Three-dimensional structure of acylphosphatase. Refinement and structure analysis

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · APRIL 1992

Impact Factor: 4.33 · DOI: 10.1016/0022-2836(92)91005-A · Source: PubMed

4 AUTHORS, INCLUDING:

Vladimir Saudek
University of Cambridge
107 PUBLICATIONS **7,263** CITATIONS

SEE PROFILE

J. Mol. Biol. (1992) **224**, 427–440

Three-dimensional Structure of Acylphosphatase

Refinement and Structure Analysis

Annalisa Pastore†

EMBL, Meyerhofstrasse 1, W-6900 Heidelberg, Germany

Vladimir Saudek

Marion Merrell Dow Research Institute, 16 rue d'Ankara, Strasbourg, France

Giampietro Ramponi

Dipartimento di Scienze Biochimiche, Università di Firenze, Firenze, Italy

and Robert J. P. Williams

Department of Inorganic Chemistry, South Parks Road, Oxford OX1 3QU, U.K.

(Received 15 May 1991; accepted 19 November 1991)

We report here the complete determination of the solution structure of acylphosphatase, a small enzyme that catalyses the hydrolysis of organic acylphosphates, as determined by distance geometry methods based on nuclear magnetic resonance information. A non-standard strategy for the distance geometry calculations was used and is described here in some detail. The five best structures were then refined by restrained energy minimization and molecular dynamics in order to explore the conformational space consistent with the experimental data. We address the question of whether the solution structure of acylphosphatase follows the general principles of protein structure, i.e. those learned from analysing crystal structures. Static and dynamic features are discussed in detail. An uncommon β–α–β motif, so far found only in procarboxypeptidase B and in an RNA-binding protein, is present in acylphosphatase.

Keywords: energy refinement; enzyme; nuclear magnetic resonance; structure

1. Introduction

Acylphosphatase (EC 3.6.1.7) is a small enzyme (98 amino acid residues) found in several tissues of vertebrates. It catalyses the hydrolysis of organic acylphosphates with substrate specificity for 1,3-diphosphoglycerate and carbamylphosphates (Ramponi, 1975). The amino acid sequences of a certain number of species (horse muscle, turkey, rabbit, rat, ox, pig and duck) have been determined. To date, neither the exact location of the active site nor the metabolic context of the function of this enzyme is entirely clear (Stefani *et al.*, 1986; Liguri *et al.*, 1984). Attempts to crystallize acylphosphatase did not provide crystals of a quality suitable for X-ray crystallography. The solution structure of acylphosphatase from horse muscle has been determined instead by nuclear magnetic resonance (n.m.r.‡), with a piece-to-piece approach. First, secondary structure determination was achieved (Saudek *et al.*, 1988); secondly, the structure of the two helices present and their relative orientations were determined (Saudek *et al.*, 1989*a*); finally, the co-ordinates of the five-strand β-sheet were deter-

† Author to whom all correspondence should be addressed.

‡ Abbreviations used: n.m.r., nuclear magnetic resonance; NOE, nuclear Overhauser effect; r.m.s., root-mean-square; MD, molecular dynamics; HPr, histidine-rich protein.

0022-2836/92/060427-14 $03.00/0

mined (Saudek *et al.*, 1989*b*). Here, we report the strategy for the complete structure determination and the energy refinement.

For many years crystallography was the only technique available for determining the structures of proteins to atomic resolution. The study, comparison and analysis of crystal structures are of great importance for decoding the very complex syntactic rules that regulate protein folding and help to decode new more complex structures. Recently, a few structures have been determined both in the crystal form and in solution. In each of these cases, how much the results obtained in the two different environments differ from each other has been considered. A number of more general questions should also be addressed, some of which may often have been assumed to be trivial. Do proteins in solution obey the same rules of packing, folding, and assembly as protein crystals? Can we classify their structures in the same way, using the same syntax rules? The answers are not straightforward.

Because acylphosphatase is a globular protein completely determined by n.m.r and with a very well-defined tertiary structure, it may be an ideal example with which to address these questions.

2. Materials and Methods

A total of 1180 distance restraints was obtained from n.m.r. data, as described elsewhere (Saudek *et al.*, 1989*a*,*b*). Of these, 1058 were distances derived from NOE data: 448 intraresidue NOEs, 212 short-range distances (i.e. between i and up to $i+4$), and 398 long-range distances. Further distance restraints were obtained from hydrogen bonds detected in the α-helices and in the β-structures. Backbone torsion angle restraints (Φ angles) were inferred from the coupling constants. A restriction of these angles between $-100°$ and $-20°$ was used for 21 residues in helices and between $-80°$ and $-160°$ for 38 residues in the β-sheet. Dihedral and hydrogen bond restraints were not used in the energy refinement. No stereo-specific assignment of the side-chains was attempted. Distance geometry calculations were carried out using the computer program DISMAN (Braun & Go, 1985). The strategy is described in some detail in Results and Discussion. Refinement was then achieved by restrained energy minimization and molecular dynamics (MD: GROMOS package; van Gunsteren & Berendsen, 1977; van Gunsteren *et al.*, 1984).

One thousand cycles of energy minimization with a conjugate gradient algorithm were performed to remove Lennard-Jones contacts. In the dynamics, the system was weakly coupled to a thermal bath (Berendsen *et al.*, 1981) at $T = 1000$ K for 5 ps, with a temperature relaxation time of 0·01 ps during the 1st picosecond of thermal equilibration and 0·1 ps during all the following runs. The temperature was then slowly decreased to 300 K and 5 ps were run to obtain thermal equilibrium. The program was run for an additional 10 ps and 1500 cycles of energy minimization were run on the final snapshot. Simulations using the time-averaged NOE restraints (Torda *et al.*, 1990) were performed at 300 K, starting from the last snapshot of the previously calculated trajectory.

All simulations were carried out *in vacuo*, with a cut-off for non-bonded interactions of 0·8 nm. The list of interactions was updated every 10 steps of 0·0005 ps each during the dynamics and every 10 cycles of the minimization, respectively. Initial velocities for the atoms were taken from a Maxwell distribution at $T = 1000$ K. The MD time step for the integration of the Newton's equation was $\delta t = 0{\cdot}0005$ ps. The bond length constraining algorithm (SHAKE: Ryckaert *et al.*, 1977; van Gunsteren & Berendsen, 1977) was used. Several preliminary attempts were carried out using different force constants for the restraint energy term. A force constant of 250 kJ mol^{-1} nm^{-2} was chosen eventually. In the simulations with time-averaged distance restraints, an initial value of $(r^0 - 0{\cdot}02)$ nm for the $r(t)$ and a time constant for the memory function of 2·5 ps were used.

The structure analysis was done with the support of the graphics program INSIGHT running on an Evans & Sutherland PS390 (Dayringer *et al.*, 1986). The co-ordinates and the list of restraints have been submitted to the Brookhaven Protein Data Bank (accession no. 1APS; Bernstein *et al.*, 1977).

3. Results and Discussion

In spite of the intrinsically different nature of the information, a parallel may be drawn between structure determinations from n.m.r. data and from crystallography (Bruenger, 1991). First, a model structure is built. Different algorithms are then applied to obtain the best agreement between the model and the experimental information. Finally, structure refinement should optimize the geometry.

Although the final structure must fulfil all the geometrical criteria and be energetically stable, the relative weight given to the experimental information and to the geometry at different stages of the calculations is critical. When the initial model is created (typically by model building in crystallography and by distance geometry methods in n.m.r.) maximum flexibility of the chains is required so that they can easily be rearranged according to the best agreement with the experimental data, even at the expense of distortions in the standard geometry or of atom co-penetration. Once a satisfactory initial model has been obtained, the most energetically stable conformations compatible with the experimental data should be sought. At this stage, too large a weight on the experimental distances might maintain large distortions in the geometry. If too low a weight is chosen, the driving force is then left to the other terms in the potential and it might result in a general increase of the distance restraint energy term (van Gunsteren, 1989). It has also been noticed that the use of high force constants in the calculation is sometimes misleading because the final conformations may be induced by the NOE pseudopotential (van Gunsteren, 1989).

Having these considerations in mind, the procedure used in the calculation and refinement of acylphosphatase is described in sections (a) and (b), below.

(a) *Distance geometry calculations*

A non-standard approach was used for the distance geometry calculations (Braun, 1987). The

structures of the five-strand β-sheet and of the two α-helices were calculated separately (Saudek *et al.*, 1989*a*,*b*). Two artificial short sequences were considered: one included the two helices connected by a link of ten glycine residues; the other included the five-strand β-sheet where each strand was connected by an eight-glycine residue link. These connections, completely unrestrained, were necessary because, in the present version, DISMAN does not deal with more than one chain. Starting structures were generated randomly. The number and the distribution of restraints were enough to determine uniquely the structures of the two pieces of secondary structure (an average r.m.s. deviation of 1·8 Å (1 Å = 0·1 nm) was obtained for the backbone atoms of the β-structures and of 1·5 Å for the 2-helix pair).

The best resulting structures were then used to start new runs in which the whole molecule was considered. Random structures were generated for the complete sequence. Each strand of the β-sheet and the two helix regions were placed with the conformations obtained from the previous runs. Five further runs of 2500 cycles of minimization of the target function that included the whole chain but turned on a progressively increasing number of NOEs were performed as follows. (1) Only the NOE list concerning the interactions within the strands of the β-sheet and within the helices was used. (2) The interactions between individual strands of the sheet and between the helices (but not between the sheet and the helices) were introduced. The first two steps led to only minor changes (caused mainly by slightly different van der Waals' interactions) in the structure of the pre-calculated β-sheet and helices. (3) All distances for residues in the sequential distance from i to $i+4$ were specified. Up to this stage, all the loops were completely unrestrained. (4) All interactions between the positions of the two main parts (α and β) of the secondary structure. (5) The complete list of all NOEs was used. The resulting structure was then submitted as a starting structure and the whole sequence (1) to (5) was repeated. Twenty structures were thus calculated and the five best results were selected for further refinement. Figure 1 shows the superposition of two consecutive results after each step.

This approach, originally required for practical reasons (namely the limitations in computer disc space and time; typically 4 h on a VAX 11/750 computer were needed), differs from the standard strategy described for the variable target function algorithm (typically 40 h of central processing unit time for 1 run; Braun, 1987). Usually, the interproton distances involving residues further and further apart in the sequence are incorporated into the target function irrespective of their location in the secondary structure elements. Our approach has the advantage of satisfying first elements of secondary structure, which are often the best defined. The loops are introduced only at a later stage, while they are initially completely unrestrained. This enables the optimization of the secondary structure units. This procedure allowed us to obtain an optimal arrangement of the structure and provided the best results. To confirm this, other more recent attempts to calculate the structure with the standard procedure always led to a higher value of the final target function.

Table 1

	Ac1	Ac2	Ac3	Ac4	Ac5
A. *Potential energy (E_{pot}) and restraint energy (E_{res}) at different stages of the calculations (after distance geometry (DM), energy minimization (EM) and further refinement (MD+EM)*†					
E_{pot}					
After DM	10^6	10^6	10^6	10^6	10^6
After EM	−4491	−4629	−4485	−4756	−4060
After MD+EM	−5935	−5958	−5752	−5989	−5971
E_{res}					
After DM	300	279	332	274	398
After EM	452	474	511	493	434
After MD+EM	334	359	389	314	302
Ac5	TAU0		TAUMD		
B. *Only 1 structure (Ac5) was chosen for 20 ps of time-averaged NOE dynamics (2·5 ps memory) and indicated here as TAUMD*‡					
E_{pot}	−3019		−3007		
E_{res}	298		233		

The energies are expressed in kJ mol, the force constant for a semiharmonic potential was 250 kJ mol^{-1} nm^{-2}.

† Comparison of the energy terms among the 5 final structures.

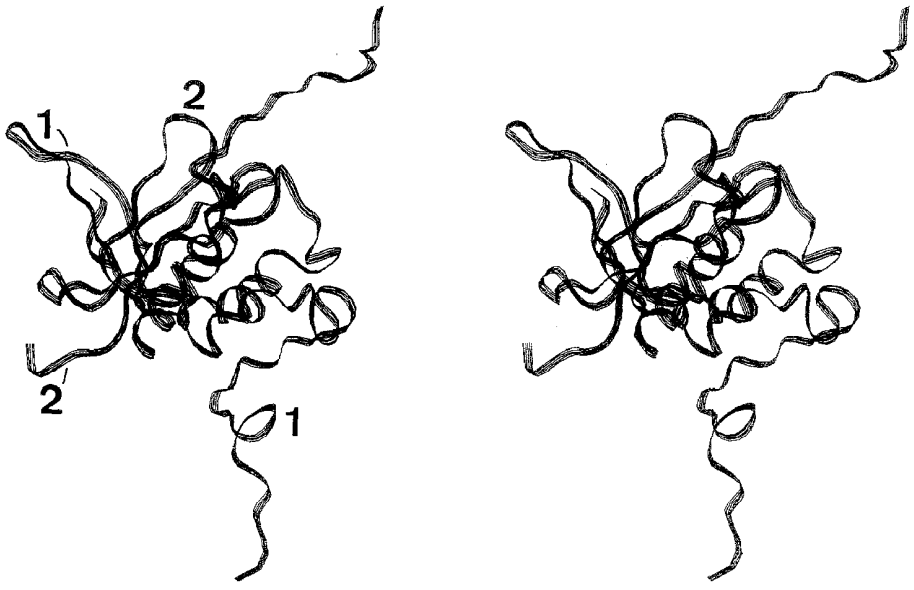
‡ The average potential and restraint energies along a 20 ps trajectory are shown. For comparison, the corresponding values for a trajectory with distance restraints enforced as static bounds (TAU0) starting from the same structure (the last snapshot of the previously calculated trajectory) are also shown.

(b) *Structure refinement*

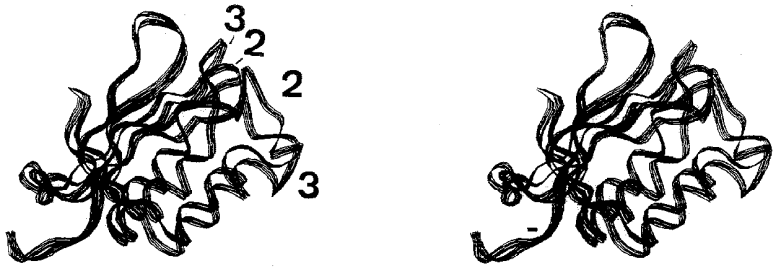
Five structures, obtained from DISMAN calculations, that satisfied the n.m.r. restraints best were refined by energy minimization and MD. As discussed above, the main aim of the dynamics is to regularize the structure and to explore better the conformational space (De Vlieg *et al.*, 1988). After MD, a much more regular hydrogen bond pattern was obtained (data not shown).

Several force constants were tried in preliminary runs. We found the best compromise was to use 250 kJ mol^{-1} nm^{-2}. With higher values, the SHAKE algorithm became unstable. We did not switch it off, however, because at high temperature and with so many n.m.r. restraints the bonds became unrealistically long. Although decreasing the temperature would cancel the effect, we preferred to run our trajectories with SHAKE and a force constant that was not too high.

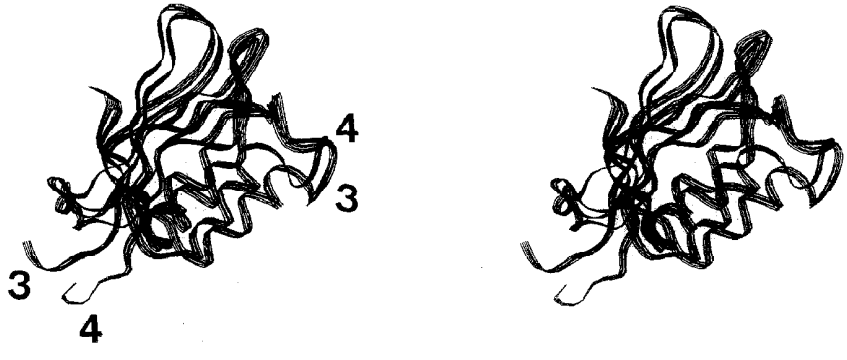
Table 1 shows the different energy terms at different stages of the refinement. After MD, relatively high residual violations could still be observed. In the large majority the violations were small (less than 0·5 Å). Some bigger violations were observed on distances involving some of the ten



(a)



(b)



(c)

Fig. 1.

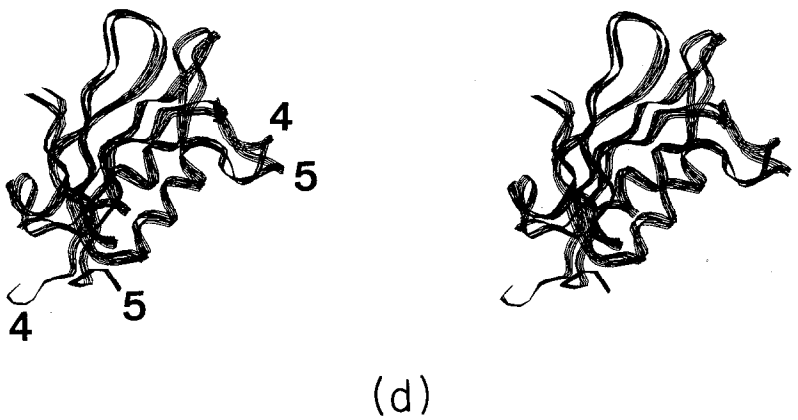


Figure 1. (a) to (d) Superposition of the structures resulting from 2 consecutive stages of the DISMAN calculation. The number of the distance restraints was increasing in 5 steps that followed the distribution of the NOEs according to their position in the secondary structure (for details, see the text). The procedure led to a progressive adjustment of the structure: (1) β-strands and helices positioned mutually at random; (2) the 5 β-strands arranged in a sheet and the 2 helices packed together, but the β and α-units still independent; (3) adjustment of all loops; (4) β and α-units packed together; (5) final arrangement according to all observed distances.

aromatic residues present (typically distances involving Y11, F14, F22, Y25, F80 and F94). They were known to be in rotation from the n.m.r. observation (Saudek *et al.*, 1989*c*). We therefore decided to run 20 picoseconds of time-averaged dynamics on one of the structures (Ac5) (Torda *et al.*, 1989, 1990), Table 1B shows the results. A smaller, averaged, residual violation energy is observed.

To prove that our results are energetically stable, an additional five picoseconds of unrestrained MD and 1000 cycles of unrestrained energy minimization were run, both starting from the final snapshot obtained of the dynamics. From the resulting structures and that of restrained MD, a r.m.s. deviation of 1·4 Å for all atoms and 1·0 Å for the backbone was obtained. These values are lower than the r.m.s. deviation between pairs of different solutions (see Table 2A). A stereo-picture of one of the final structures is shown in Figure 2.

(c) *Structure analysis*

The analysis of a structure is similar to an archival search. In archives, the rules followed by lost languages may be reconstructed from the study of old documents. In protein studies, the rules that determine protein folding and protein function may be elucidated by the analysis of known structures.

In this and in the following sections, we discuss: (1) whether there are new structural elements that have still not been recognized in crystals; (2) how the secondary structure elements pack together; (3) what we can say about the protein mobility; (4) whether the folding can be compared with any other structure; and (5) unresolved questions.

The first step in the analysis of molecular structure is the assignment of secondary structure elements, using consistent criteria. When the results are obtained from n.m.r. data, a family of solutions more or less all equally satisfying the experimental information is obtained. Even when the r.m.s. deviation between structures is small, it is still impossible, and probably meaningless, to talk about a unique structure. The use of Ramachandran plots may help, together with the r.m.s. deviation, in evaluating the differences between the structures.

Table 2A shows the average r.m.s. deviation between pairs of structures at different stages of the calculation. Almost no difference is found between

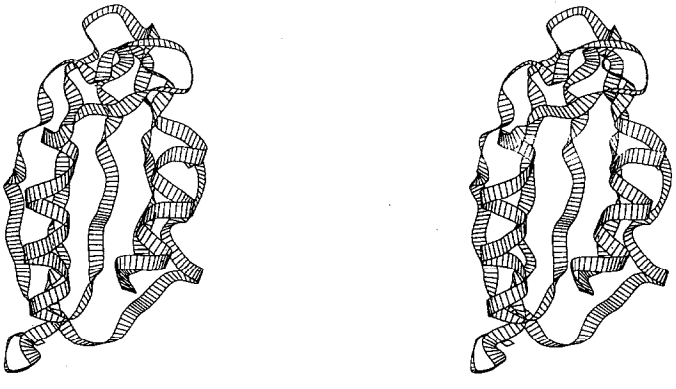


Figure 2. Stereo picture of one of the final structures of acylphosphatase obtained after refinement. This, as well as Fig. 4, was produced using the ARPLOT program kindly provided by A. M. Lesk.

Table 2

Average of the r.m.s. deviation (in Å) obtained comparing structure pairs

r.m.s.d.	DISMAN	EM	MD+EM
A. *Fitting all atoms, only the backbone (N, C^{α}, C), only the side-chains and only the backbone atoms of the helix pair and the β-sheet at different stages of the calculations*			
All	2·3±0·7	2·4±0·7	3·3±0·3
Backbone	1·9±0·6	2·0±0·6	2·5±0·3
Side-chains	2·6±0·8	2·7±0·8	3·9±0·3
H1, H2†	0·8±0·1	0·9±0·1	1·3±0·1
Sheet‡	0·7±0·1	0·8±0·1	1·5±0·1

	Range	r.m.s.d.
B. *Fitting the backbone atoms for each element of secondary structure*		
	1–6	1·7
B1	7–13	0·5
	14–22	1·9
H1	22–34	0·7
B2	35–42	0·8
	42–45	0·8
B3	46–53	0·9
	53–55	0·6
H2	55–65	0·9
	66–77	1·8
B4	77–85	0·9
	86–92	0·9
B5	93–97	0·7

r.m.s.d., root-mean-square deviation. Other abbreviations as in Table 1.

† Positions 22 to 35, 55 to 66 (backbone atoms only).

‡ Positions 7 to 14, 35 to 42, 46 to 53, 77 to 86, 94 to 97 (backbone atoms only).

the value before and after minimization, whereas an increase of the r.m.s. deviation after MD is a sign of a conformational search among conformations approximately equally consistent with the n.m.r. data.

Figure 3 shows the Ramachandran plots of two representative structures before and after refinement (the other plots are available as additional material). Two effects are detectable. First, a more regular distribution in the expected regions is observable for the residues in helical and in β-sheet conformation. Secondly, a small number of residues in the "non-allowed" regions can be observed after molecular dynamics. Whether these residues are real exceptions to the Ramachandran distribution should be considered very carefully. Of the eight glycine residues present in the sequence, four are present consistently in this region (G34, 37, 45 and 49). The other violations arise from residues occurring in turns or loops where there are fewer NOEs. Often the exact position of these violations is not conserved.

For most of our purposes, we use the best solution according to the final distance violations (Ac5), checking at the same time the general validity of our statements on the different solutions. The secondary structure assignment of acylphosphatase is summarized in Table 3. The criteria used for defining helices, β-strands and hydrogen bonds are those used in the program DSSP (Kabsch & Sander, 1983). Bond occurrence in the five structures is listed. Water accessibility along the amino acid sequence is also shown and compared with the location of the secondary structure elements.

(d) *Packing of secondary structure elements*

It has been widely demonstrated that the patterns of residue–residue contacts in the interfaces between packed elements of secondary structure in globular protein cores is one of the best preserved features in protein families (Lesk & Chothia, 1980). Unusual packings may be diagnostic of hidden similarities between even very distantly related families (Pastore & Lesk, 1990).

In acylphosphatase, the two helices run almost parallel to each other. In H1, two ridges are formed by residues Y25, E29 and I33, and by the less bulky residues A26, A30 and G34. W64 is placed at the centre of the groove formed by H1. In H2, two shorter ridges are formed by W64 (in contact with Y25, A26 and E29) and S60 (which faces I33) and by M61 (in contact with A26, E29 and A30) and K57 (in contact with I33). The position of the side-chains in this interface is well determined for W64, while residues F22, Y25 and M61 show some variability in the five final structures. Overall, the two helices pack together with a distorted $\pm 4n$ ridge (Chothia *et al.*, 1981).

The main-chain geometry of β-sheets is regular with a β-bulge at residue 81. A right-handed twist is shown by the β-sheet, as commonly found in parallel and antiparallel sheets in globular proteins (Janin & Chothia, 1980). The helix axis is parallel to the β-strands ($\omega = 0$, $\theta = 0$). The two helices pack onto the placed sheet with the helix axes parallel to the strands of the sheet, since it is only in this orientation that the two contact surfaces have complementary twists and can pack closely. Both the sheet and the helices are amphiphilic. The β-sheet provides one side of the enzyme structure, while the helices form the other side. They interact by means of a hydrophobic core built from the side-chains coming from both. The side-chains involved in the packing are selected to approximate a smooth surface. Figure 4 shows two pictures of the interfaces between the two helices and between helices and the sheet. The hydrophilic side-chains point into the solvent.

The acylphosphatase packing patterns may provide an important guide for future comparison with other proteins.

(e) *Mobility*

Protein motions cover a wide range of frequencies. The dynamics of major interest in the study of biological systems are often those with time constants of the order of a second down to one

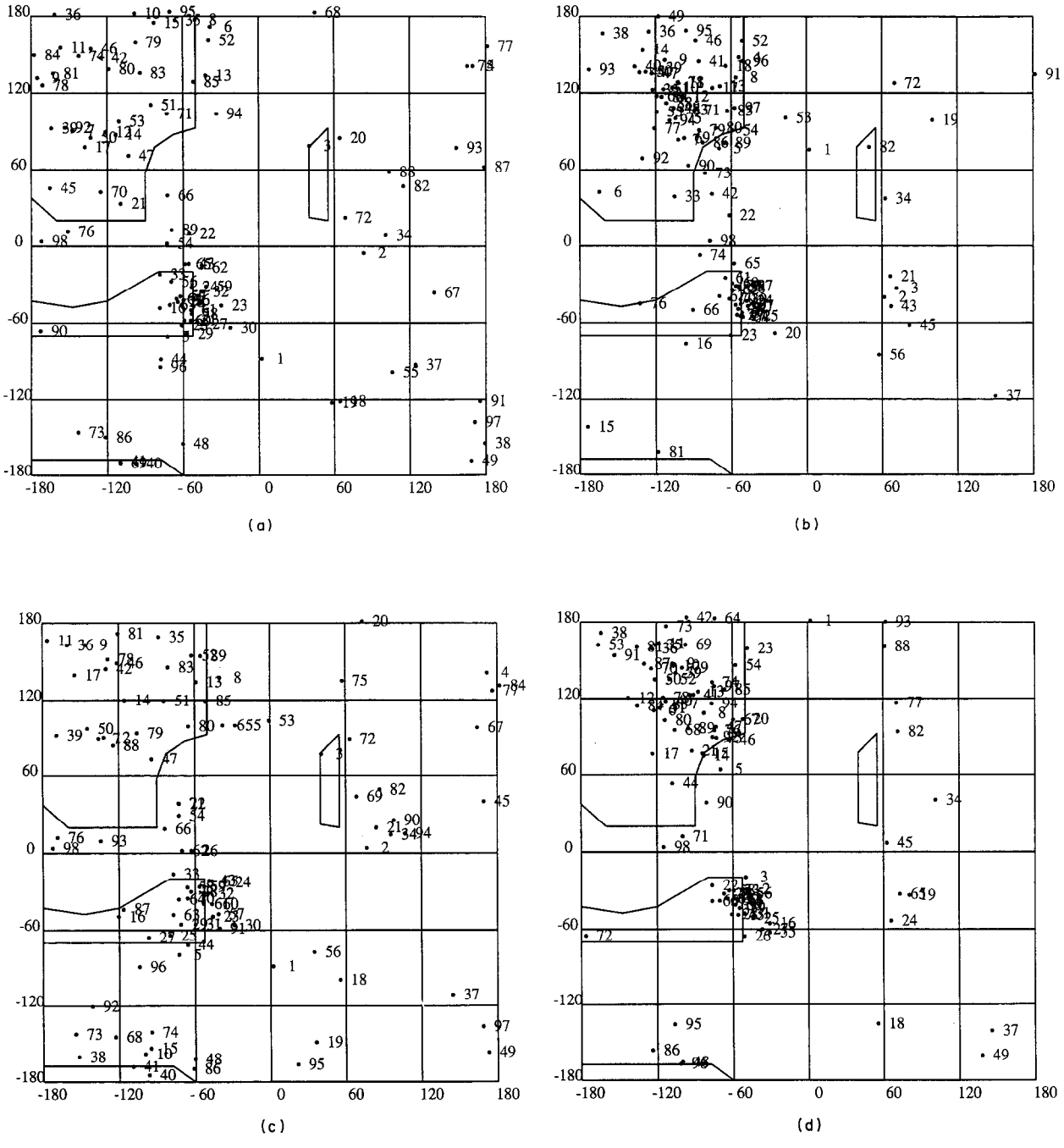


Figure 3. Ramachandran plot of 2 representative structures (Ac1 and 2) before ((a) and (c)) and after ((b) and (d)) energy refinement. Residues 15, 19, 34, 37, 45, 49, 53 and 69 are glycine. Four of them appear consistently in the region with positive ϕ angles. The other residues present in this half of the plot are involved in loop regions.

nanosecond (10^{-9} s). It is in this range that n.m.r. is more useful.

The possibility of a correlation between n.m.r. information and antigenic sites of acylphosphatase has been discussed elsewhere (Saudek *et al.*, 1989*c*,*d*). Here, we show additional indications of higher flexibility in some regions. Table 2B shows the r.m.s. deviation among the family of the final structures. A consistently lower variability is shown by the two helices and the β-sheet. The r.m.s. deviation has been correlated sometimes to the resolution achieved in the structure determination. However, we could take it also as a measure of the differential mobility of some parts of the structure. Figure 5 shows the NOE distribution along the sequence and the r.m.s. deviation of the α-carbon atoms along the sequence during an MD run. An inverse correlation between the two plots can be observed and it shows that the most mobile regions are also those that produce fewer NOE interactions.

The regions of maximum variability are positions 1 to 6, 14 to 22, 53 to 55, 66 to 77 and 86 to 92. Proteins have been described as "breathing". Figure 6(a) and (b) shows the behaviour during an MD run

Table 3

Secondary structure assignment, solvent accessibility and hydrogen bonds involving backbone atoms for Ac5

Res.	AA	Structure	Acc.	H-bonds
1	S		e	
2	T	T	e	
3	A	T	e	85
4	R	S	e	
5	P		e	
6	L	E	e	4
7	K	E	e	53
8	S	E	b	84
9	V	E	b	51
10	D	E	b	82
11	Y	E	b	49
12	E	E	e	79
13	V	E	b	47
14	F	E	e	77
15	G		b	
16	R	T	e	
17	V	T	e	
18	Q	S	e	
19	G		e	
20	V		b	
21	C		e	19
22	F		b	
23	R	T	e	
24	M	T	e	21
25	Y	H	b	21
26	A	H	b	22
27	E	H	b	23
28	D	H	e	24
29	E	H	e	25
30	A	H	b	26
31	R	H	e	27
32	K	H	e	28
33	I	H	e	29
34	G	T	b	
35	V		b	30
36	V	E	b	52
37	G	E	b	93
38	W	E	e	50
39	V	E	b	95
40	K	E	e	48
41	N		e	98
42	T		e	46
43	S	T	e	18
44	K	T	e	
45	G		b	
46	T	E	e	43
47	V	E	b	13
48	T	E	b	40
49	G	E	b	11
50	Q	E	e	38
51	V	E	b	9
52	Q	E	b	36
53	G	E	b	7
54	P	E	b	
55	E	H	e	5
56	E	H	e	
57	K	H	e	
58	V	H	b	54
59	N	H	e	55
60	S	H	e	56
61	M	H	b	57
62	K	H	e	58
63	S	H	e	59
64	W	S	b	60
65	L	S	b	61
66	S	S	e	62
67	K	S	e	63
68	V		e	
69	G		b	74
70	S		e	68
71	P	S	e	
72	S	S	e	70
73	S	S	b	17
74	R		e	
75	I		b	73
76	D		e	
77	R		e	75
78	T	E	b	65
79	N	E	e	12
80	F	E	e	
81	S	E	e	10
82	N	E	e	
83	E	E	e	
84	K	E	e	8
85	T	E	e	
86	I	E	e	6
87	S	S	b	
88	K	S	e	
89	L		e	
90	E		e	
91	Y	S	e	
92	S	S	e	
93	N	S	e	
94	F	B	b	
95	S		e	37
96	V		e	
97	R		e	
98	Y		e	39

Small variations (within ± 1 residue) are found for the structure assignment in the other 4 structures. The convention used is the same as that in the program DSSP (Kabsch & Sander, 1983): T, turn; E, extended; H, helical; S, bend. Lower-case letters in the 4th column indicate exposed (e) and buried (b) residues. The threshold used is 30 Å^3. Only hydrogen bonds involving backbone atoms are reported. They are marked when their occurrence in 20 ps of dynamics is higher than 50%. The number marked in the 5th column refers to the acceptor atom (oxygen) involved in the bond. Res., residue; AA, amino acid; Acc., accessibility.

of two representative distances between two β-sheet strands (HN of residues 12 and H^{β} of 46) and between the two helices (H^{γ} of residue 25 and H^{β} of 65). To allow more variability, the distances are taken from the trajectory with time-averaged distance restraints. In spite of the dense packing between the elements of secondary structure, motions are allowed.

(f) *Acylphosphatase folding*

Proteins are constructed from a limited number of architectural elements. The most common arrangement of α-helices and β-pleated-sheets in proteins is for the β-sheet to be at the centre of the molecule, with the α-helices packed on both its faces.

Acylphosphatase is a β–α–β protein, forming an "open-sandwich" fold. Two β–α–β units with a right-handed crossover intercalate so that the four strands of β-sheets are all antiparallel and the two helices pack against each other (Fig. 7). This leads to 4-1-3-2 topology of the β-strands. When the acylphosphatase structure was first published, only

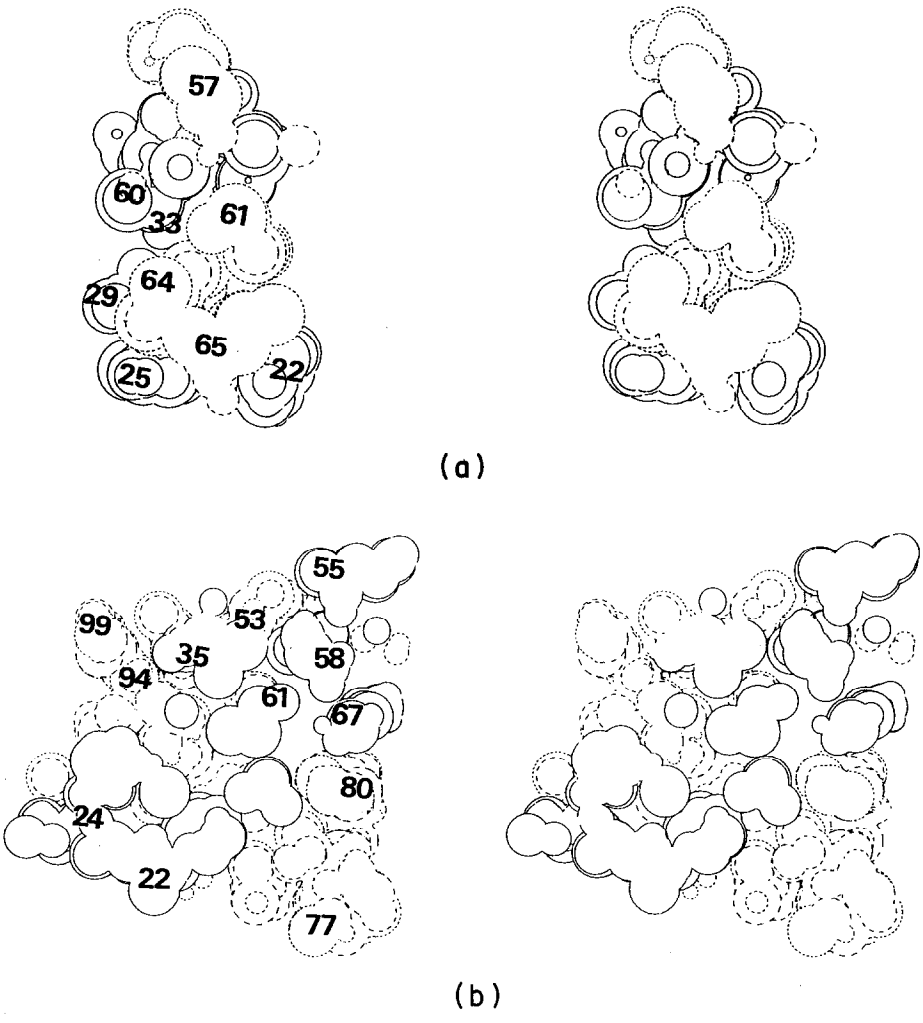


Figure 4. (a) Helix packing and (b) helix–sheet interface.

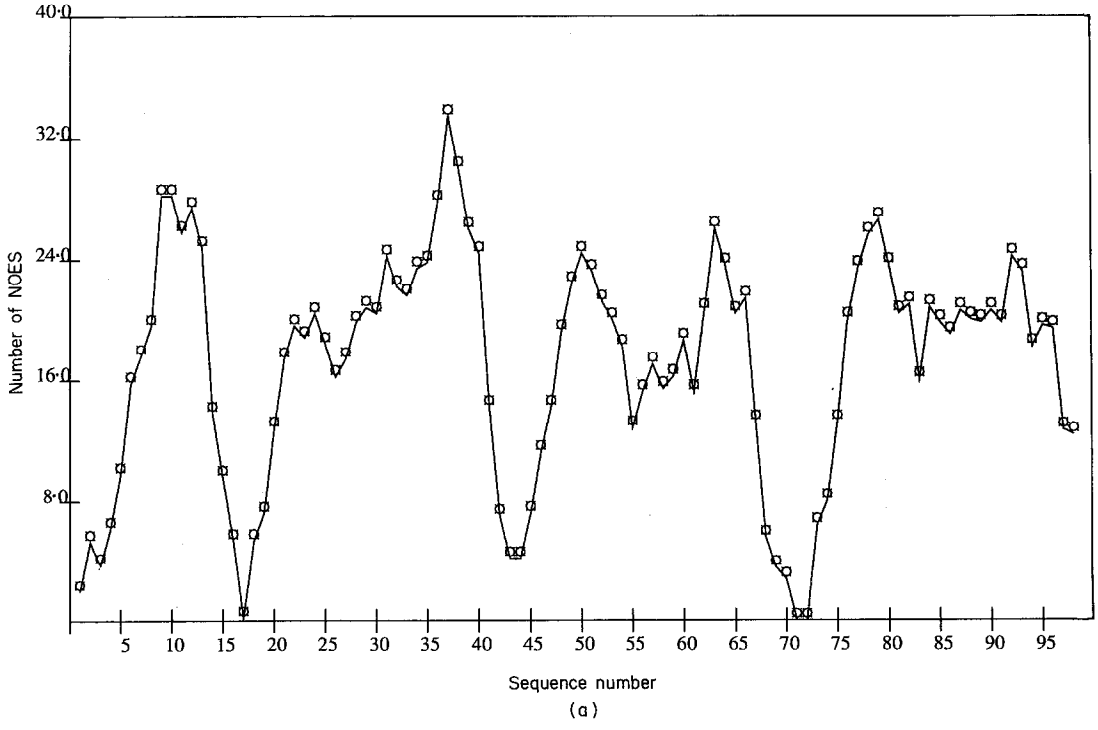


Fig. 5.

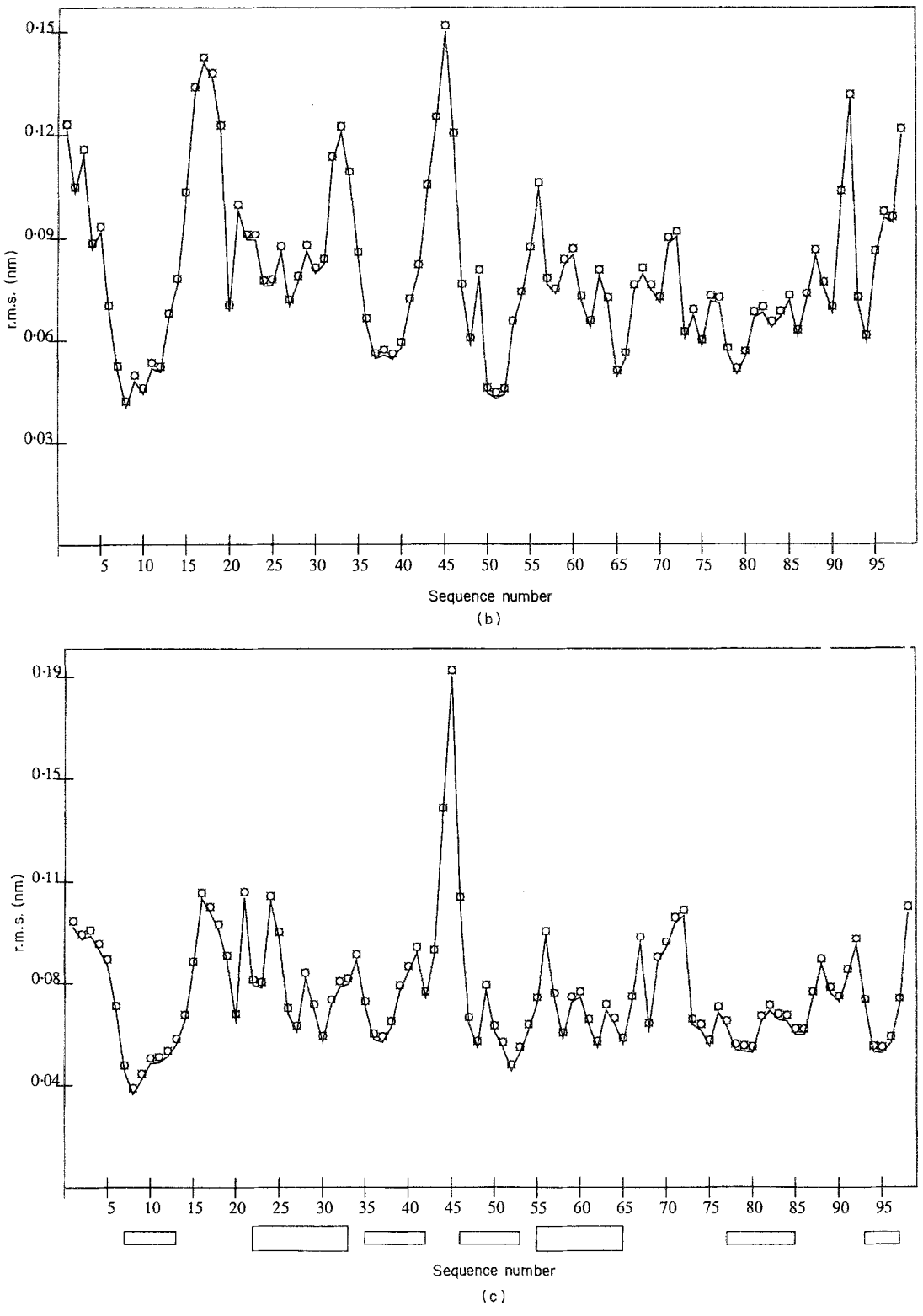


Figure 5. (a) NOE distribution and (b) and (c) r.m.s. deviation of C^{α} atomic positions along the sequence during trajectory (b) with and (c) without time averaging (see the text). Thin and thick boxes indicate the sheet and the helices, respectively.

ferredoxin presented a similar folding (Adman *et al.*, 1973).

A few more examples have appeared recently in the literature. The activation domain of procarboxypeptidase B, solved both by n.m.r. (Vendrell *et al.*, 1991) and by X-ray crystallography (Coll *et al.*, 1991), and the RNA-binding domain of the small nuclear ribonucleoprotein A (U1; Nagai *et al.*, 1990), show the same folding and the same topology as acylphosphatase. Of the two proteins, procarboxy-

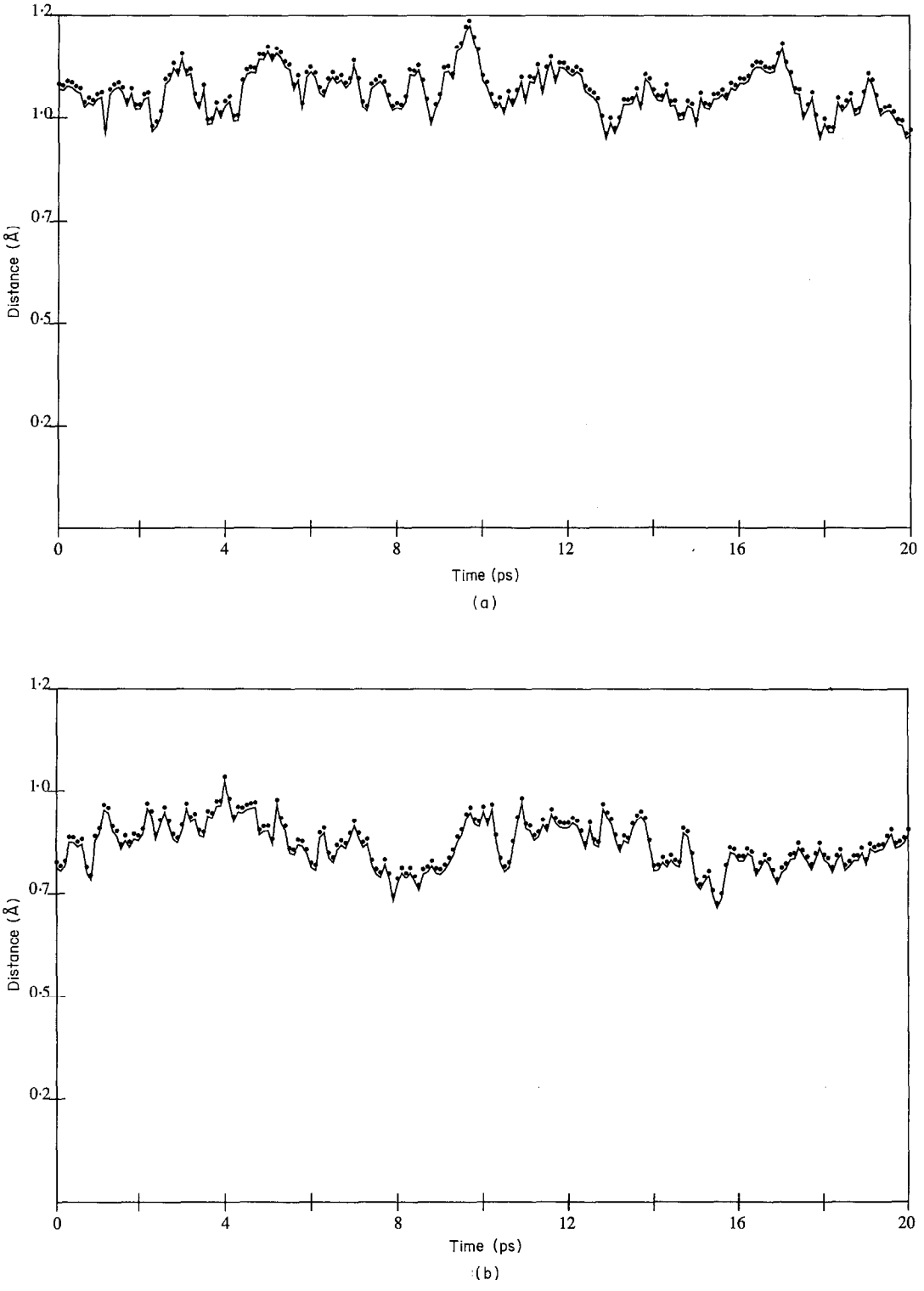


Figure 6. Interatomic distances (in Å) as a function of time during a 20 ps trajectory of structure Ac5 using time-averaged distance restraints: (a) distance between the HN of residue 12 and H^{β} proton of residue 46 (strands B1 and B3); (b) distance between the H^{γ} proton of residue 25 and H^{β} of residue 65 (belonging to helices H1 and H2, respectively).

peptidase shows a much more compact packing, with the two helices, as in acylphosphatase, parallel to each other and packed together into a concave sheet. In U1 ribonucleoprotein, only one helix packs into the sheets, while the other is almost perpendicular to the first. As very little, if any, sequence identity is shared by acylphosphatase, procarboxypeptidase B and ribonucleoprotein A, it is worth trying to establish whether the proteins are evolutionarily related or whether they have a common solution to different functional requirements.

The second possible intercalation of the two units

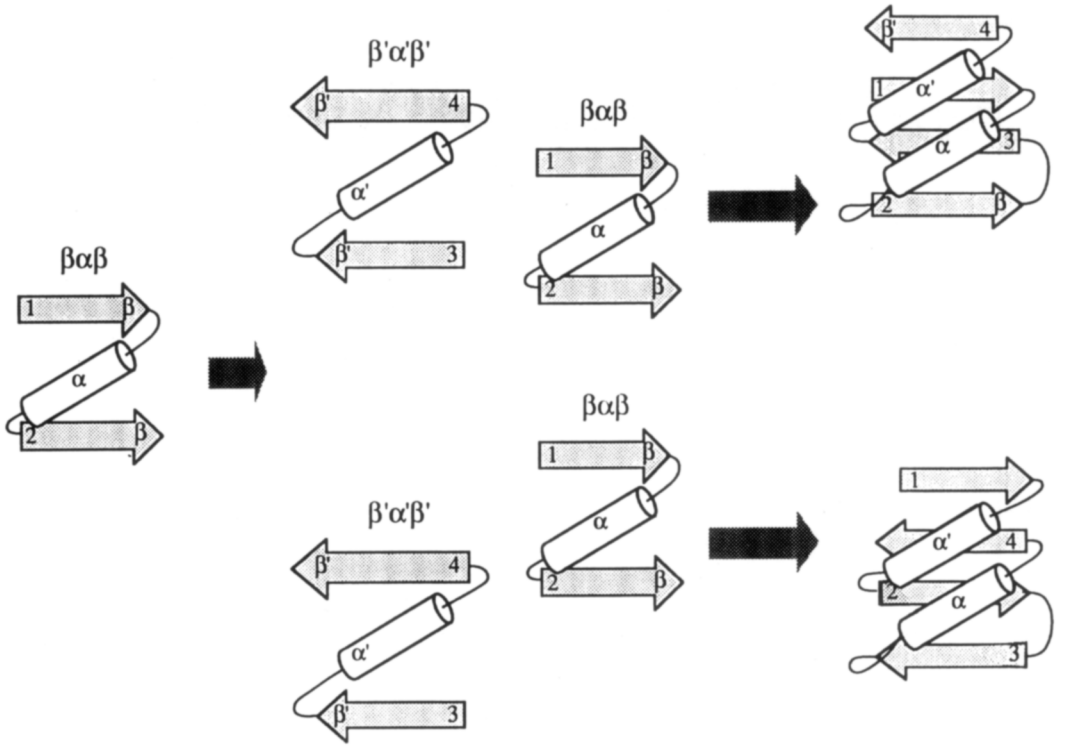


Figure 7. Two folding β–α–β units are intercalated in 2 different foldings. The first, displayed here, leads to a 4-1-3-2 topology of the β-strands. Acylphosphatase, the activation domain of procarboxypeptidase B and ribonucleoprotein A share this folding. The other intercalation of the 2 units is present in HPr proteins. The 2 foldings are not topologically equivalent because one cannot be generated from the other by simply breaking and repositioning a link between strands.

in Figure 7 gives a 1-4-2-3 topology, which has been found in HPr (histidine-rich protein; Witterkind *et al.*, 1989). This folding, although not at all topologically equivalent to the previous, shows the same intercalation of the two β–α–β motifs. The topology of these proteins depends on the two possible relative positions of the β–α–β units and not on the way in which they are connected. Ribosomal proteins from *Escherichia coli* L7/L12 (Leijonmarck & Lijias, 1987) and L30 (Wilson *et al.*, 1986) share the same folding as HPr but the first strand and the first helix are missing.

In all cases a hydrophobic core is sandwiched between the helices and the β-sheet and the domain exhibits the same type of autonomy as a β-barrel or a helical bundle.

(g) *Unresolved questions: glutathione and the active site*

Acylphosphatase was isolated with a glutathione group covalently linked by a disulphide bond at C21 (Stefani *et al.*, 1986). No indication of the presence of this group *in vivo* has been found. Glutathione is not important for activity and it has been suggested that it becomes attached only during the isolation procedure. The position of glutathione relative to the rest of the sequence could not be determined by n.m.r. because virtually no NOEs across the two chains could be detected.

The position of the active site is still under investigation. Because of the nature of the substrate, the active site should involve positively charged groups (R or K, since no H is present). Chemical modifications on the arginine residues make the enzyme inactive (M. Stefani, personal communication). Figure 8 shows the alignment of other sequences of the family, with the completely conserved residues marked. Preliminary experiments with paramagnetic probes seemed to indicate that the C-terminal regions should be involved.

4. Conclusions

We have presented here the structure refinement of horse muscle acylphosphatase based on n.m.r. data and discussed a number of structural features. Acylphosphatase is a globular protein and follows most of the general principles known from protein studies. Two β–α–β units are intercalated into each other in the folding. The packing of two helices and of a five-strand β-sheet is compact. A hydrophobic core holds the two units. A number of other proteins show a similar folding motif. As with β-barrels or four-helix bundles, it might be challenging to design this folding by protein engineering.

We thank Dr M. Billeter, Mrs M. Holmes, Professor A. M. Lesk and Dr N. Taddei for critical reading of the manuscript.

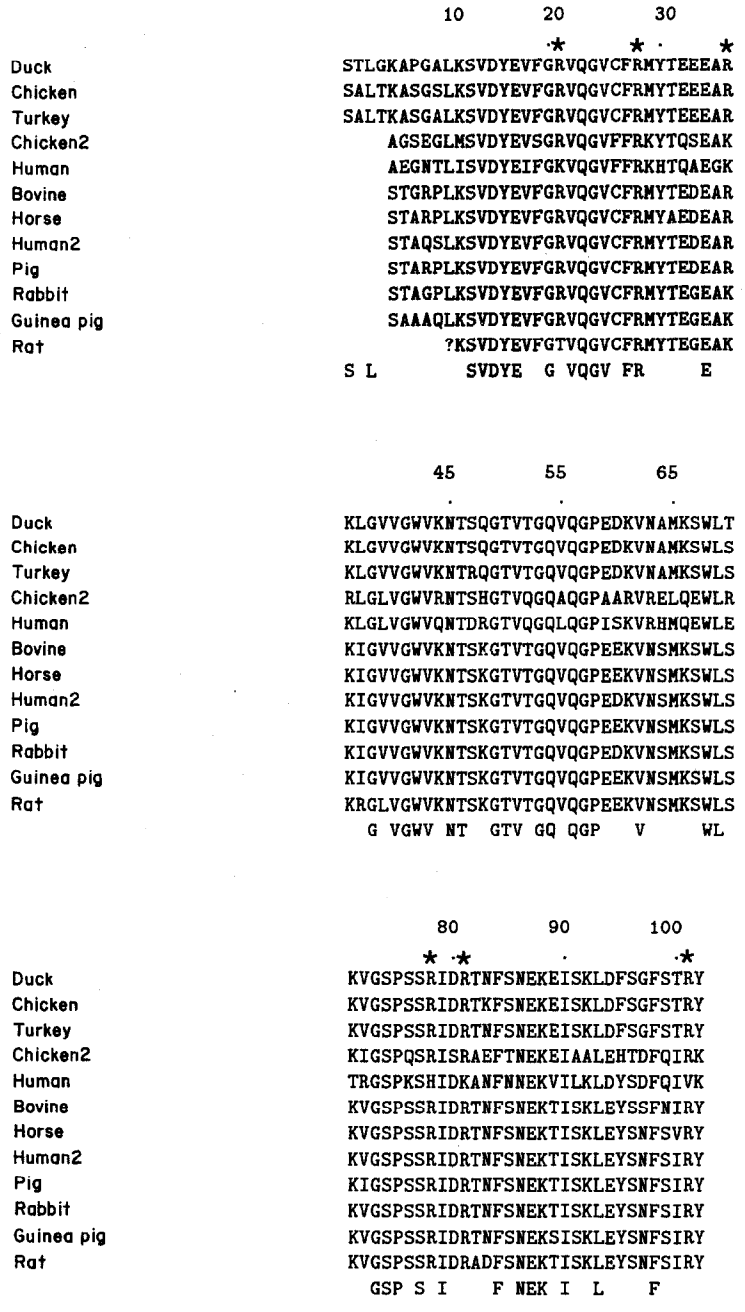


Figure 8. Alignment of 12 acylphosphatase sequences from mammals and birds. Alignment of other sequences of the family with the residues completely conserved are indicated on the bottom line. A star marks the positions where arginine residues are highly conserved.

References

Adman, E. T., Sieker, L. C. & Jensen L. H. (1973). Structure of a bacterial ferredoxin. *J. Biol. Chem.* **248**, 3987–3996.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. (1981). Interaction models for water in relation to protein hydration. *Jerusalem Symp. Quantum Chem. Biochem.* **14**, 331–342.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Braun, W. (1987). Distance geometry and related methods for protein structure determination from NMR data. *Quart. Rev. Biophys.* **19**, 115–157.

Braun, W. & Go, N. (1985). Calculation of protein conformations by proton–proton distance constraints: a new efficient algorithm. *J. Mol. Biol.* **186**, 611–626.

Bruenger, A. T. (1991). Refinement of three-dimensional structures of proteins and nucleic acids. In *Molecular Dynamics – Applications in Molecular Biology* (Goodfellow, J. M., ed.), pp. 137–178, MacMillan Press, London.

Chothia, C., Levitt, M. & Richardson, D. (1981). Helix to helix packing in proteins, *J. Mol. Biol.* **145**, 215–250.

Coll, M., Guasch, A., Aviles, F. X. & Huber, R. (1991).

Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity. *EMBO J.* **10**, 1–9.

Dayringer, H. E., Tramontano, A., Sprang, S. R. & Fletterick, R. J. (1986). Interactive program for visualization and modeling of proteins, nucleic acids and small molecules. *J. Mol. Graph.* **4**, 82–87.

De Vlieg, J., Scheek, R. M., van Gunsteren, W. F., Berendsen, H. J. C., Kaptein, R. & Thomason, J. (1988). Combined procedure of distance geometry and restrained molecular dynamics techniques for protein structure determination from NMR data: application to the DNA binding domain of Lac repressor from *Escherichia coli*. *Proteins: Struct. Funct. Genet.* **3**, 209–218.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Janin, J. & Chothia, C. (1980). Packing of alpha-helices onto beta-pleated sheets and the anatomy of alpha/beta proteins. *J. Mol. Biol.* **143**, 95–128.

Leijonmarck, M. & Liljas, A. (1987). Structure of the C-terminal domain of the ribosomal protein L7/L12 from *Escherichia coli* at 1·7 Å. *J. Mol. Biol.* **195**, 555–580.

Lesk, A. M. & Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **136**, 225–270.

Liguri, G., Nassi, P., Camici, G., Manao, G., Cappugi, G., Stefani, M., Berti, A. & Ramponi, G. (1984). Studies on synthesis and degradation rates and some molecular properties of guinca-pig muscle acylphosphatase. *Biochem. J.* **217**, 499–505.

Pastore, A. & Lesk, A. M. (1990). Comparison of the structures of globins and phycocyanins: evidence for evolutionary relationship. *Proteins: Struct. Funct. Genet.* **8**, 133–155.

Ramponi, G. (1975). 1,3-Diphosphoglycerate phosphatase. *Methods Enzymol.* **42**, 409–426.

Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Compt. Phys.* **23**, 327–341.

Saudek, V., Williams, R. J. P. & Ramponi, G. (1988). Secondary structure of acylphosphatase from rabbit skeletal muscle. A nuclear magnetic resonance study. *J. Mol. Biol.* **199**, 233–237.

Saudek. V., Atkinson, R. A., Williams, R. J. P. & Ramponi, G. (1989*a*). Identification and description of alpha-helical regions in horse muscle acylphosphatase by ^{1}H nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **205**, 229–239.

Saudek, V., Wormald, M. R., Williams, R. J. P., Boyd, J., Stefani, M. & Ramponi, G. (1989*b*). Identification and description of beta-structure in horse muscle acylphosphatase by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **207**, 405–415.

Saudek, V., Williams, R. J. P. & Ramponi, G. (1989*c*). The structure and properties of horse muscle acylphosphatase in solution. Mobility of antigenic and active site regions. *FEBS Letters*, **242**, 225–232.

Saudek, V., Williams, R. J. P., Stefani, M. & Ramponi, G. (1989*d*). Mobility of secondary structure units of horse-muscle acylphosphatase. Relation to antigenicity. *Eur. J. Biochem.* **185**, 99–103.

Stefani, M., Modesti, A. Camici, G. Manao, G., Cappugi, G. Berti, A. & Ramponi, G. J. (1986). Duck skeletal muscle acylphosphatase: primary structure. *J. Protein Chem.* **5**, 307–321.

Torda, A. E., Scheek, R. M. & van Gunsteren, W. F. (1989). Time-dependent distance restraints in molecular dynamics simulations. *Chem. Phys. Letters*, **157**, 289–294.

Torda, E. A., Scheek, R. M. & van Gunsteren, W. F. (1990). Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. *J. Mol. Biol.* **214**, 223–235.

van Gunsteren, W. F. (1989). On testing theoretical models by comparison of calculated with experimental data. In *Studies in Physics and Chemistry*, vol. 71, pp. 463–478, Proceedings of an International Meeting, Nancy, France.

van Gunsteren, W. F. & Berendsen, H. J. C. (1977). Algorithms for macromolecular dynamids and constraint dynamics. *Mol. Phys.* **334**, 1311–1327.

van Gunsteren, W. F., Kaptein, R. & Zuiderwerg, E. R. P. (1984). Use of molecular dynamic computer simulations when determining protein structure by 2D NMR. In *Nucleic Acid Conformation and Dynamics* (Olson, W. K., ed.), pp. 79–92, Proceedings of NATO/CECAM Workshop, Orsay, France.

Vendrell, J., Billeter. M., Wider, G., Aviles, F. X. & Wuethrich, K. (1991). The NMR structure of the activation domain isolated from porcine procarboxypeptidase B. *EMBO J.* **10**, 11–15.

Wilson, K. S., Appelt, K., Badger, J., Tanaka, I. & White, S. W. (1986). Crystal structure of a prokaryotic ribosomal protein. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 7251–7255.

Wittekind, M., Reizer, J., Deutscher, J., Saier, M. H. & Klevit, R. E. (1989). Common structural changes accompany the functional inactivation of HPr by seryl phosphorylation or by serine to aspartate substitution. *Biochemistry*, **28**, 9908–9912.

Edited by M. Moody