KGKLRFVRNCFPHHGYIHNYGAFPQTWEDPNVSHPETKA....

          *    *                          ** * * *                          22
SOL2  GDNDPIDVLEIGETIAYTGQVKQVKALGIMALLDEGETDWKVIAIDINDPLAPKLNDIEDVEKYFPGLLRATNEV
SPO   GDSDPLDVCEIGEARGYTGQVKQVKVLGVMALLDEGETDWKVIVIDVNDPLAPKLNDIEDVERHMPGLIRATNEV
MIT   GDNDPLDCCEIGSDVLEMGSIKKVKVLGSLALIDDGELDWKVIVIDVNDPLSSKIDDLEKIEEYFPGILDTTRE
SOL1  GDNNPIDVLQIGETIAYTGQVKEVKALGIMALLDEGETDWKVIAIDINDPLAPKLNDIEDVEKYFPGLLRATDE

             **    *                                                        30
SOL2  FRIYKIPDGKPENQFAFSGEAKDKKYALDIIKETHDSWKQLIAGKSSDSKGIDLTNVTLPDTPTYSKAASDAIPI
SPO   FRIYKIPDGKPENSFAFSGECKNRKYAEEVVRECNEAWERLITGKTDAKSDFSLVNVSVTGSVANDPSVSSTIPI
MIT   FRKYKVPAGKPLNSFAFHEQYQNSNKTIQTIKKCHNSWKNLISGSLQE.........KYDNLPNTERAGNGVTLI
SOL1  FRIYKIPDGKPENQFAFSGEAKNKKYALDIIKETHNSWKQLIAGKSSDSKGIDLTNVTLPDTPTYSKAASDAIPI


SOL2  ASLKADAPIDKSIDKWFFISGSV
SPO   AQELAPAPVDPSVHKWFYISGSP
MIT   DSVKPPSQIPPEVQKWYYV....
SOL1  ASPKADAPIDKSIDKWFFISGSV
```

Fig. 1. Alignment of the sequences for *S. cerevisiae* soluble cytoplasmic PPase (SOL1 is protein sequenced and SOL2 is derived from DNA sequence), *S. cerevisiae* mitochondrial PPase (MIT) and *S. pombe* cytoplasmic PPase (SPO). The secondary structural elements according to Terzyan *et al.* [9] are indicated in the SOL1 sequence and those found by analysis with the program DSSP [23] are indicated in the SOL2 sequence. Underlining denotes β-sheets and bold underlining denotes α-helices.

The sequences were analysed both on the primary and secondary structural levels. No significantly similar sequences, except the PPase sequences, were found on sequence database searches. The sequence of the *S. pombe* enzyme is very similar (80.7 %) to that of PPA1, whereas PPA2 is only moderately similar (65.8 %) (Table I). Both enzymes are similar enough to PPA1 for computer aided molecular modeling. Most of the dissimilar residues in the *S. pombe* enzyme are located in the C-terminal region whereas they are more scattered along the entire sequence of PPA2. Alignment of the *S. pombe* sequence with that of *S. cerevisiae*, required only one gap of one residue.

Alignment of PPA1 and PPA2 requires four gaps including three insertions in PPA2 of one, one and four residues, respectively, and one deletion of nine residues, to the cytoplasmic enzyme (Fig. 1). These gaps are not found within the highly conserved regions and do not disturb secondary structural elements (Fig.1). Terzyan *et al.* [10] have presented the existence of numerous β-strands and α-helices in the PPA1 3D-structure, whereas the program DSSP (23), which analyzes several features based on geometry of the molecule, indicates only a few secondary structures (see Fig. 1). Secondary structural regions are generally very conserved, also here.

TABLE I. Sequence similarities and identities of PPases. Similarities are shown in the lower triangle and identities in the upper. Sequences were aligned so that the conserved regions were on the main diagonal when compared with the program MULTICOMP [19].

	1	2	3	4	5	6
1		96.8	48.5	68.1	82.5	21.3
2	99.6		48.7	69.6	84.3	21.8
3	65.8	65.9		48.7	50.4	22.4
4	80.7	80.8	66.2		71.1	20.6
5	89.8	90.2	66.8	80.5		26.3
6	43.1	43.1	43.7	40.0	48.6	

1) *S. cerevisiae*, cytoplasmic (protein sequence) [6].
2) *S. cerevisiae*, cytoplasmic (derived from DNA sequence) [7].
3) *S. cerevisiae*, mitochondrial [15].
4) *S. pombe*, cytoplasmic [11].
5) *Kluyveromyces lactis*, cytoplasmic [10].
6) *E. coli* [12].

The 3D-structure of cytoplasmic yeast PPase at 3 Å resolution was used as a basis for modeling PPA2 and the *S. pombe* cytoplasmic PPase. The backbone and space filling representations including side chain atoms of the PPA1 molecule and the two modeled PPase structures are shown in Fig. 2 and 3. The high sequence similarity of the *S. pombe* cytoplasmic PPase and PPA1 made the modeling of the former very straightforward. There was only one insertion of one residue and the replacements usually involved chemically similar residues. Modeling of the mitochondrial enzyme required building of four loops and several substitutions. The models were minimised using Amber force fields (26) as implemented in the program Sybyl. In the beginning stage, the Kollman united force field was used and in the final stage, the Kollman all parametrization. Reliability of the models was estimated with the program Poldiag [24], which analyses the distribution of residues between the surface and the interior of the protein. Analysis of the distribution of polar and nonpolar residues indicates that the PPA2 model is acceptable on all the analyzed levels. The minimized structures of PPA1 and *S.pombe* PPase have exceptional polar fractions. PPA1 is a dimeric enzyme and when the dimer was analyzed normal distribution of residues was obtained because the residues of the interface being on the surface in the monomer were hided in the dimer. There is no biochemical data on the*S. pombe* enzyme, but exceptional distribution of charge could suggest a multimeric structure. The relative minimized energies, which are of the same magnitude in the known and modeled structures, also suggest that the models are reliable. Because of the much higher sequence similarity the model of the *S. pombe* enzyme is presumably more reliable.

The model of the mitochondrial enzyme is most accurate in the region of the active site cleft which possesses the highest sequence similarity. The insertions and the deletions that had to be built into the model are exclusively on the surface of the protein

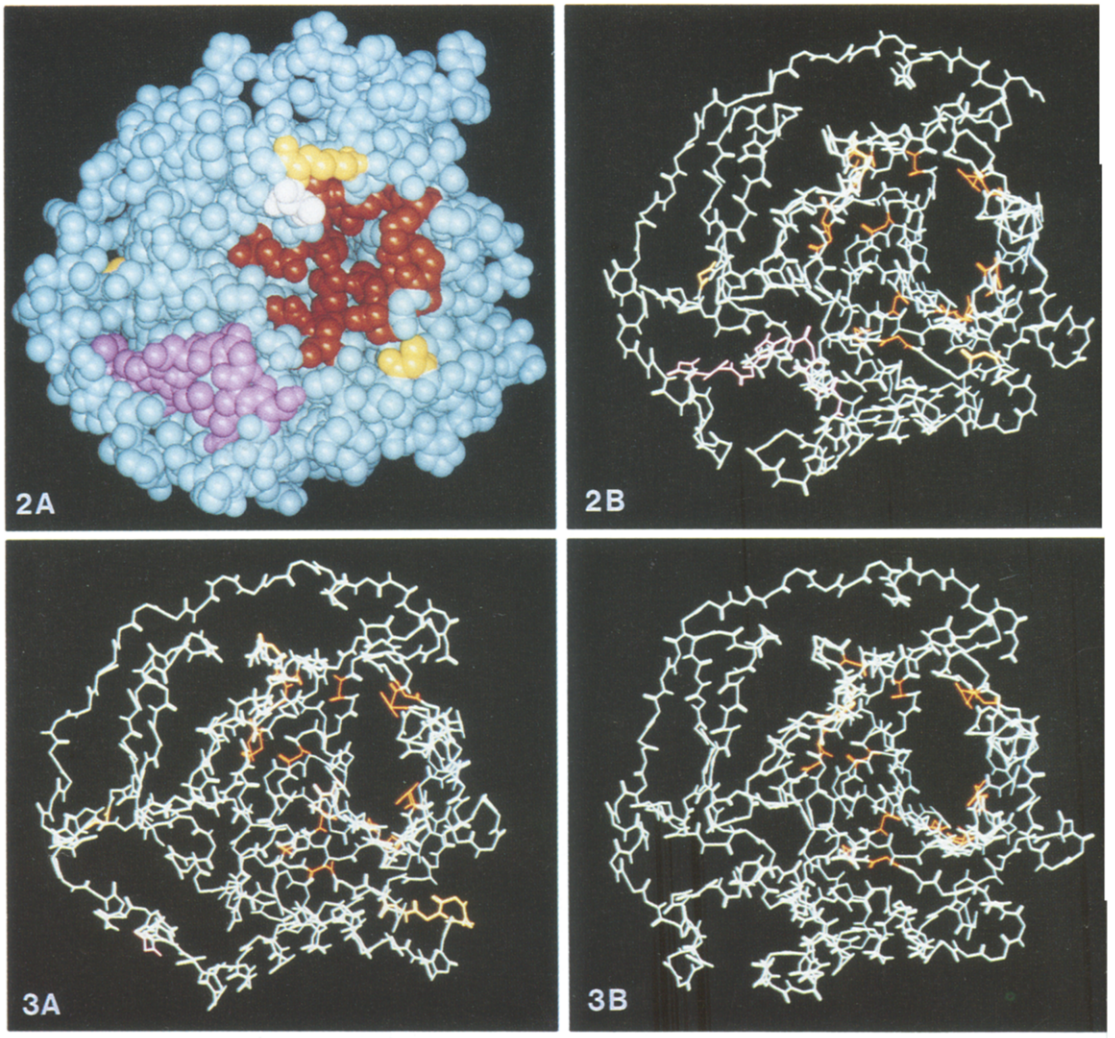

Fig. 2. Structure of the *S. cerevisiae* cytoplasmic PPase (PPA1) A) on space fill presentation and B) only backbone atoms. Sites of insertions in the mitochondrial enzyme are in yellow and deletion in magenta. The insertion site in S. *pombe* is in white. Residues known to be important for activity [14] are in red.

Fig. 3. Backbone of A) yeast mitochondrial and B) *S. pombe* PPase structural models. The counterpart for catalytically important residues in yeast soluble PPase are in red. Inserted amino acids are in yellow and deletions in magenta. Cysteines 158 and 159 in the yeast mitochondrial PPase are in white and cysteine 145 is in green.

(Fig.1 and Fig. 2). The overall structure of the PPA2 model does not differ very much from the PPA1 three dimensional structure. The modeling revealed that two Cys residues, 158 and 159, are present in the bottom of the catalytic cleft of PPA2. The second of these cysteines is also found in the *S. pombe* protein. The tight structure around these residues suggests that they are important for the structure. The four-residue insertion in PPA2 is very close to a highly conserved sequence 151-157, which lines the

active site inner surface (Fig. 1 and Fig. 2). Adjacent to the insert, one cysteine is located at position 145. These four inserted amino acids extend a loop near the entrance to the active site (Fig. 1. and Fig. 2.). It can be noted that a highly conserved His residue is situated at the tip of this large loop. Between the His residue and one highly conserved Lys residue, adjacent to the four residue insertion, there are three unique substitutions in PPA2. In contrast, this sequence is identical in the three cytoplasmic yeast PPases. The critical location of the four residue insertion, just at the entrance to the catalytic cleft, and their unique appearance in the mitochondrial enzyme suggest them as possible candidates for the assumed interaction with a membrane protein and/or a coupling of hydrolysis of PPase to proton-pumping.

An important question is whether the unique Cys residues flanking the active site of PPA2 might form a disulfide bridge. Since the distance between the Cys 145 and the Cys 159 in the model is not less than 20 Å, a disulfide bridge between them seems unlikely.

The model of the *E. coli* PPase (Vihinen, unpublished data) and the present models suggest that the folding of the core of PPases (about the size of bacterial enzymes [14]) is very conserved and the N- and C-terminal extensions have been introduced later during evolution. The terminal loops are very flexible whereas the conserved core structure is more rigid and stable (14). These terminal regions of the PPases may well be dispensable also from the eukaryotic enzymes. We have investigated the effect of the protease trypsin on the PPA2 protein. Trypsin cleaves polypeptides specifically after Arg and Lys residues. Such residues are found in the terminal regions of PPA2.

We measured the PPase activity in the presence and absence of trypsin of the mitochondrial matrix fraction since most of the PPA2 is found in the matrix after subfractionation of mitochondria, which includes sonication. Cells overproducing PPA2 were used in this experiment. After pre-incubation of the matrix fraction with trypsin for 30 min at 0°C the PPase activity was stimulated 2-3 fold (not shown). It is likely that under these conditions trypsin cuts the PPA2 protein at the amino terminus and carboxy terminus. The PPase activity is not only retained but stimulated by the trypsin treatment.

The apparently unique role, among the PPases sequenced, of the yeast mitochondrial PPase in membrane-linked energy transfer may well have led to the insertions and substitutions of residues around the active site, conferring to this membrane-associated enzyme an energy transferring capacity as well as a binding capacity to the mitochondrial membrane counterpart.

Acknowledgment: This work was supported by the Nordic Yeast Research Program (NYRP).

REFERENCES

1. Kornberg, A. (1962). In Horizons in biochemistry (Kasha, H. and Pullman, P., eds.) Academic press, New York 251-264.

2. Baltscheffsky, M. and Nyrén, P. (1984) In New Comprehensive Biochemistry: Bioenergetics (Ernster, L., ed.) vol 9, pp. 187-206, Elsevier Amsterdam.
3. Volk, S.E., Baykov, A.A., Kostenko, E.B., and Avaeva, S.M. (1983) Biochim. Biophys. Acta 744, 127-134.
4. Mansurova, S.E. (1989) Biochim. Biophys. Acta 977, 237-247.
5. Lundin, M., Deopujari, S.W., Lichko, L., Pereira da Silva, L., and Baltscheffsky, H. (1992) Biochim. Biophys. Acta., 1098, 217-223.
6. Cohen, S.A., Sterner, R., Keim, P.S., and Henrikson, R.L. (1978) J. Biol. Chem., 253, 889-897.
7. Kolakowski, Jr.L.F., Schloesser, M., and Cooperman, S., (1988) Nucleic Acids Res. 22, 10441-10452.
8. Kuranova, I.P., Terzyan, S.S, Voronova, A.A., Smirnova, E.A., Vainshtein, B.K., Höhne, W.E., and Hansen, G. (1983) Bioorg. Khim. 9. 1611-1619.
9. Terzyan, S.S., Voronova, A.A., Smirnova, E., Kuranova, I.P., Nekrasov, Y.V., Arutyunyun, E.G., Vainstein, B.K., Höhne, W., and Hansen, G. (1984) Bioorg. Khim. 10, 1469-1482.
10. Stark, M., J.R., and Milner, J.S. (1989) Yeast, 5, 35-50.
11. Kawasaki. I., Adachi. N., and Ikeda, H. (1990) Nucleic Acids Res. 18, 5880.
12. Lahti, R., Pitkäranta, T., Valve, E., Ilta, I., Kukko.Kalske, E., and Heinonen, J. (1988) J. Bact. 170, 5901-5907.
13. Ichiba T., Takenaka, O., Samejima, T., and Hachimori, A. (1990) J. Biochem. 108, 572-578.
14. Lahti, R., Kolakowski, Jr.L.F., Heinonen, J., Vihinen, M., Pohjanoksa, K., and Cooperman, B.S. (1990) Biochim. Biophys. Acta. 1038, 338-345.
15. Lahti, R., Pohjanoksa, K., Pitkäranta, T., Heikinheimo, P., Salminen, T., Meyer, P., and Heinonen, J. (1990) Biochemistry 29, 5761-5766.
16. Lundin, M., Baltscheffsky, H., and Ronne, H. (1991), J. Biol. Chem. 266, 12168-12172.
17. Lipman, D.J., and Pearson, W.R. (1985) Science 227, 1435-1441.
18. Devereux, J.P., Haeberli, P., and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395.
19. Needleman, S.B., and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443-453.
20. Vihinen, M. (1988) Comput. Appl. Biosci. 4, 89-92.
21. Chou. P. Y., and Fasman G. D. (1978) Adv. Enzymol. 47, 45-148.
22. Kytes, J., and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
23. Kabsch, W., and Sander, C. (1983) Biopolymers 22, 2577-2637.
24. Baumann, G., Frömmel, C., and Sander, C. (1989) Prot. Engin. 2, 329-334.
25. Shatton, J.B., Ward, C., Williams,A., and Weinhouse, S.(1982) Anal. Biochem. 130, 114-119.
26. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P.K. (1984) J. Am. Chem. Soc. 106, 765-784.