The EMBO Journal vol.10 no.7 pp.1607-1618, 1991

Crystal structure of *Penicillium citrinum* P1 nuclease at 2.8 Å resolution

A.Volbeda[1,3], A.Lahm[1,4], F.Sakiyama[2] and D.Suck[1]

[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, Germany and [2]Osaka University, Institute for Protein Research, 3-2 Yamadaoka, Suita, Osaka 565, Japan

Present addresses: [3]Laboratoire d'Ingénierie des Protéines, DSV/LIP - CEN.G 85X, 38041 Grenoble Cedex, France and [4]Instituto di Richerche di Biologia Molecolare (IRBM), Via Pontina, I-00040 Pomezia (Rome), Italy

Communicated by K.C.Holmes

P1 nuclease from *Penicillium citrinum* is a zinc dependent glyco-enzyme consisting of 270 amino acid residues which cleaves single-stranded RNA and DNA into 5'-mononucleotides. The X-ray structure of a tetragonal crystal form of the enzyme with two molecules per asymmetric unit has been solved at 3.3 and refined at 2.8 Å resolution to a crystallographic R-factor of 21.6%. The current model consists of 269 amino acid residues, three Zn ions and two *N*-acetyl glucosamines per subunit. The enzyme is folded very similarly to phospholipase C from *Bacillus cereus*, with 56% of the structure displaying an α-helical conformation. The three Zn ions are located at the bottom of a cleft and appear to be rather inaccessible for any phosphate group in double-stranded RNA or DNA substrates. A crystal soaking experiment with a dinucleotide gives clear evidence for two mononucleotide binding sites separated by ~20 Å. One site shows binding of the phosphate group to one of the zinc ions. At both sites there is a hydrophobic binding pocket for the base, but no direct interaction between the protein and the deoxyribose. A cleavage mechanism is proposed involving nucleophilic attack by a Zn activated water molecule.

Key words: binding of dinucleotide/cleavage mechanism/P1 nuclease/X-ray structure/Zn enzyme

Introduction

Knowledge of the structural basis of the recognition of nucleic acids by proteins has increased considerably through a number of high resolution X-ray crystallographic studies on protein-double-stranded DNA complexes (see Steitz, 1990, for a review), and recently the first X-ray structure of an aminoacyl-tRNA synthetase complexed to its cognate tRNA was solved (Rould *et al.*, 1989). Relatively little is known, however, about the binding interactions between proteins and single-stranded DNA or RNA. This has motivated us to start an X-ray crystallographic analysis of P1 nuclease, an enzyme capable of cleaving single-stranded RNA and DNA molecules.

P1 nuclease is a glycoprotein isolated from the mold *Penicillium citrinum*. Its mol. wt is 36 kd, ~20% of which stems from covalently bound carbohydrate. P1 belongs to a family of zinc dependent nucleases with a high cleavage specificity for single-stranded nucleic acid conformations (for a review see Shishido and Ando, 1982), but which are also capable of recognizing a range of non-A/non-B/non-Z conformations in double-stranded nucleic acids (see e.g. Pulleyblank *et al.*, 1988). The enzyme is a phosphodiesterase, cleaving the bond between the 3'-hydroxyl and 5'-phosphoryl group of adjacent nucleotides, and at the same time acts as a phosphomonoesterase, removing the 3'-terminal phosphate group (Fujimoto *et al.*, 1974a). Both activities depend on the presence of zinc and both show very little sequence dependence (Fujimoto *et al.*, 1974b,c). Single-stranded nucleic acids are cleaved at least 200 times faster than double-stranded conformations. Whereas the phosphodiesterase activity is rather similar for single-stranded DNA and RNA, the 3'-nucleotidase activity appears to be significantly higher for RNA substrates. The final cleavage products of P1 are solely 5'-mononucleotides.

In its active form, P1 contains three zinc ions per molecule (Fujimoto *et al.*, 1975a). Complete removal of zinc by EDTA treatment causes a loss of activity which is only partially restorable by readdition of zinc (Fujimoto *et al.*, 1974a). This suggests an important function for the zinc ions not only in catalysis but also in stabilizing the folding of the enzyme. In a previous report we described the results of a crystallographic analysis of P1 at 4.5 Å resolution (Lahm *et al.*, 1990), showing that the Zn ions are arranged in a trinuclear cluster very similar to the Zn cluster observed in the 1.5 Å resolution X-ray structure of phospholipase C from *Bacillus cereus* (Hough *et al.*, 1989).

Recently the amino acid sequence of P1 was determined (Maekawa,K., Tsunsawa,S., Dibo,G. and Sakiyama,F., manuscript in preparation), showing that the enzyme consists of 270 residues. Two disulfide bonds and four carbohydrate attachment sites were characterized. P1 nuclease from *P.citrinum* is highly homologous to S1 nuclease from *Aspergillus oryzae*, showing ~50% sequence identity (Iwamatsu,A., Aoyama,H., Dibo,G., Tsunasawa,S. and Sakiyama,F., to be published). The close relationship is also evident from the fact that S1, like P1, needs three zinc ions for activity (Shishido and Habuka, 1986). Studies with thiophosphorylated substrates have shown that both S1 and P1 nuclease cleave the phosphodiester bond with inversion of configuration at the phosphorus. In addition they show the same stereo-selectivity for these substrates (Potter *et al.*, 1983a,b). All these similarities strongly suggest that the overall structures of the two enzymes are very closely related as well.

The X-ray structure of P1 nuclease described in this paper at 2.8 Å resolution represents the first atomic model for the class of zinc dependent single strand specific nucleases. Based on the three-dimensional structure and first results from crystal soaking experiments with a dinucleotide, we discuss the interaction of P1 with single- and double-stranded substrates and the mechanism of their cleavage.

Results

Refinement

The refined dimer of the tetragonal crystal form of P1 nuclease presented in this paper contains 269 amino acid residues, three Zn ions and two *N*-acetyl glucosamines per subunit. The C-terminal Lys270 is the only residue for which no density is observed. The crystallographic R-factor of the refined model is 21.6%, using all 22653 observed reflections between 8.0 and 2.8 Å resolution. From a Luzzati plot (Luzzati, 1952), an r.m.s. co-ordinate error of 0.35 Å is estimated, which may be compared to the r.m.s. deviation of 0.32 Å obtained after superposition of all 1956 atoms present in subunit 1 onto subunit 2. All peptide torsional angles fall in or very close to the 'allowed' regions (see Ramachandran *et al.*, 1963). The quality of the model before and after refinement is illustrated in Figure 1 by a real space R-factor plot (Brandén and Jones, 1990). Further refinement including more sugar residues and ordered water molecules is in progress.

Three-dimensional folding

The folding of P1 nuclease is characterized by a high amount of α-helices (see Figures 2 and 8). Using the classification procedure developed by Kabsch and Sander (1984), 14 α-helices may be distinguished, varying in length from four to 31 residues. Eleven of these (labeled A, B, C, E, F, G, I1, I2, J, L and N) are found in both subunits of the non-

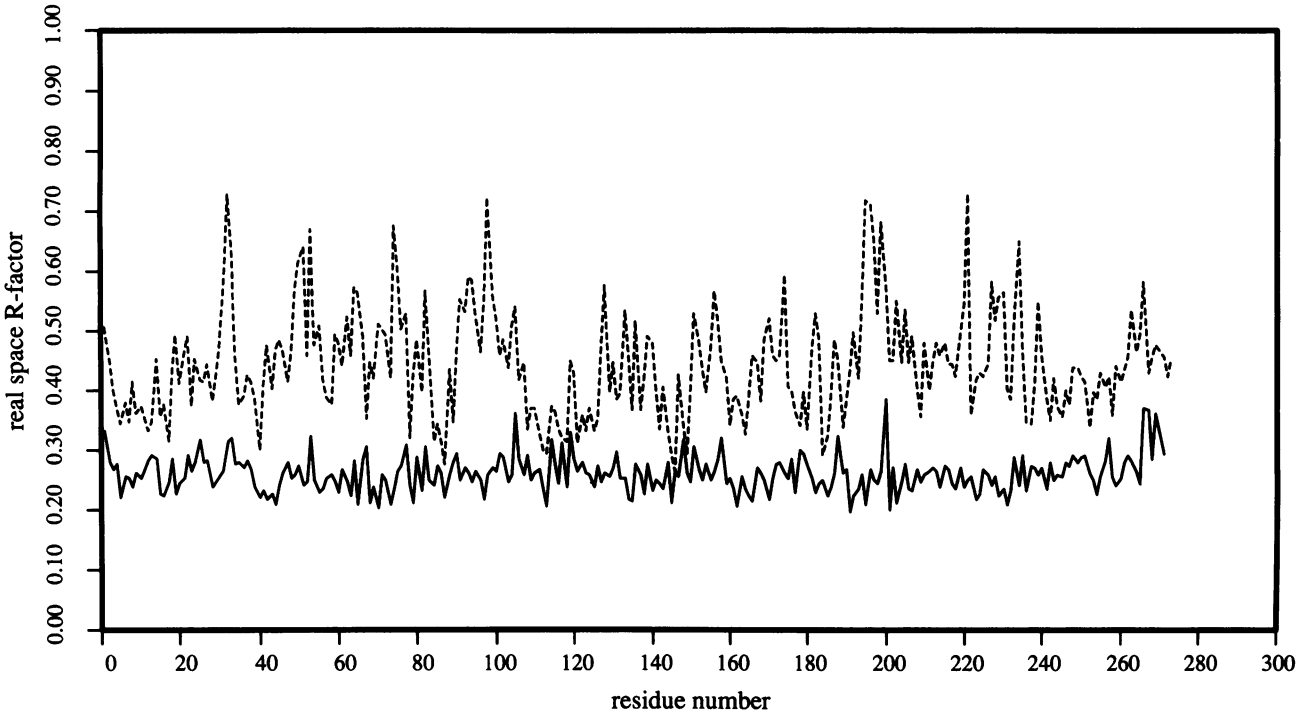

Fig. 1. Real space R-factor plot, showing the fit (averaged for all atoms) of each residue in a $2mF_o-DF_c$ map (Read, 1986) calculated for the refined model. The top curve shows the result for the unrefined starting model, the bottom curve for the refined model. Real space residuals were calculated by the program O (Jones *et al.*, 1991).

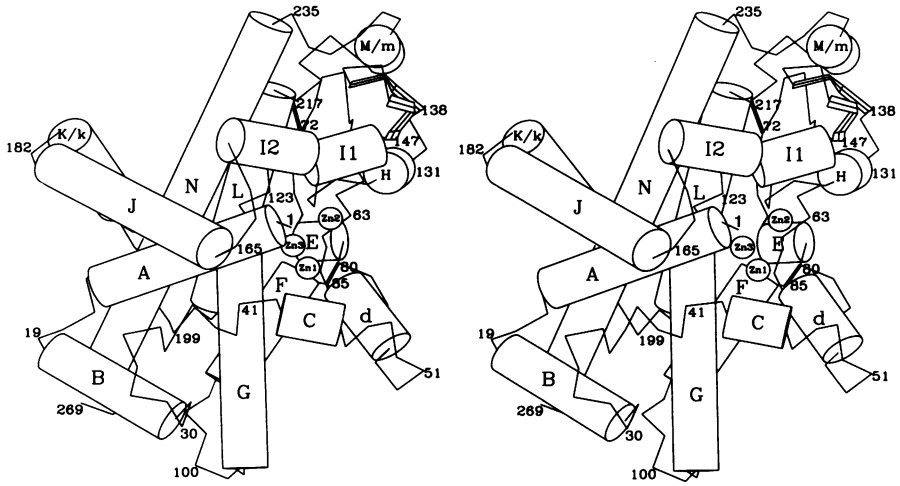

Fig. 2. ARTPLOT representation (Lesk and Hardman, 1982) of the folding of nuclease P1. Cylinders indicate α- or 3_{10}-helices containing at least four residues, as traced by the program DSSP (Kabsch and Sander, 1983). β-Structure is indicated by a thick ribbon. Heavy lines depict the two disulfide bonds connecting Cys80 to Cys85 and Cys72 to Cys217. The three metal ions labeled Zn1, Zn2 and Zn3 are represented by open spheres. α-Helices are labeled with capital letters, 3_{10}-helices by small letters. Boundaries of all secondary structure elements are given in Figure 8.

crystallographic dimer. K and M are classified as a regular α-helix in subunit 1, but as a short 3_{10}-helix (labeled k and m, respectively) in subunit 2. α-Helix H only occurs in subunit 2. The secondary structure differences between the two subunits might be correlated with different crystallographic environments. A 7 residue long 3_{10}-helix (labeled

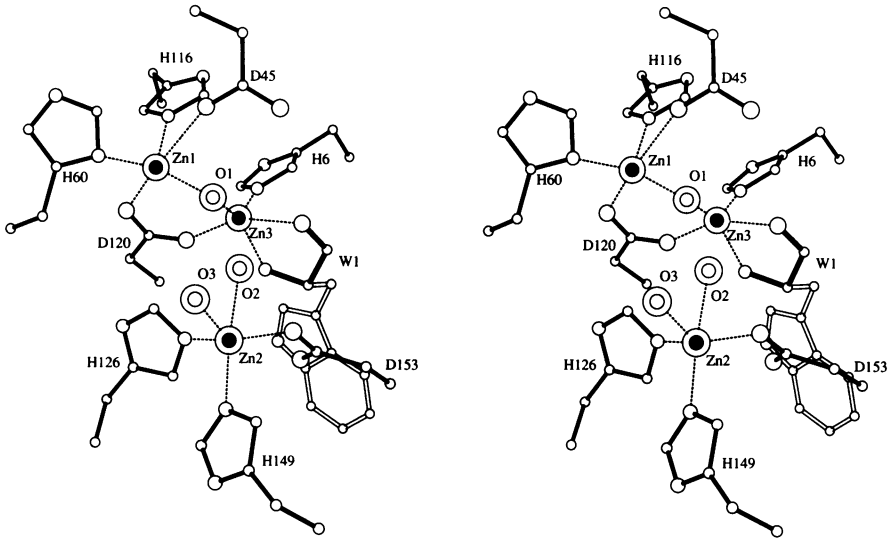

Fig. 3. Zn co-ordination. Carbon, nitrogen and oxygen atoms are represented by open spheres with the radius depending on atom type. Zn ions (Zn1, Zn2, Zn3) and water ligands (O1, O2, O3) are depicted by two concentric spheres. Dashed lines indicate Zn−ligand interactions.

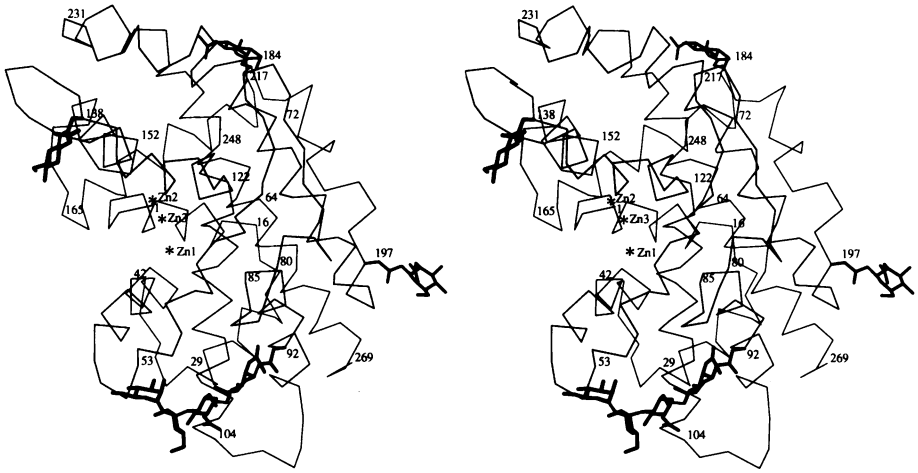

Fig. 4. C_α trace of a P1 molecule, showing in heavy lines the sugar residues visible in the electron density at the four carbohydrate attachment sites Asn92, Asn138, Asn184 and Asn197.

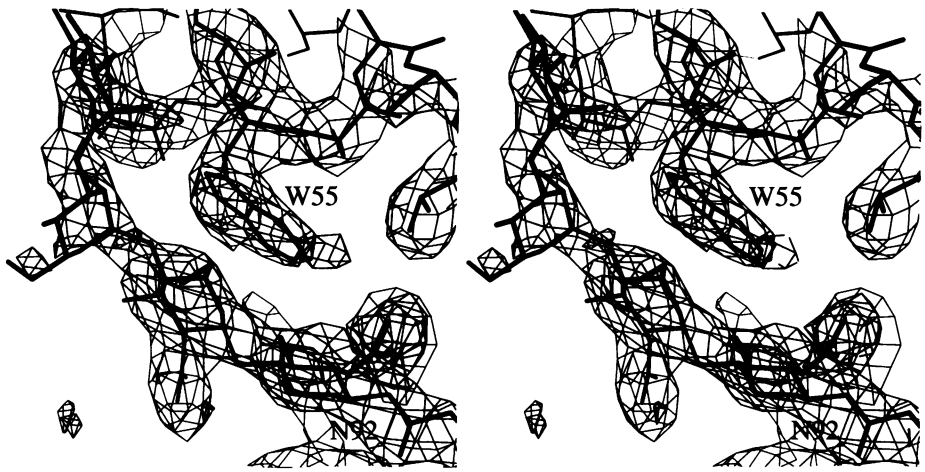

Fig. 5. Part of a $2mF_o-DF_c$ map showing the packing of four sugar residues of the carbohydrate moiety attached to Asn92 against the protein, specifically Trp55.

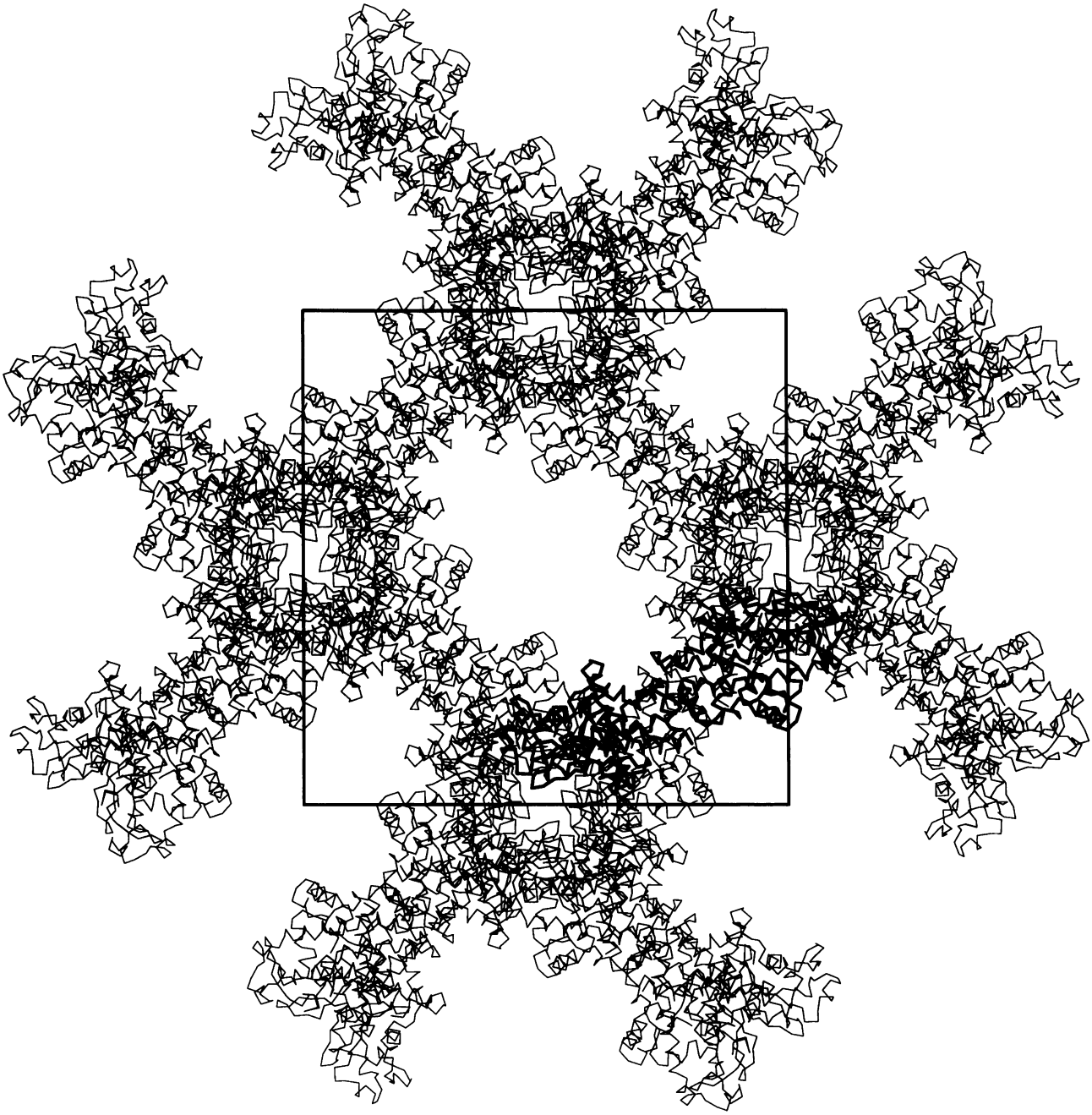

Fig. 6. Packing diagram for the tetragonal crystal form of P1, viewed along the four-fold screw axis. The square box shows the a and b unit cell edges. P1 molecules are represented by their C_α traces. One non-crystallographic dimer is indicated in heavy lines.

d) and a single β-hairpin are found in both subunits. Together all observed α-helices represent 56% of the structure; 4% of it is in a 3_{10}-helix conformation and only 3% displays β-structure.

A P1 molecule can be described as an ~57 Å long, 43 Å wide and 23 Å thick ellipsoid with two 12 Å long protrusions perpendicular to its long axis (Figure 2). One of the 'handles' protruding from the center of the molecule is formed by helices J and K/k. The other handle is stabilized by the 7 residue long 3_{10}-helix d. The two handles are separated by the central core helices A and G. Along the long axis of the molecule two pairs of anti-parallel helices are oriented, F−G and L−N, which together form an approximately anti-parallel four-helical bundle. All the other α-helices make large angles with the long axis. Helix B is noteworthy for

Table I. Data statistics

Data set[a]	d (Å)	N_{meas}	N_{unique}	R_{merge}(%)[b]	% complete
native	2.8	75339	23699	8.4	95.7
apo	3.3	47179	12817	7.7	83.9
K_2PtCl_4	3.3	48677	14922	9.0(6.7)	95.0
Hg·CTP	2.8	70958	21430	7.9(5.7)	84.4
$KAu(CN)_2$	4.5	34821	5848	12.1	95.0
dA·P(S)·dA	3.0	68318	19651	5.9	97.4

[a]$KAu(CN)_2$ data were collected on a Xentronics detector (EMBL Heidelberg); all other data were collected on an image plate, using synchrotron radiation (EMBL outstation, Hamburg). Each data set was collected using a single crystal.
[b]On $<I>$ (respectively, $<I^+>$, $<I^->$).

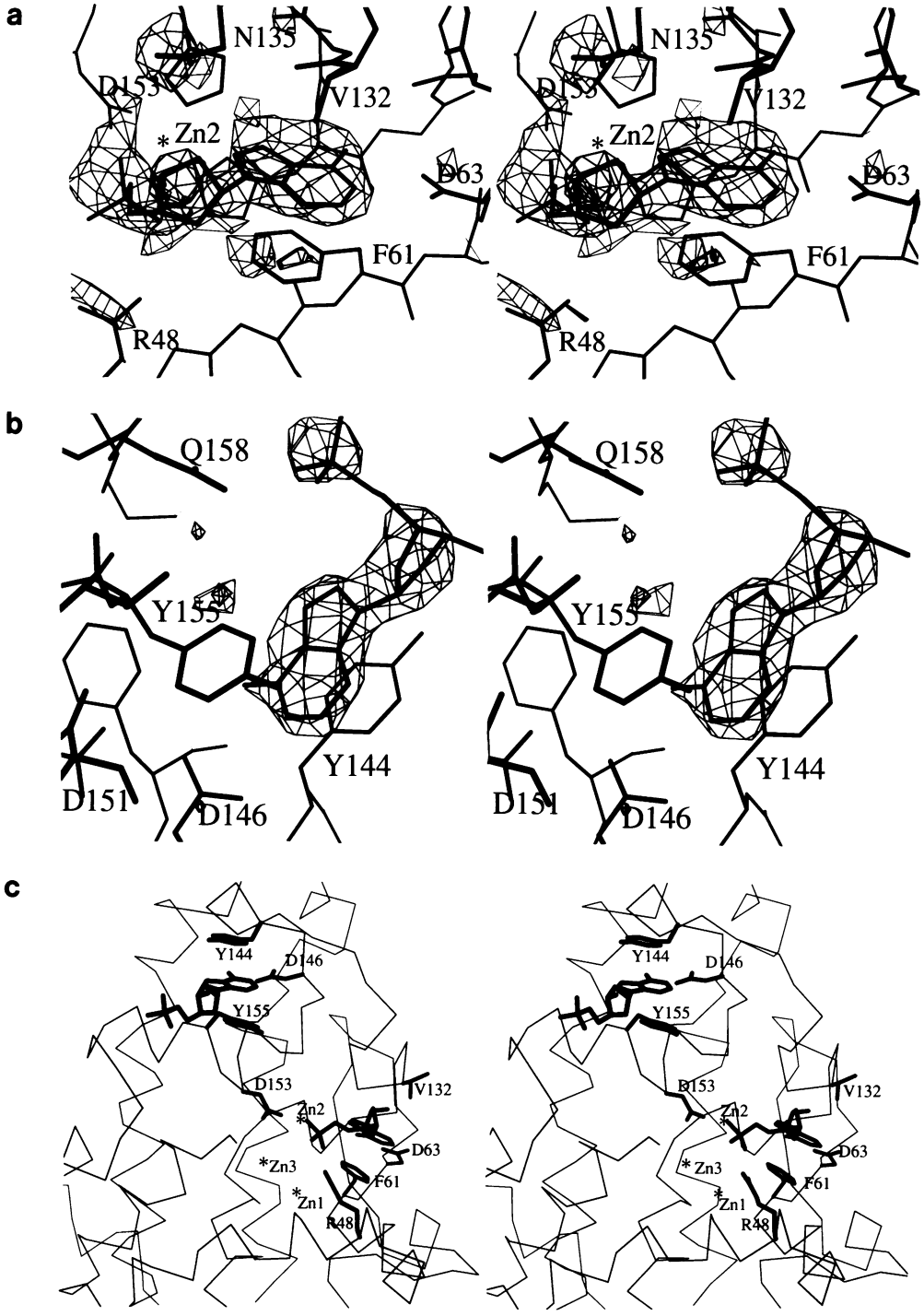

Fig. 7. (**a**) Part of a $m(F_{complex} - F_{native})$ map calculated using phases from the refined native model, m determined by the program SIGMAA (Read, 1986) and $F_{complex}$ denoting the data set measured for a crystal soaked with the thiophosphorylated dinucleotide dA·P(S)·dA. The part shown indicates a mononucleotide binding site close to the Zn cluster. (**b**) Part of the same map showing a second mononucleotide binding site ~20 Å removed from the first one. (**c**) Combined view of the two DNA binding sites, highlighting in heavy lines the residues directly involved in binding. The Zn ions are denoted by asterisks.

its involvement in inter-subunit contacts across the local two-fold axis. Helices I1 and I2 can be described as being part of one larger kinked helix. The three Zn ions are located at the bottom of the large cleft between helices A, C, d, E, H and I1. As described below, inside this cleft substrate molecules bind.

The longest loop without regular secondary structure elements is formed by the 15 residues between helix d and E. Its conformation is stabilized by a disulfide bond between Cys72 and Cys217. A second disulfide bond occurs in the short loop between helices E and F, connecting Cys80 to Cys85. The segment between helices H and I1 contains the short β-hairpin involving residues 138–141 and 144–147. This is the only piece of β-structure in P1. The loop connecting helices G and H contains His126, which serves as a ligand for Zn2. The precise conformation of the non-helical segments just described is probably quite important, as they all lie around the substrate binding cleft.

We have also obtained a trigonal and an orthorhombic crystal form of P1, the latter one diffracting to 2.2 Å resolution (see Lahm *et al.*, 1990). Both forms have been solved by molecular replacement (Lahm,A., Volbeda,A. and Suck,D., to be published). They are being refined independently and show only minor structural differences from the tetragonal form.

Zinc co-ordination

The Zn cluster in P1 consists of a dinuclear site containing Zn1 and Zn3 at a distance of ~3.2 Å from each other, and a single site containing Zn2, which is ~5.8 Å removed from Zn1 and 4.7 Å from Zn3 (see Figure 3). Two histidine nitrogens, His60N$^{\delta}$ and His116N$^{\epsilon}$, and two asparate oxygens, Asp45O$^{\delta 1}$ and Asp120O$^{\delta 1}$, serve as ligands for Zn1. Asp120 is also a ligand of Zn3 by its O$^{\delta 2}$ atom. Zn3 is further co-ordinated by His6N$^{\delta}$ and the main chain N and O of Trp1. Zn2 has three amino acid ligands: His126N$^{\epsilon}$, His149N$^{\epsilon}$ and Asp153O$^{\delta 1}$.

Besides the nine amino acid residues mentioned above, three water molecules might be involved in Zn co-ordination. Zn2 probably has two water ligands, which show up as significant peaks in a 2.2 Å resolution $F_o - F_c$ map calculated for the refined orthorhombic model (Lahm,A., Volbeda,A. and Suck,D., to be published). Another possible water ligand is seen in this map in a bridging position between Zn1 and Zn3, opposite to Asp120. All three Zn ions therefore seem to have three oxygen and two nitrogen ligands in what may be described as a trigonal bipyramidal co-ordination. Metal ligand distances vary between 1.9 and 2.5 Å. The arrangement of the Zn ions and their co-ordination is almost the same as in phospholipase C (Hough *et al.*, 1989), the only difference being a Glu/Asp substitution at position 153.

Carbohydrate structure

Four asparagine-linked carbohydrate units have been detected during the amino acid sequence analysis of P1. At least three molecular species were determined by laser ionization mass spectrometry, corresponding to molecular masses of 35.7, 36.3 and 36.7 kd (Maekawa,K., Tsunasawa,S., Dibo,G. and Sakiyama,F., to be published). From the amino acid sequence alone a molecular mass of 29.2 kd is calculated. These results show that the carbohydrate content of P1 ranges between 18 and 20% of the total molecular mass. The apparent heterogeneity is possibly due to extracellular processing of the carbohydrate units.

The electron density map clearly shows extended density features at the side chain positions of Asn92, Asn138, Asn184 and Asn197 which, according to the sequencing results, are the *N*-glycosylation sites (Figures 4 and 8). For the oligosaccharides linked to Asn138, Asn184 and Asn197, density is only visible for the first sugar residue. This indicates that these carbohydrate moieties are disordered, a common observation for glycoproteins. In contrast, for the oligosaccharide attached to Asn92, significant density is visible for at least two *N*-acetyl glucosamines and two following mannose residues (Figure 5). This higher degree of order is obviously due to the fact that these sugar residues pack against the protein. Without them, the side-chain of Trp55 would be completely exposed to solvent, an entropically unfavorable situation for such a large hydrophobic residue. Similar observations have been made in other cases, e.g. in the structure of a human F_c fragment, where an ordered carbohydrate moiety packs against several large hydrophobic side-chains at the outside of the protein (Deisenhofer, 1981).

Substrate binding

Analysis of the packing of the P1 molecules in the tetragonal crystal form reveals the presence of very large solvent channels (Figure 6), which can easily accommodate larger substrate molecules. This makes the crystals of P1 well suited for soaking experiments with single- or even double-stranded oligonucleotides. In order to study their binding, however, one has to prevent their cleavage. An additional practical problem is the small size of the P1 crystals, making collection of good data and therefore the screening of soaking conditions only possible with a synchrotron source.

One way to prevent enzymatic cleavage is to use an uncleavable substrate analog. We use the R-diastereomer of the thiophosphorylated dinucleotide dA·P(S)·dA, which in contrast to the S-diastereomer is not cleaved by P1 (see Potter *et al.*, 1983a). It was possible to collect a good 3 Å resolution data set for tetragonal crystals of P1 soaked in a solution of this substrate analog (see Table I). A difference Fourier with data of the native enzyme (Figure 7a and b) shows two well defined mononucleotide binding sites ~20 Å apart (Figure 7c). As we found no cleavage of the substrate analog within a period of a few days in solution, it seems that the second half of the dinucleotide is disordered at both sites in the crystal, and therefore invisible in the map.

The first binding site is situated close to the Zn cluster (Figure 7a). Density for an adenine base is clearly visible between the side-chains of Phe61 and Val132. If a 5′-AMP is modeled into the calculated difference Fourier (as in Figure 7a), the carboxylate group of Asp63 as well as the carbonyl oxygens of Leu125 and Glu128 are available as hydrogen bond acceptors for the N6 group of the adenine. An extra hydrogen bond might be formed between N1 of the adenine and the side-chain of Asp63, but this requires protonation of either the base or the Asp. Alternatively the 3′-AMP part of the dinucleotide can be modeled inside the binding pocket, but this fits less well into the calculated density. In addition it cannot form as many hydrogen bonds with the protein compared with the 5′-AMP model. The 5′-AMP model is therefore more likely to be correct, but this has to be checked by refinement. There is clear density for the deoxyribose, which is not involved in direct contacts with the enzyme. A high peak close to Zn2 may be interpreted as a phosphate peak, indicating co-ordination of the phosphate group to the single Zn. The phosphate group appears also to be involved in a salt bridge interaction with the guanidinium group of Arg48. Apart from the ligands of Zn2 there are no other protein side-chains in the direct neighborhood of the phosphate.

At the second site (Figure 7b), which coincides with one of the binding sites of the Hg·CTP derivative used for MIRAS phasing, clear density is present for an adenine base stacking between the two exposed side-chains of Tyr144 and Tyr155. Hydrogen bond formation between the carboxylate group of Asp146 and the N1 and N6 of the adenine base is possible in a very similar way as observed for the other binding site. As discussed below, the same carboxylate is involved in a close contact with the carboxylate group of Asp151. The observed density for the deoxyribose suggests that it is not involved in direct interactions with the

protein. A 5′-phosphate group is tentatively modeled into an unconnected density feature which is within hydrogen bonding distance of the side-chain of Gln158. More detailed information about the mode of binding at both sites might become available after careful refinement of the model of the enzyme–dinucleotide complex.

Discussion

Comparison with phospholipase C

The three-dimensional structures of two other enzymes utilizing three Zn ions are known: phospholipase C (PLC) from *B.cereus* (Hough *et al.*, 1989) and alkaline phosphatase from *Escherichia coli* (Sowadski *et al.*, 1985). The latter probably contains one Mg and two Zn ions *in vivo*, but is also active with three Zn ions. Comparison of amino acid sequences did not show any similarity of the phosphatase with P1. We did find low sequence similarity between P1 and PLC, but only in short segments containing the Zn ligands of the latter enzyme. Because the rest of the sequence did not show any obvious similarity, we concluded that the structure of P1 might be similar to the structure of PLC only in the metal binding region (Lahm *et al.*, 1990).

After completion of the chain tracing for P1, we were therefore surprised to see that the structural similarity with PLC extends to the major part of the molecule (Figure 8). The two enzymes appear to have essentially the same folding. Using the partially refined 2.8 Å resolution P1 model (R = 21.6%) and a PLC model refined at 1.5 Å resolution (Hough *et al.*, 1989), 145 C_α-atoms may be superimposed, giving a positional r.m.s. deviation of 1.8 Å. According to the structural superposition, the overall sequence identity of the superimposed segments is only 18%. Apart from the loop regions, the main structural differences of P1 are the handles containing helices J, K/k and d, the short β-hairpin and the short helices H and M, all of which have no counterpart in PLC. Nevertheless, eight of the nine α-helices in PLC appear to correspond to α-helical segments in P1 (Figure 8). Moreover, the Zn co-ordination is essentially the same in the two enzymes, including the presence of three water ligands. The only difference in Zn co-ordination is an Asp at position 153 in P1, which is a Glu in PLC. In spite of the overall low sequence identity, the extensive similarities in folding, Zn co-ordination and function (both enzymes are phosphodiesterases) strongly suggest that P1 and PLC are evolutionarily related.

Function of the Zn ions

The metal cluster in P1 is located at the bottom of the substrate binding cleft and consists of a relatively inaccessible dinuclear metal site and a more exposed single site (Figure 3). The co-ordination of the dinuclear pair of Zn ions appears to involve four amino acid ligands and one bridging water molecule. A bridging water or OH^- molecule is also observed in other proteins with dinuclear metal sites, e.g. in myohemerythrin, which has a dinuclear Fe site (Sheriff *et al.*, 1987), and in ascorbate oxidase, which contains a trinuclear Cu site (Messerschmidt *et al.*, 1989). The co-ordination of Zn2 on the other hand resembles that observed in all so far structurally characterized catalytically active Zn sites (reviewed by Vallee and Auld, 1990), sharing the common feature of three amino acid ligands—with two of them separated by a short spacer and the third further away in the primary sequence—and at least one water ligand. This suggests that Zn2 is directly involved in catalysis.

The dinuclear Zn site probably has an important function in the stabilization of the folding of P1. Evidence for this is the drastic change of the CD spectrum of the enzyme when all Zn ions are removed by EDTA (see Fujimoto *et al.*, 1975b), a procedure which inactivates the enzyme. It was reported that the activity may be partially restored by readdition of Zn, but we were not able to reactivate the

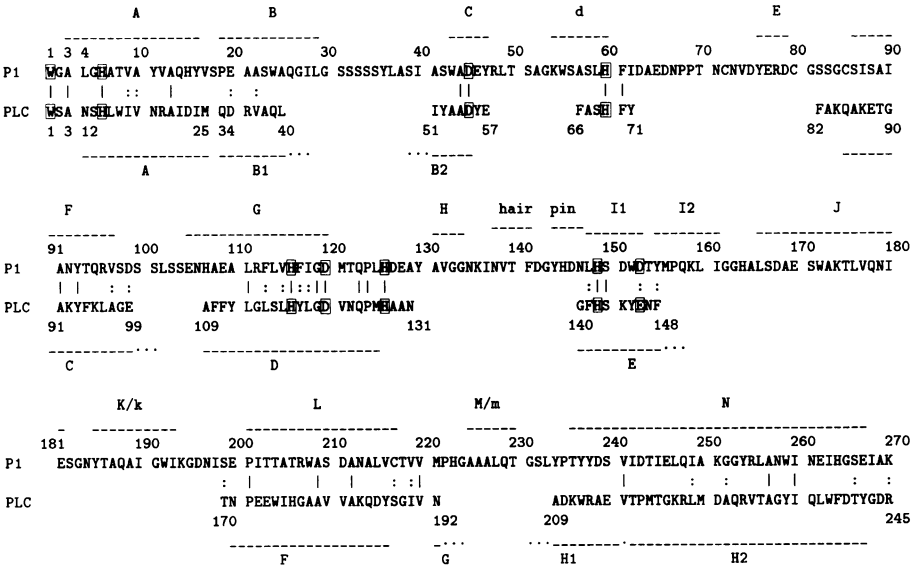

Fig. 8. The complete amino acid sequence of P1 and an alignment of those 145 residues of *B.cereus* phospholipase C (sequenced by Johansen *et al.*, 1988) for which the X-ray structure (determined by Hough *et al.*, 1989) may be superimposed on the P1 model with a r.m.s. deviation of 1.8 Å. All Zn ligands are boxed. Sequence identities are highlighted by a vertical bar, conservative substitutions by a colon. Secondary structure elements of P1 and PLC are indicated above and below the sequence, respectively, using capital letters for the labeling of α-helices and small letters for 3_{10}-helices.

enzyme after prolonged EDTA treatment. An important structural function of the dinuclear pair of Zn ions is also suggested by the observation that they tightly link together three regions which are far apart in the amino acid sequence (Figures 3 and 8).

A 6 Å resolution difference Fourier, calculated using data from crystals grown in phosphate and in acetate buffer, shows a high peak located in the center of the three Zn ions. This peak probably indicates a binding site for inorganic phosphate, which is a reaction product of the phosphomonoesterase activity of the enzyme. At the same position with respect to the trinuclear Zn cluster, phosphate binding has been observed in PLC (Hough *et al.*, 1989). According to this result all three Zn ions may be directly involved in the 3′-nucleotidase reaction.

Interacting carboxylates

A remarkable feature of the structural model of P1 is the presence of two buried pairs of carboxylate groups at the edge of the substrate binding cleft. The first pair involves Asp66 and Glu128 (Figure 9), the second one Asp146 and Asp151 (Figure 7b). None of these residues is involved in an internal salt bridge. In order to prevent repulsive electrostatic interactions which would disturb the conformation of the enzyme, the interacting carboxylates must be protonated. A typical pK for the carboxylate group of aspartic and glutamic acid residues on the surface of a protein is 4.4 (Stryer, 1988). This may shift considerably, however, when such a residue is buried inside the protein.

All interacting Asp and Glu residues appear to fit well into the electron density (see e.g. Figure 9) and refine to relatively low temperature factors, as expected for buried residues. This is also true for the other two crystal forms. In addition, an omit map calculated from a model which does not include these residues shows density in virtually the same positions. All neighboring residues show a good fit to the density as well. Moreover, the observed Phi−Psi combinations are all within or very close to the allowed regions in a Phi−Psi plot. We could not find a different tracing of the polypeptide chain with a good fit to the electron density. All these criteria indicate that we can trust our atomic model, and that the buried carboxylate pairs are indeed present in P1 nuclease.

A systematic search of well refined structures in the protein data base (Bernstein *et al.*, 1977) revealed only six other examples of buried uncompensated Asp−Asp or Asp−Glu pairs. They occur in porcine pepsin (Sielecki *et al.*, 1990; Cooper *et al.*, 1990), *Rhizopus chinensis* pepsin (Suguna *et al.*, 1987), *Endothia parasitica* pepsin (Veerapandian *et al.*, 1990) and bovine chymosin (Gilliland *et al.*, 1990). These enzymes, all being aspartate proteinases, utilize a pair of aspartate side-chains for catalysis. In *Rhizopus* pepsin and

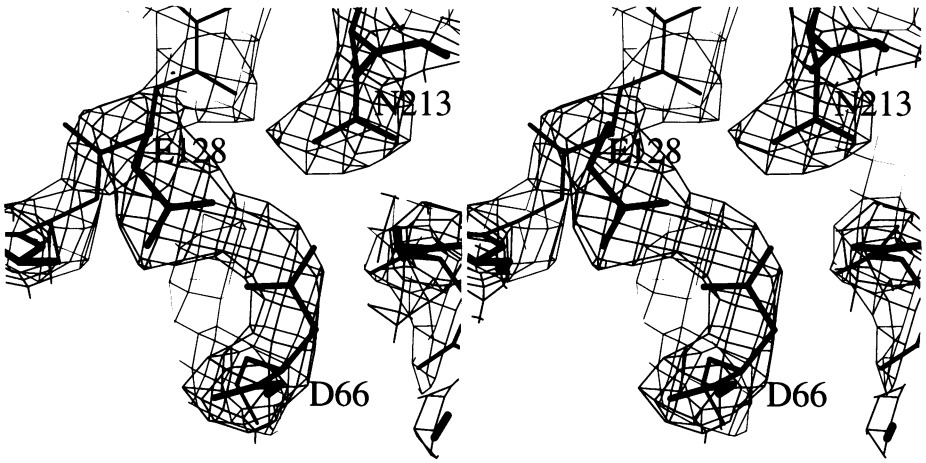

Fig. 9. Part of $2mF_o-DF_c$ map showing the fit of the interacting side-chains of Asp66 and Glu128.

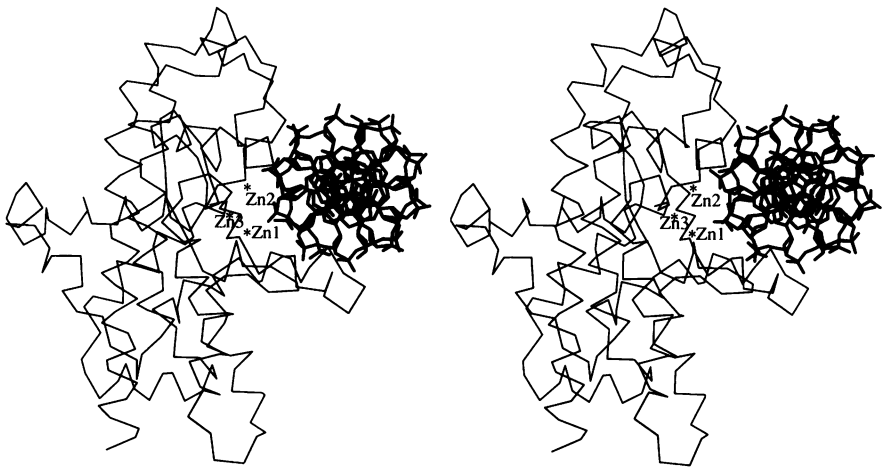

Fig. 10. Docking of a 14 bp fragment of a regular B-DNA structure into the substrate binding cleft of P1.

Endothia pepsin, an additional uncompensated buried Asp–Glu pair distant from the active site is observed. The crystal structures of the aspartic proteinases mentioned above were determined at pH values between 2.0 and 6.0. As the P1 crystals used for the structure determination were buffered at pH 5.3, it is probably safe to conclude that the Asp66–Glu128 and Asp146–Asp151 pairs are protonated in the crystal structure.

Binding and cleavage of substrates

A characteristic feature of the enzymatic activity of P1 is the strong preference for single-stranded nucleic acid conformations. A clue for the origin of this preference is obtained from the docking of double-stranded B-DNA into the substrate binding cleft of the enzyme (Figure 10). If the structure of the enzyme and the DNA is not disturbed, the closest distance any phosphate group can approach Zn2, the most exposed Zn ion, is ~7 Å. However, as double-stranded DNA is known to be rather flexible (see e.g. Steitz, 1990), it cannot be excluded that after some conformational change a much closer approach to the Zn cluster is possible. An additional clue is provided by the results of our binding studies with the thiophosphorylated dinucleotide dA·P(S)·dA, showing two hydrophobic binding pockets for an unpaired base 20 Å away from each other. This suggests that the DNA has to be unwound before cleavage can occur, thus explaining the high preference of P1 for single-stranded substrates.

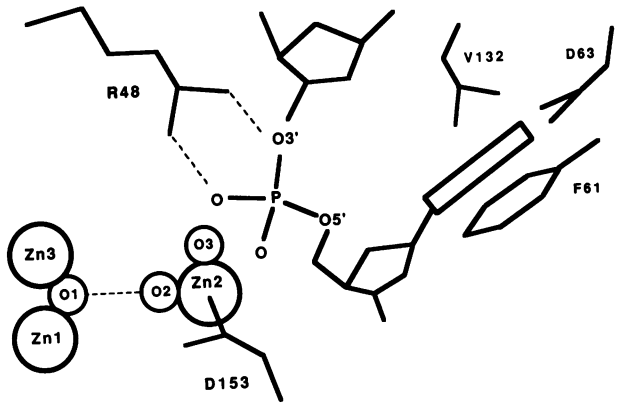

Fig. 11. Schematic view of the active site of P1. Zn ions are depicted by large open circles, their water ligands by smaller circles. The ribose attached to the O3' is dotted in order to indicate that it is not visible in the electron density. Dashed lines denote hydrogen bonding interactions likely to occur in a P1–substrate complex. The base of the nucleotide, represented by a rectangle, is stacked between a phenylalanine (Phe61) and a valine (Val132) and in addition is hydrogen bonded to Asp63.

A comparison between P1 and other single-stranded DNA binding proteins shows interesting common features. For example, the complex of RNase T1 and 2'-GMP like P1 shows a binding site in between two tyrosine side-chains (Arni *et al.*, 1988), an additional similarity being the apparent hydrogen bonding interaction between the purine and a carboxylate group, belonging to a Glu in RNase T1 and to an Asp in P1. The mononucleotide binding site near the Zn ions in P1 resembles the binding of single-stranded DNA at the 3',5'-exonuclease site of DNA polymerase I (Klenow fragment, see Beese and Steitz, 1991). Whereas an exposed phenylalanine and valine residue are involved at the active site of P1, a phenylalanine and a leucine form a hydrophobic binding pocket for a base in the polymerase. As discussed by Beese and Steitz, the exposed Phe and Leu probably play a key role in the observed unwinding of double-stranded DNA by the Klenow fragment (see Freemont *et al.*, 1988). In view of the fact that P1 has two hydrophobic binding pockets for unpaired bases it appears not unlikely that it can unwind double-stranded substrates as well. This might explain the ready degradation by P1 of certain non-A/non-B conformations in double-stranded DNA, as these are probably less stable than the A and the B forms and therefore easier to unwind.

An interesting feature of the two DNA binding sites in P1 is the possible formation of two hydrogen bonds between the adenine and Asp63 and Asp146 respectively (see Figure

Table II. MIRAS phasing statistics (20–3.3 Å resolution)

Resolution (Å)	12.25	8.83	6.90	5.67	4.80	4.17	3.68	3.30	All
$N(F)_{nat}$	176	431	835	1324	1877	2536	3262	4084	14525
$<m>_{nat}$	0.64	0.70	0.74	0.71	0.67	0.60	0.57	0.55	0.61
$N(F)_{apo}$	156	332	685	1106	1625	2236	2931	3721	12792
$<m>_{apo}$	0.75	0.82	0.83	0.80	0.77	0.70	0.67	0.66	0.71
$N(F)_{com}$	198	447	855	1340	1895	2564	3295	4090	14684
$<m>_{com}$	0.75	0.80	0.82	0.80	0.77	0.69	0.66	0.66	0.71
$kemp(Pt_{nat})$	8.0	8.5	7.8	5.7	4.7	4.4	3.7	2.7	3.9
$kemp(Pt_{apo})$	8.7	9.4	9.0	6.5	5.9	5.0	4.0	2.9	4.4
$kemp(Hg_{nat})$	8.7	7.3	6.8	5.7	6.5	6.3	5.7	5.0	5.7
$kemp(Hg_{apo})$	9.8	8.6	9.2	7.2	8.3	7.1	6.5	5.5	6.7
$pp(Pt_{nat})$	1.4	1.7	2.5	2.2	2.0	1.4	1.1	0.9	1.4
$pp(Pt_{apo})$	1.6	2.1	3.0	3.2	2.9	2.5	2.0	1.7	2.2
$pp(Hg_{nat})$	2.5	3.1	3.0	3.8	3.0	2.4	2.4	2.5	2.6
$pp(Hg_{apo})$	2.5	2.9	3.3	3.4	3.1	2.6	2.5	2.8	2.8
$pp(Au_{apo})$	1.4	3.1	2.3	1.8	1.6	1.4	–	–	1.7
$pp(Zn_{apo})$	4.6	2.9	2.1	2.3	1.9	1.7	1.5	1.6	1.7

N(F) = number of phased reflections; <m> = average figure of merit; nat = native, apo = apoenzyme, com = combined; kemp = 2Δ(iso)/Δ(ano) where Δ(iso) and Δ(ano) indicate isomorphous and anomalous differences respectively; Pt_{nat} = Pt versus native data set (analagous definitions for Pt_{apo}, Hg_{nat} and Hg_{apo}): pp (phasing power) = r.m.s. heavy atom F/r.m.s. residual.

7). This requires protonation of either the carboxylate or the N1 atom of the adenine. As far as we know, this specific mode of hydrogen bonding has not been observed before in a protein–DNA complex. As discussed above, Asp146 might already be protonated before binding of substrate because of its interaction with Asp151. Protonation of Asp63 might be induced by substrate binding. Alternatively protonation of the adenine may be considered, although the normal pK of the N1 nitrogen is only 3.5. However, an example of protonated adenines at a pH of 6.6 is found in the crystal structure of a DNA duplex containing A(anti)–G(syn) base pairs (Brown *et al.*, 1989). This clearly shows that protonation between the adenine N1 and the carboxylate is very well possible at pH 5.3 (the pH used for crystallization) in the buried environment of the two substrate binding pockets.

A common feature in the catalytic mechanism of nucleases seems to be the formation of a pentacovalent phosphate intermediate (see e.g. Cotton *et al.*, 1979; Arni *et al.*, 1988; Breslow *et al.*, 1989; Beese and Steitz, 1991). The catalytic mechanism of all zinc dependent enzymes on the other hand seems to involve in at least one step a nucleophilic attack by a Zn-activated water molecule (see e.g. Vallee and Auld, 1990). Activation and/or orientation of the attacking water is assisted in most cases by protein side-chains, e.g. a Thr and a Glu in carbonic anhydrase (Eriksson *et al.*, 1988), a Tyr and a Glu in the Klenow fragment of DNA polymerase I (Derbyshire *et al.*, 1991; Beese and Steitz, 1991), and a Glu alone in thermolysin (Holmes and Matthews, 1981) and carboxypeptidase A (Kim and Lipscomb, 1990). All these Zn enzymes have a pH optimum in the alkaline region, supporting the idea that activation of the water molecule involves a deprotonation.

A remarkable feature of P1 is that the pH optimum for its phosphodiesterase activity lies between 4.5 and 6.0, depending on the substrate (Fujimoto *et al.*, 1974b). There are three possible structural reasons which might explain a pH optimum in the acid region. The first is the presence of two probably protonated carboxylate pairs at the edge of the substrate binding cleft. It is not unreasonable to assume that they will become unprotonated under alkaline conditions. This should give rise to a conformational rearrangement which is expected to have a significant effect on the structure of the substrate binding cleft and therefore also on the enzymatic activity. Second, according to our modeled enzyme–substrate complex, the binding of DNA might be facilitated, at least for adenine bases, by a protonation step. A third contributing factor to the low pH optimum might be the absence of any suitable protein side-chain in the direct neighborhood of the phosphate group bound at the active site which could act as a proton donor for the leaving R–O3$^-$ group formed after cleavage of the P–O3′ bond.

On the basis of biochemical evidence a reaction mechanism has been proposed for nuclease S1 (Witzel *et al.*, 1986), an enzyme closely related to P1, involving nucleophilic attack of the phosphorus by the carboxylate group of a Glu. This mechanism appears to be very unlikely for P1, as the observed binding of the dA·P(S)·dA substrate analog shows only Arg48 and Asp153 to be within hydrogen bonding distance of the bound phosphate at the active site (Figure 11), the closest glutamate (Glu46) being >10 Å away. The mode of binding of the substrate analog in contrast strongly suggests that nucleophilic attack occurs by a Zn activated water molecule.

There are two likely candidates for the attacking water molecule: O2 and O3 (Figure 11). If it is O2, our binding model suggests that as a first step in the reaction O3 must be substituted by the phosphate group. If it is O3, nucleophilic attack of the phosphate can start immediately. As cleavage proceeds with inversion of configuration at the phosphorus (Potter *et al.*, 1983a), the attack should be approximately in line with the P–O3′ bond. We see only half of the dinucleotide in the electron density, so it is impossible from the present results to decide which of the two mechanisms is correct. More might be said after refinement of the complex, when this provides evidence on the location of P–O3′ bond. The carboxylate group of Asp153, although already functioning as a ligand of Zn2, might assist in the activation of either O2 or O3. Arg48 may serve to stabilize the pentacovalent phosphate intermediate which is likely to be formed. Interestingly, O2 is within hydrogen bonding distance of both O3 and O1, the bridging water ligand of the dinuclear Zn site. This provides a possibility for a direct role of the latter in the activation of the attacking water. A contribution of all three positively charged Zn ions in the activation process is especially attractive in view of the low pH optimum of the enzyme.

Materials and methods

Data collecting and processing

Tetragonal crystals of P1 nuclease from *P.citrinum* were obtained as described previously (Lahm *et al.*, 1990). Their space group is $P4_32_12$ with a = b = 133.7 Å, c = 108.3 Å. Phasing at 4.5 Å resolution was performed by using a combination of SIRAS information from the Pt derivative (two sites/asymmetric unit) and solvent flattening (see Lahm *et al.*, 1990). This revealed that there is a dimer in the asymmetric unit, giving rise to a V_M of 3.4 Å^3/dalton (Matthews, 1968) or a solvent content of ~64%.

Five data sets were collected and used for phasing at 3.3 Å resolution (Table I). A native, two derivative (Pt and Hg) and an apo data set (obtained by EDTA soaking of a native crystal) were collected at the X11 beamline of the DESY–EMBL outstation in Hamburg, using the Fuji image plate detector developed in-house (by J.Hendrix and A.Lentfer) at a wavelength of ~0.96 Å. Data reduction was performed with a suite of film processing programs from the CCP4 software package. A data set for a Au derivative was collected on a Xentronics detector, using Cu K_α radiation, and processed with the program XDS (Kabsch, 1988). Scaling of derivative to native data was done with the program KBRANI from the BIOMOL package developed in Groningen.

MIRAS phasing

Using the 4.5 Å resolution SIRAS phases $F_{derivative} - F_{native}$ and $F_{derivative} - F_{apo}$ difference maps were calculated, revealing four heavy atom sites for the Hg derivative, three for the Au derivative and six Zn positions per asymmetric unit. In the next stage, phasing was performed at 3.3 Å resolution using the program PHARE (written by Dr G.Bricogne). The native data were phased using the isomorphous and anomalous differences from the Pt and Hg derivatives. The same was done for the apoenzyme, but using in addition the isomorphous differences with the Au derivative (which gave only weak phasing power for the native enzyme) and the holo (i.e. native) enzyme. In the phasing of the apoenzyme, the Zn contribution was taken into account for all the derivatives.

The results summarized in Table II show that the final statistics are much better for the apoenzyme. The phasing power obtained from the Zn ions alone is remarkable. The apoenzyme phase set is, however, much less complete than the native phase set, which may be an important disadvantage for the subsequent solvent flattening step. The native phases were therefore improved as follows: from the refined Zn parameters, ABCD coefficients were calculated (see Hendrickson and Lattmann, 1970) and added to ABCD(apo), giving ABCD(apo+Zn). Because the phasing sources of the two phase sets are interdependent, addition of ABCD(apo+Zn) to

ACBD(native) would give much too high figures of merit. Therefore they were averaged instead, and the final set of MIRAS phases and figures of merit were reconstructed from the resulting ACBD(av).

Solvent flattening

A 3.3 Å resolution MIRAS map was calculated which clearly showed the outline of the protein region. This map was improved by 10 cycles of solvent flattening (Wang, 1985; Leslie, 1987) using a solvent level of 54%, an 8 Å radius for local density averaging and no truncation of negative protein density inside the envelope. After each cycle the inverted map phases were combined with the MIRAS phases using the program SIGMAA (Read, 1986). At the end, the average phase change with respect to the MIRAS phases was 43.8° and the disagreement factor between inverted map and observed structure factor amplitudes was 17.6%. Skeletonization of the solvent flattened map (Jones and Thirup, 1986) showed very clear helical features. These were, however, left-handed, indicating that the initial space group assignment ($P4_12_12$) had to be changed to $P4_32_12$.

Chain tracing and starting model

A new map was calculated and 44 equivalent C_α positions were traced independently for each of the two subunits in the asymmetric unit. Together with the known heavy atom positions these were used to determine the local two-fold, giving χ = 182.4, ϕ = 42.1 and φ = 70.1°, and a translation component of 0.25 Å parallel to the rotation axis. Two-fold density averaging (Bricogne, 1976) produced a very good map allowing a complete chain tracing. Side-chain and peptide atoms were generated by the automatic procedure developed by Jones and Thirup (1986) and the resulting model was subjected to energy minimization. After manual adjustment of the worst van der Waals collisions, this produced a starting model for refinement with a crystallographic R-factor of 48.4% (for 13745 reflections in the resolution range between 8.0 and 3.3 Å).

Refinement

Refinement was started with the program XPLOR (Brünger, 1988), using a protocol consisting of a prepstage (40–80 cycles of conventional refinement), a heat stage (0.5–1 ps of molecular dynamics refinement at 2000K), a cool stage (0.25 ps of MD refinement with gradual cooling to 300K) and a final stage (50–100 cycles of conventional refinement). The two subunits were refined independently. The R-factor after one round of positional refinement at 3.3 Å resolution using this protocol was 26.5%. After an extensive manual rebuilding step using σ_A weighted $2F_o - F_c$ and $F_o - F_c$ type maps (see Read, 1986), one round of positional refinement at 2.8 Å and 20 cycles of individual temperature factor refinement, the R-factor dropped to 23.5% (22653 Fs, 8.0–2.8 Å resolution). After a second manual rebuilding step and addition of two glucosamines per subunit, refinement was continued with the program PROLSQ (Hendrickson and Konnert, 1980), this time using moderate restraints for the non-crystallographic symmetry. The present R-factor is 21.6% for all 22653 Fs between 8.0 and 2.8 Å resolution, and 19.7% for the 20024 Fs above the 3σ level, with an overall good geometry (r.m.s. deviation from ideal values for bond lengths: 0.013 Å, for planar groups: 1.8°). Details of the refinement which is to be continued further will be published elsewhere.

Substrate binding

A tetragonal crystal of P1 nuclease was soaked for 1 day in a solution containing 70% saturated ammonium sulfate, 40 mM sodium acetate and 0.5 mM of the R-diastereomer of the thiophosphorylated dinucleotide dA·P(S)·dA, a cleavage-resistant substrate analog (Spitzer and Eckstein, 1988). The pH of the soaking solution was 5.3. Processing of an image plate data set of this crystal gave good statistics to a resolution of 3.0 Å (Table I). The binding of the dinucleotide was studied by inspection of a $F_{complex} - F_{native}$ map with σ_A weighted phase information from the refined native model.

Acknowledgements

The assistance of the staff of the EMBL outstation of the Hamburg synchrotron radiation facility with data collection is gratefully acknowledged. Dr G.Vriend was of great help in obtaining the starting model for refinement by using automatic procedures incorporated in his program WHATIF for generating side-chain and peptide atoms from C_α positions. We thank Drs E.Hough and L.-K.Hansen for making the co-ordinates of phospholipase C from *B.cereus* available to us and Dr F.Eckstein for kindly providing the R-diastereomer of the thiophosphorylated dinucleotide dA·P(S)·A.

References

Arni,R., Heinemann,U., Tokuoka,R. and Saenger,W. (1988) *J. Biol. Chem.*, **263**, 15358–15368.
Beese,L.S. and Steitz,T.A. (1991) *EMBO J.*, **10**, 25–33.
Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanoucki,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.
Brandén,C.-I. and Jones,T.A. (1990) *Nature*, **343**, 687–689.
Breslow,R., Huang,D.-L. and Anslyn,E. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1746–1750.
Bricogne,G. (1976) *Acta Crystallogr.*, **A32**, 832–847.
Brown,T., Leonard,G.A., Booth,E.D. and Chambers,J. (1989) *J. Mol. Biol.*, **207**, 455–457.
Brünger,A. (1988) In Isaacs,N.W. and Taylor,M.R. (eds), *Crystallographic Computing 4: Techniques and New Technologies*. Clarendon Press, Oxford, pp. 127–140.
Cooper,J.B., Khan,G., Taylor,G., Tickle,J.J. and Blundell,T.L. (1990) *J. Mol. Biol.*, **214**, 199–222.
Cotton,F.A., Hazen,E.E. and Legg,M.J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 2551–2555.
Deisenhofer,J. (1981) *Biochemistry*, **20**, 2361–2370.
Derbyshire,V., Grindley,N.D.F. and Joyce,C. (1991) *EMBO J.*, **10**, 17–24.
Eriksson,A.E., Kylsten,P.M., Jones,T.A. and Liljas,A. (1988) *Proteins*, **4**, 283–293.
Freemont,P.S., Friedman,J.M., Beese,L.S., Sanderson,M.R. and Steitz,T.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8924–8928.
Fujimoto,M., Kuninaka,A. and Yoshino,H. (1974a) *Agr. Biol. Chem.*, **38**, 785–790.
Fujimoto,M., Kuninaka,A. and Yoshino,H. (1974b) *Agr. Biol. Chem.*, **38**, 1555–1561.
Fujimoto,M., Fujiyama,K., Kuninaka,A. and Yoshino,H. (1974c) *Agr. Biol. Chem.*, **38**, 2141–2147.
Fujimoto,M., Kuninaka,A. and Yoshino,H. (1975a) *Agr. Biol. Chem.*, **39**, 1991–1997.
Fujimoto,M., Kuninaka,A. and Yoshino,H. (1975b) *Agr. Biol. Chem.*, **39**, 2145–2148.
Gilliland,G.L., Winborne,E.L., Nachman,J. and Wlodawer,A. (1990) *Proteins*, **8**, 82–101.
Hendrickson,W.A. and Konnert,J.H. (1980) In Diamond,R., Rameshan,S. and Vekatesan,K. (eds), *Computing in Crystallography*. Indian Institute of Science, Bangalore, India, pp. 13.01–13.23.
Hendrickson,W.A. and Lattman,E.E. (1970) *Acta Crystallogr.*, **B26**, 136–143.
Holmes,M.A. and Matthews,B.W. (1981) *Biochemistry*, **20**, 6912–6920.
Hough,E., Hansen,L.K., Birknes,B., Jynge,K., Hansen,S., Hordvik,A., Little,C., Dodson,E. and Derewenda,Z. (1989) *Nature*, **338**, 357–360.
Johansen,T., Holm,T., Guddal,P.H., Sletten,K., Haugli,F.B. and Little,C. (1988) *Gene*, **65**, 293–304.
Jones,T.A. and Thirup,S. (1986) *EMBO J.*, **5**, 819–822.
Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr.*, **A46**, in press.
Kabsch,W. (1988) *J. Appl. Crystallogr.*, **21**, 916–924.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Kim,H. and Lipscomb,W.N. (1990) *Biochemistry*, **29**, 5546–5555.
Lahm,A., Volbeda,A. and Suck,D. (1990) *J. Mol. Biol.*, **215**, 207–210.
Lesk,A.M. and Hardman,K.D. (1982) *Science*, **216**, 539–541.
Leslie,A.G.W. (1987) *Acta Crystallogr.*, **A43**, 134–136.
Luzzati,V. (1953) *Acta Crystallogr.*, **5**, 802–819.
Matthews,B.W. (1968) *J. Mol. Biol.*, **33**, 491–497.
Messerschmidt,A., Rossi,A., Ladenstein,R., Huber,R., Bolognesi,M., Gatti,G., Marchesini,A., Petruzelli,R. and Finazzi-Agro,A. (1989) *J. Mol. Biol.*, **206**, 513–529.
Potter,B.V.L., Connolly,B.A. and Eckstein,F. (1983a) *Biochemistry*, **22**, 1369–1377.
Potter,B.V.L., Romaniuk,P.J. and Eckstein,F. (1983b) *J. Biol. Chem.*, **258**, 1758–1760.
Pulleyblank,D.E., Glover,M., Farah,Ch. and Haniford,D.B. (1988) In Wells,R.D. and Harvey,S.C. (eds), *Unusual DNA Structures*. Springer, Heidelberg, pp. 23–44.
Ramachandran,G.N., Ramakrishnan,G. and Sasisekharan,V. (1963) *J. Mol. Biol.*, **7**, 95–99.
Read,R.J. (1986) *Acta Crystallogr.*, **A42**, 140–149.
Rould,M.A., Perona,J.J., Söll,D. and Steitz,T. (1989) *Science*, **246**, 1135–1142.

Sheriff,S., Hendrickson,W.A. and Smith,J.L. (1987) *J. Mol. Biol.*, **197**, 273–296.
Shishido,K. and Ando,T. (1982) In Linn,S.M. and Roberts,R.J. (eds), *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 155–185.
Shishido,K. and Habuka,N. (1986) *Biochim. Biophys. Acta*, **884**, 215–218.
Sielecki,A.R., Fedorov,A.A., Boodhod,A., Andreeva,N.S. and James,M.N.G. (1990) *J. Mol. Biol.*, **214**, 143–170.
Sowadski,J.M., Handschumacher,M.D., Krishna Murthy,H.M., Foster,B.A. and Wyckoff,H. (1985) *J. Mol. Biol.*, **186**, 417–433.
Spitzer,S. and Eckstein,F. (1988) *Nucleic Acids Res.*, **16**, 11691–11704.
Steitz,T.A. (1990) *Q. Rev. Biophys.*, **23**, 205–280.
Stryer,L. (1988) *Biochemistry*. W.H.Freeman and Company, New York.
Suguna,K., Bott,R.R., Padlan,E.A., Subramanian,E., Sheriff,S., Cohen,G. and Davies,D.R. (1987) *J. Mol. Biol.*, **196**, 877–900.
Vallee,B.L. and Auld,D.S. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 220–224.
Veerapandian,B., Cooper,J.B., Sali,A. and Blundell,T.L. (1990) *J. Mol. Biol.*, **216**, 1017–1029.
Wang,B.C. (1985) *Methods Enzymol.*, **115**, 90–112.
Witzel,H., Berg,W., Creutzenberg,O. and Karreh,A. (1986) In Bertini,I. (ed.), *Zinc Enzymes*. Birkhäuser, Boston, pp. 295–306.

Received on March 6, 1991