Eur Biophys J (1995) 24: 1–11

European Biophysics Journal

Structure and atomic fluctuation patterns of potato carboxypeptidase a inhibitor protein

A molecular dynamics study in water

B. Oliva[1], X. Daura[1], E. Querol[1], F. X. Avilés[1], O. Tapia[2]

[1] Institut de Biologia Fonamental i Departament de Bioquímica, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain
[2] Department of Physical Chemistry, University of Uppsala, Box 532, S-751 21 Uppsala, Sweden

Received: 26 October 1994 / Accepted in revised form: 26 January 1995

Abstract. Molecular dynamics (MD) simulation methods were applied to the study of the structural and dynamic fluctuation properties of the potato carboxypeptidase A inhibitor protein (PCI) immersed in a bath of 1259 water molecules. A trajectory of 200 ps was generated at constant temperature and pressure. The crystallographic structure of PCI, as found in its complex with bovine carboxypeptidase A (CPA), was used to seed the MD simulation. Analyses show that the structure of the PCI core is fairly rigid and stable in itself, and that little deformation is caused by the protein-protein interactions found in the PCI-CPA complex. The N-terminal tail fluctuates to approach the core structure and appears as a relatively disordered region. In contrast, the conformations of the C-terminal tail, which is involved in the inhibitory mechanism, fluctuates in the neighborhood of the X-ray structure in orientations which facilitate CPA binding. Comparison with the structural entries for PCI in water obtained from both 2D-NMR experiments and X-ray data shows that important features of the MD structural results fluctuates between the initial crystal values and those obtained from the NMR solution structure. This fluctuation is not uniform; minor regions move away from the X-ray conformation while they do not approach the NMR conformation. The results reported suggest that the trajectory is long enough to show a behavior that is consistent with the conformational space available to the protein in solution.

Key words: Carboxypeptidase inhibitor – Molecular dynamics – Computer simulation – NMR/X-Ray comparison – Solvent effects

Abbreviations: (CPA), Carboxypeptidase; (DG), Distance Geometry; (NMR), Nuclear Magnetic Resonance; (NIS) Non Inertial Solvent; (MD), Molecular Dynamics; (PBC), Periodic Boundary Conditions; (PCI), Potato Carboxypeptidase Inhibitor; (RMSD), Root Mean Square Deviation; (a.m.u.) Atomic mass units

Correspondence to: O. Tapia

1. Introduction

Potato carboxypeptidase inhibitor (PCI) is a protein of 39 residues involved in defense mechanisms in vegetables (Hass and Ryan 1981). It may have biotechnological applications in relation to this biological function as well as biomedical uses (Ryan 1989). The X-ray crystal structure of a PCI complex with carboxypeptidase-A (Rees and Lipscomb 1982) and the NMR structure of PCI alone in water (Clore et al. 1987) are known. Potato carboxypeptidase inhibitor is one of the smallest globular proteins known and a particularly suitable system for studies combining computer simulation of mutants and site-directed mutagenesis. Molecular dynamics studies of PCI using an implicit solvent representation (NIS model) were previously carried out to explore the possibilities offered by this technique as a tool to help design mutants (Oliva et al. 1991 a, b). In spite of the relative success in identifying structural determinants using these *quasi*-in-vacuum simulations, a systematic and more realistic approach is required to check the effects that explicit water and counterions may have on the structure and fluctuation patterns. In addition, and in a more technical vein, such detailed simulations may help as a benchmark to improve simple solvation models. Here, a 200ps molecular dynamics trajectory has been studied for PCI in a box with 1259 water molecules and 8 counterions. The main objectives are threefold: (1) to characterize this protein in detail with a more realistic model; (2) to compare its structural features with those derived from X-ray and 2D-NMR; and (3) to check the reliability of the NIS-MD methodology in simulating protein conformations and fluctuations patterns.

Several methods have been used in the past to represent aqueous solvent effects in protein simulations (Karplus and Petsko 1990; van Gunsteren and Mark 1992; Berendsen et al. 1984; Aqvist et al. 1985; Straatsma and McCammon 1990; van Gunsteren and Berendsen 1990). The most rigorous approach – the direct inclusion of water molecules around the protein – is highly time-consuming and has been limited to its use with small peptides/proteins or by research teams with large computing facilities.

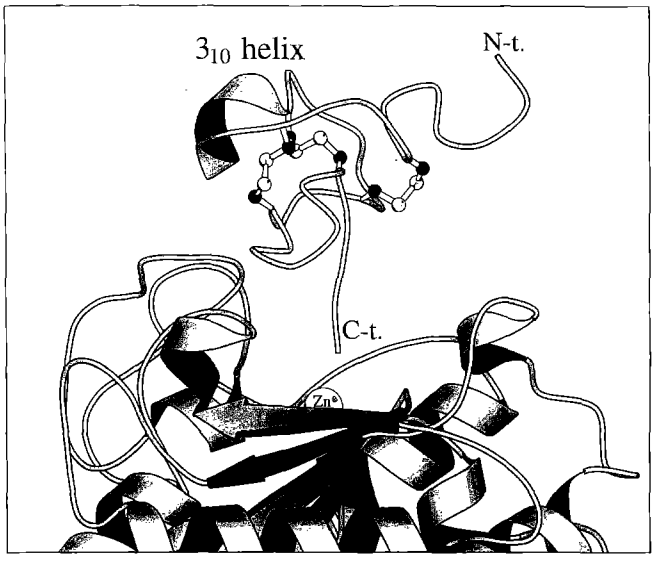

Fig. 1. Three-dimensional structure of the PCI-CPA complex. A ribbon-like representation of the complex was derived from the crystallographic coordinates. The main characteristic regions of PCI are indicated. Residues 1 and 39 (Glu and Gly, respectively) were not taken into account in the MD-analyses presented in this work

The analyses of trypsin inhibitors using an empirical solvation model are examples (Chiche et al. 1989; Vila et al. 1991). Developing efficient algorithms to simulate the presence of water in a simple and rapid fashion is therefore one of the challenges in this field. Examples of such algorithms are the following: stochastic dynamics (Yunyu et al. 1988), hydrophobic potentials (Casari and Sippl 1992; von Freyberg and Braun 1993), solvation force (Gilson and Honig 1988; Wesson and Eisenberg 1992, Stouten et al. 1993) and hydration prediction (Roe and Teeter 1992). The development of simpler methods, however, requires prior knowledge of the behaviour of proteins in full solvent representation and the selection of proteins of a size and composition suitable for the comparative studies of the different approaches.

The protein here analyzed, PCI, fulfils these requirements and, at the same time, provides an interesting model for MD simulations given its low percentage of regular secondary structures, high percentage of cysteines, and folding in three regions of different physico-chemical properties. The recent finding that this protein is representative of a structural folding module in small cysteine-rich domains, identified in several distinct protein families (Holm and Sander 1993), increases the biological interest of this study.

Description of PCI structure and functional determinants

Structurally, PCI can be divided into three parts: the core (residues 8–34), the N-terminal tail (residues 2–7) and the C-terminal tail (residues 35–38; see Fig. 1). The core is constrained by three disulphide bridges: Cs-8 to Cys-24, Cys-12 to Cys-27 and Cys-18 to Cys-34. Both tails are attached to the core through a proline residue. The core contains a 3_{10} helix (residues 14–18), generated by two consecutive type III reverse turns, as visualized by X-ray crystallography (Rees and Lipscomb 1982), whilst the NMR studies (Clore et al. 1987) position this helix in a slightly shifted region (residues 16–21). Moreover, the core is structured by five other reverse turns. The core also contains the secondary binding region to CPA, residues His-15 and Cys-27 to Ser-30 (Rees and Lipscomb 1982).

The N-tail does not seem to be involved in the PCI function, although it is involved in the crystal structure through the contacts that Gln-2 establishes with Arg-2 of a CPA molecule that belongs to a neighbouring symmetric cell (Rees and Lipscomb 1982). The C-tail (residues 35–39) is the primary binding region to CPA (Hass and Ryan 1981; Rees and Lipscomb 1982). The inhibitory mechanism of PCI acts through strongly competitive binding to the enzyme. Inhibitor binding constant to CPA (K_i) is in the nM range (Hass and Ryan 1981, Rees and Lipscomb 1982). The functional importance of the primary and secondary contact sites to the CPA enzyme have been experimentally proved by chemical modification studies (Hass and Ryan 1981), as well as by X-ray crystallography (Rees and Lipscomb 1982).

2. Methods

Computational details

The MD-simulations were performed with the GROMOS package (van Gunsteren and Berendsen 1987). Initial coordinates for the simulations were taken from the X-ray structure of the PCI-IIa isoform in the complex with CPA. Two residues were removed from the model: Glu-1 in the N-terminus and Gly-39 in the C-terminus (see Fig. 1). The former is located at some distance from the inhibitory region, is not well refined in the X-ray structure, and is lacking in certain isoforms. The latter is severed from PCI by CPA; it is not required for inhibition, and is lacking in homologous forms. The protein model contains 349 atoms. The total potential energy for the uncomplexed PCI from the crystallographic structure is large (positive) in the GROMOS NIS force field (Oliva et al. 1991 a, b). This is to be expected because stabilizing electrostatic interactions between the inhibitor and the carboxypeptidase have been removed. Similar problems were found by Clore et al. (1987) in their NMR study of PCI in aqueous solution.

Special care was exercised to obtain a low-energy initial conformation for the uncomplexed PCI before starting simulations in water. First, the isolated protein was cooled by steepest descent energy optimization (1800 steps; NIS field; no SHAKE). The potential energy decreased by 4000 kJ/mol, with no significant structural changes, and reached a negtive value. Water molecules were placed around the PCI to fill a rectangular box of 30.4 Å×32.2 Å×43.6 Å. The number of solvent molecules within the box, after the removal of those with oxygen atoms within 2.3 Å of a non-hydrogen protein atom, was 1259. At this point the solvent was energy-optimized using 300 steps of steepest descent minimization and with

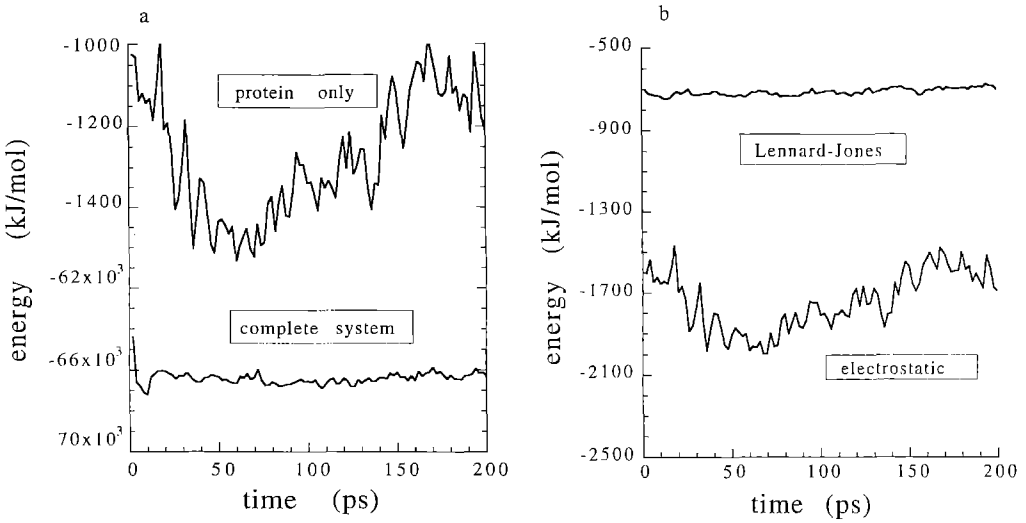

Fig. 2a, b. Time series of the energetic terms in the MD simulation of PCI. **a** Total potential energy with respect to the time of the PCI molecule; 4 Na, 4 Cl and 1259 water molecules were included as solvent, and Potential energy of the PCI molecule only. **b** Potential energy of the nonbonded interaction (electrostatic and Lennard-Jones) and Potential energy of the bonded interactions of the protein (bond angles, proper and improper dihedrals)

the protein atoms constrained. Counterions, 4 Na^+ and 4 Cl^-, were added by substituting 8 water molecules to obtain electroneutrality. The solvent conformation was relaxed again using 300 steps of steepest descent minimization with the protein atoms constrained. The resulting system was used to seed the MD simulation.

The original GROMOS C4 force field parameters (van Gunsteren and Berendsen 1987), and the new ones recently modified by Berendsen and coworkers (van Buuren et al. 1993), were used for the simulations in water. Here we report only the results obtained with the new set of parameters, whilst the comparison with the old ones will be reported elsewhere. In both cases, ionizable residues are given net charges of –1 and +1 (–1 for the carboxylic group of the aspartates, the glutamates and the C-terminal; +1 for the ammonium group of the lysines, guanidinium group of the arginines and ammonium group of the N-terminal) while partial charges are given for each atom. In contrast, the GROMOS D4 force field in the NIS model, used in previous studies (Oliva et al. 1991 a, b) gives only partial charges for each atom while it keeps the complete chemical groups electroneutral. In the new parameter set, the charges on water were those of the SPC/E model (van Buuren et al. 1993). The Lennard-Jones interactions between the oxygen of water with carbon atoms have been modified to better describe hydrophobic effects.

Periodic boundary conditions (PBC) were used. The pressure was kept constant at 10^5 Pa with a relaxation time of 0.05 ps. The temperature was kept constant at 293 K, with independent thermal baths for the protein (τ=0.1 ps) and the solvent (τ=0.1 ps). Electrostatic forces were evaluated using a clean twin-range cut-off technique. All nonbonded forces within 8 Å were evaluated with every simulation step. Electrostatic forces at distances 8–13 Å were reevaluated every 10 steps. The integration time was 2 fs. Bonds were constrained using SHAKE (van Gunsteren and Berendsen 1987).

3. Results

a. Energies

Energy entries. Time series showing the quality of the MD simulation in water are briefly presented. Energy entries illustrate its behaviour during the trajectory. After 10 ps the total potential energy of the whole system reaches equilibrium at an average of –66 MJ/mol with fluctuations of ±1 MJ/mol (Fig. 2a).

The potential energy of the protein alone is depicted in Fig. 2a. The fluctuations along the trajectory show that this is an open system exchanging energy with the surrounding water molecules. The Lennard-Jones energy is fairly well stabilized (Fig. 2b). Not surprisingly, it is the protein internal electrostatic interaction term that undergoes the greatest changes. On the other hand, bond and dihedral angles of the protein are also equilibrated after a frew picoseconds (not shown).

The results indicate a (stochastic) energy exchange between the protein and the solvent due to solvent collisions. These results also indicate that, within a cut-off of 8 Å, local changes are small. As the electrostatic forces have a cut-off of 13 Å, they are able to sense energy fluctuations that most likely come from interactions between the protein and the counterions. No attempts were made to sense the effect of periodic boundary conditions. The shortest box-side being larger than the electrostatic cut-off, the protein images do not interact among themselves.

b. Structural fluctuations

Time series of moments of inertia. The principal moments of inertia (pmi), calculated by the diagonalization of the tensor of the moments of inertia for all the atoms of PCI (where the polar hydrogens are included and the mass of

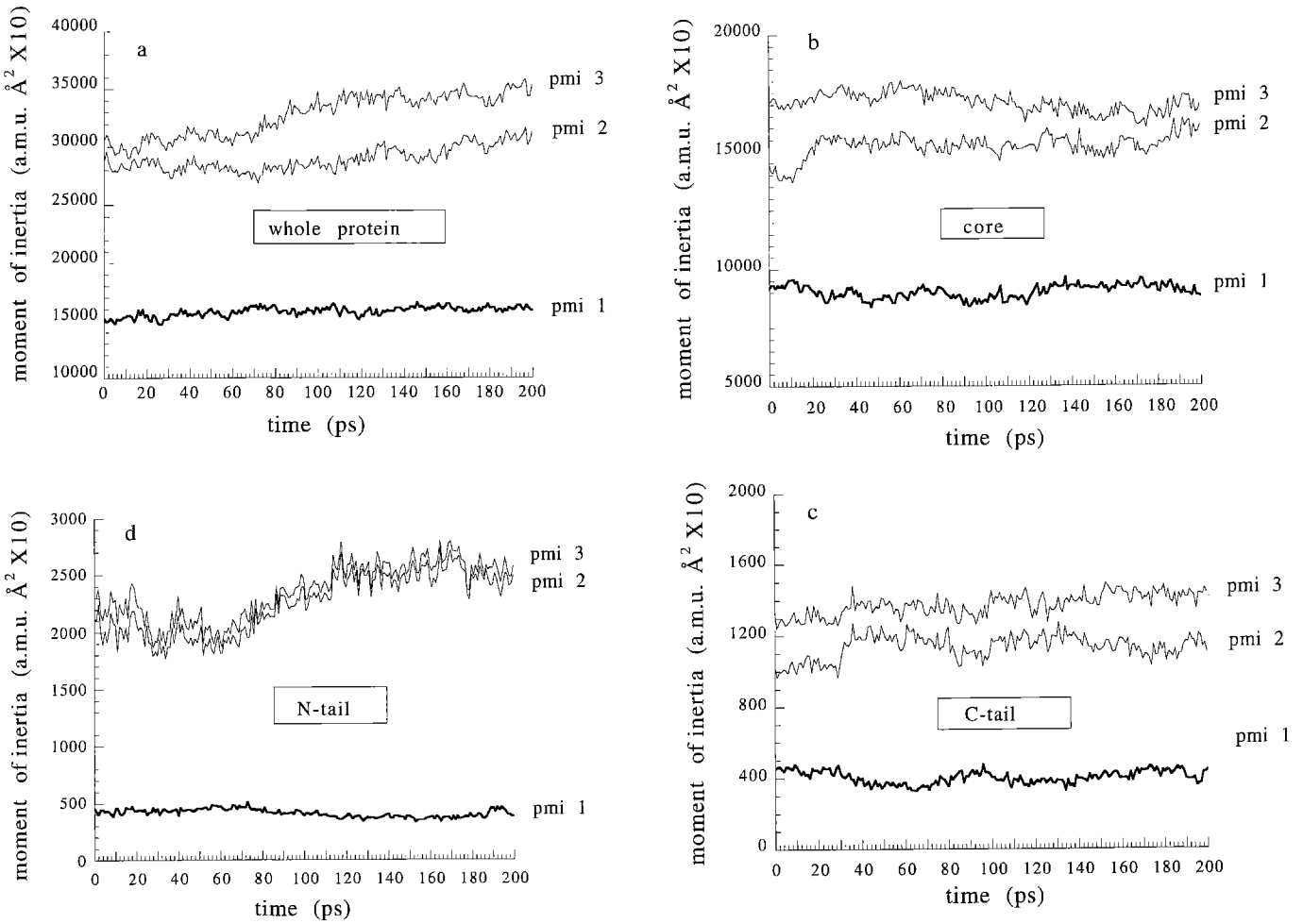

Fig. 3. Time series of the MD-derived moments of inertia of PCI in the main axes. **a** Whole PCI molecule. **b** PCI core **c** PCI N-tail. **d** PCI N-tail(pmi 1 ° in thick line, pmi 2 ° and pmi 3° in thin line. Gauged in atomic mass unit (a.m.u.) by square Ångstrom)

the apolar hydrogens is added to the carbon atom to which they are bonded), yield information on the overall shape behaviour of the protein as a function of time. For instance, the presence of one small and two large, similar pmi's is associated with a more or less cylindric distribution of matter, e.g. a helix. The changes in shape during the MD trajectory for all atoms and the three regions defined for this molecule are reported in Fig. 3.

The time evolution of the principal moments of inertia for the whole molecule show some minor changes in the overall shape. The structure is always more or less elongated. Between the time points 40 ps and 80 ps the structure is *quasi* cylindrical. The system has one invariant axis along the first principal moment of inertia. According to the present results, the core of PCI is stable. Once the time span between 0 and 30 ps is exceeded, the structure of the core persists, with small fluctuations in the Cα's. Despite these fluctuations, it preserves its shape along the 200 ps MD trajectory. In contrast, the N-tail is apparently compressed along the MD trajectory, but still keeps its elongated shape. Observation of the structure using computer graphics shows that this tail approaches the core along the trajectory. Interestingly, the C-tail preserves its shape during the trajectory, but with relatively large fluctuations.

c. Comparisons with X-ray data

Overall structural displacements. Since the structure of PCI used to seed the MD simulation is obtained from the crystallographic complex of PCI with CPA, one would expect differences in structure and fluctuation patterns for the isolated inhibitor, given that the system in water is not constrained by zinc-binding, protein-protein or crystal forces. Despite this fact, the r.m.s. for all Cα's of the MD structures (taken in steps of 2 ps), superimposed on the PCI crystallographic structure by fitting the Cα's of the cystines, is found to be approximately 2 Å, while for all atoms it is under 3 Å (not shown). The fluctuation pattern and time-dependent structure for the core appear to be fairly stable, reaching a r.m.s. value for its Cα's of 1.15 ± 0.2 Å. In constrast, the N-tail shows a much greater pattern of change, with average r.m.s. values for its Cα's of 3.9 ± 1.3 Å. The C-tail shows an intermediate behaviour, with values around 1.6 ± 0.6 Å.

r.m.s. deviations in Cα's residue. A comparison between Cα atoms of the MD average structure ⟨20–200 ps⟩ and the X-ray crystal structure (Fig. 4) shows that most of the PCI residues within the core remain in a configuration topo-

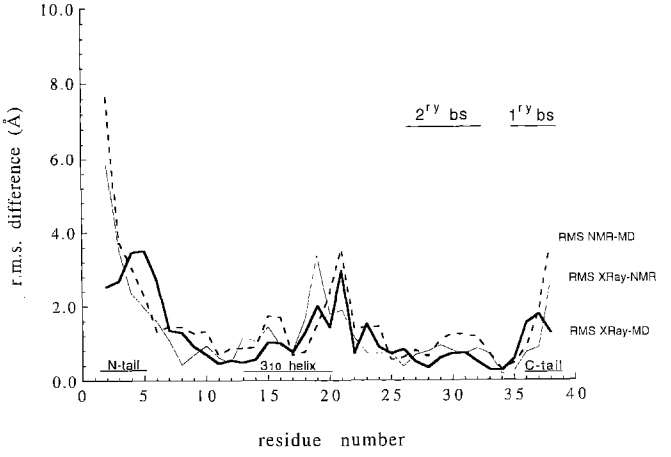

Fig. 4. Comparison between the ⟨20–200⟩ MD averaged structures of PCI and the experimental structures. Comparison between the r.m.s. difference of the Cα's of the MD average structure of PCI with respect to the crystallographic structure (*thick line*), r.m.s. difference of the NMR reproduced structure (*RDDG*) with respect to the crystallographic structure (*thin line*), and r.m.s. difference of the Cα's of the MD average structure of PCI with respect to the *RDDG* structure (*dashed line*). The main structural features are also indicated: primary binding site (1^{ry} bs), secondary binding site (2^{ry} bs), N-tail, 3_{10} helix and C-tail

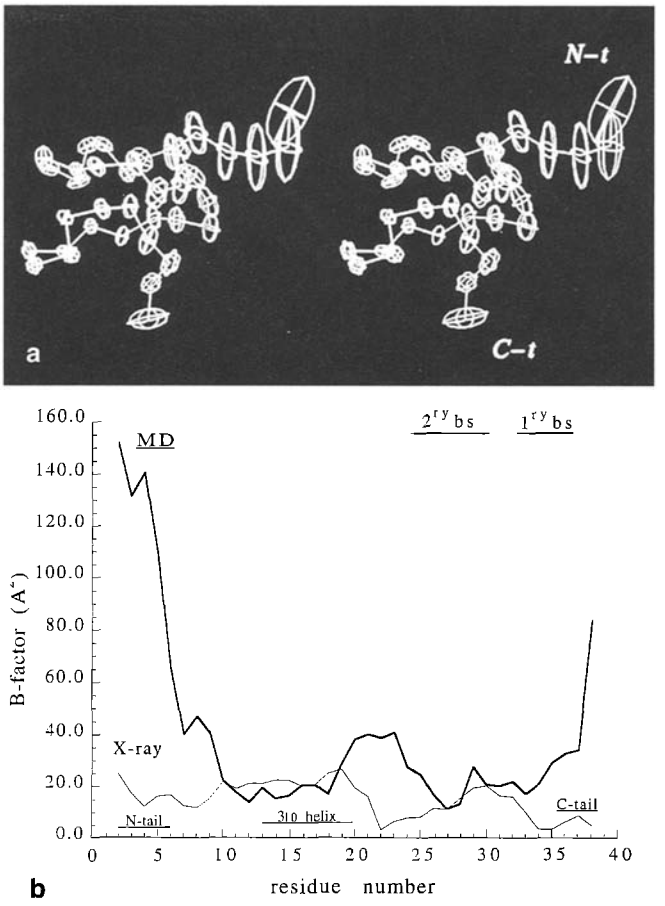

Fig. 5a, b. MD r.m.s. fluctuations and temperature B-factors of PCI. **a** Ellipsoid representation of the anisotropic r.m.s. fluctuation of the Cα atoms of PCI in the ⟨20–200 ps⟩ MD averaged structure of PCI. **b** Cα temperature B-factors of the PCI molecule. The thin line shows the crystallographic temperature B-factors of PCI in the crystal of CPA-PCI complex. The thick line shows the temperature B-factors obtained in the MD simulation with solvent. The main structural features are also indicated: primary binding site (1^{ry} bs), secondary binding site (2^{ry} bs), N-tail, 3_{10} helix and C-tail

logically similar to that found in the experimental complex, with displacement differences under 2 Å. Interestingly, the in-water simulation results are similar to the in-pseudo-vacuum (NIS) simulation results (Oliva et al. 1991 a, b). However, the drifts in the tails, and mainly in the N-terminal tail, are smaller in the model with explicit water molecules than in the NIS simulations. This result is probably due to solvent inertial effects and collisions that preclude large displacements.

Temperature B-factors. The simulation in water gives rise to a well-defined average structure in the 20–200 ps time window and atomic fluctuations along the three main axes of the fluctuation were calculated for this time average. The MD B-factors derived from the matrix of the atomic fluctuations were also calculated, using the equation B-factor $= 8\pi^2 \| \text{r.m.s.} \|^2/3$ (Fig. 5).

The values obtained are fairly comparable to those found in the PCI-CPA complex crystal structure (Rees and Lipscomb 1982). The differences correlate with the crystal contacts and CPA-PCI contacts.

The PCI core Cα's in the MD simulation show the smallest temperature factor values, comparable to X-ray values and all below 40 Å^2. Only at the border with the tails and at the loop formed by residues 20–26 (between the 3_{10} helix and the secondary binding site) are the core MD-temperature factors greater than those derived from the PCI-CPA crystal structure. The largest differences, with respect to the fluctuations in the crystal complex, are found at the N- and C-terminal tails. The large fluctuations of the N-terminus (with MD-B factors around 160 Å^2) are indicative of important disorder. On the other hand, the intermediate value of the MD-B factors for the C-tail (reaching 80 Å^2) indicates that this tail is also disordered although to a lesser degree than the N-tail.

The lower values of the X-ray derived B factors at the N- and C-tails with respect to the MD-derived factors are worthy of comment. Crystal contacts, e.g. the electrostatic interaction found for the Gln-2 of PCI with the Arg-2 of the CPA in the neighboring symmetric cell, may be sufficient to stabilize mobile parts, such as loops or tails, and this may decrease the values of these crystallographic B-factors at the N-tail. In contrast, the low values of this factor at the C-tail are better explained by the fact that the tail motion is dampened by its coordination with zinc in the PCI-CPA complex in a larger extent as the water cage does in isolated PCI.

d. Comparisons with NMR data

r.m.s. deviations per residue from the X-ray structure. Although the structure reported by Clore et al. (1987) is somewhat old, the comparisons with our MD results are still useful since both structures are independent. The

NMR structure revealed 11 converged structures, referred to as DG structures, obtained from restriction geometry calculations using the program DISGEO (Havel and Wüthrich 1985). Refinement was carried out in two stages. In the first, restrained energy optimization generated the so-called DGm structures. In the second, a restrained MD was used to obtain the final structures referred to as RDDG. In addition, the mean structures *DG*, *DGm*, *RDDG* of each class were calculated. For the sake of comparison, the data reported have been retrieved and used in conjunction with our own data.

The r.m.s. difference between the *RDDG* structure and the X-ray structure of PCI and the r.m.s. difference between the average ⟨20–200 ps⟩ of the MD aqueous simulation of PCI and the experimental structures: X-ray structure and NMR structure (*RDDG*), are reported in Fig. 4. The shapes and values shown for the two conformations (*RDDG* and ⟨20–200 ps⟩ MD average) with respect to the crystallographic structure show a remarkable similarity, with the exception of the residues at the tails and the above-mentioned loop between residues 20–26. It is worth noting the coincidence of high r.m.s. difference values at residue 19 in both MD- and NMR (RDDG) structures, although this difference is also present for the r.m.s. of the MD averaged structure with respect to the NMR structure, this shows the digression for the three structures around this residue. On the whole, the comparison between the *RDDG* structure and the average structure of the MD simulation in aqueous solution shows that the PCI structural backbone deviates from the X-ray and from the NMR structure holding the features of both.

r.m.s. differences. The atomic r.m.s. differences presented in Table 1 show that the MD average structure between 20 ps and 200 ps has a well-defined conformation. The r.m.s. differences obtained for each 10 ps from 20–200 ps (in the set of 18 MD conformations) is smaller than any of those obtained in the sets reported by Clore et al. (1987). On the other hand, the comparison between the MD set of these 18 conformations and the X-ray conformation shows a small r.m.s. of about 1.1 Å for the backbone atoms of the 2 to 38 residues and of about 1.6 Å for the heavy atoms. The r.m.s. differences in the MD set versus the X-ray conformation shows that the MD conformations deviate from the X-ray structure as do the *RDDG* reported set of conformations.

NMR inter-proton distances versus MD average structure. A more accurate structural characterization can be obtained by calculating the interproton distances of the atoms with an upper limit distance smaller than the one found in the crystal (Rees and Lipscomb 1982). These are reproduced in Table 2, and include the values found for the ⟨20–200 ps⟩ MD averaged structure in aqueous solution. The interproton distances of the structure in the crystal complex of PCI with CPA are compared with the result of a minimization using distance restraints of the crystallographic structure, and both are compared with the NMR data. Taking the difference between the column Δ2 and Δb from the X-ray RM reported by Clore et al. (1987), and reproduced in Table 2, the absolute difference in interatomic distance between the upper limit distance and our MD in water simulation is obtained.

Table 1. Atomic r.m.s. differences (Å) between the structures of PCI

	atomic r.m.s. differences (Å)			
	residues 2–39		residues 5–36	
COMPARISON	*backbone*	*all atoms*	*backbone*	*all atoms*
⟨MD⟩ vs ⟨MD⟩	1.05±0.29	1.52±0.39	0.85±0.21	1.30±0.32
⟨MD⟩ vs X-ray	1.61±0.25	2.36±0.23	1.45±0.22	2.09±0.20
⟨RDDG⟩ vs ⟨RDDG⟩[a]	2.10±0.40	2.80±0.40	1.30±0.20	2.00±0.30
⟨RDDG⟩ vs X-ray[a]	2.10±0.20	3.00±0.30	1.50±0.20	2.40±0.30

RDDG = molecular dynamics of the 11 DGm conformations using the NOE distance restraints
MD = molecular dynamics simulation of PCI in aqueous solution
⟨A⟩ = indicates the set of A structures
[a] As reported by Clore et al. (1987) [6]

Agreement with short-range inter-residue distances is remarkably good. With regard to NH(i)–NH(i+1), the greatest difference (1.5 Å) is found for G20-A21, while all other pairs reported have differences under 1 Å. In these case of NH(i)-NH(i+2), the difference amounts to 0.3 Å. The agreement for the CαH(i)-NH(i+1) case is also fairly good. Not surprisingly, the long-range inter-residue distances present larger differences than the short-range ones. The greatest difference between the upper limit and the interatomic distance in the ⟨20–200 ps⟩ MD averaged structure (5 Å) concerns the proton pair F23 (CδH)-D5(CβH).

e. Shape changes from X-ray to solution structure

Radius of gyration of PCI. If the PCI conformation changes its shape from the crystallographic structure to the structure in solution during the simulation, then this should be clearly detectable from the radius of gyration of the whole molecule calculated for all the atoms (defined as the r.m.s. of the radii of gyration in the main axes of the normalized inertial tensor). This can also be calculated using the equation:

$$RG = (\sum r_i^2 m_i / \sum m_i)^{1/2}$$

Figure 6 shows the time series of the radius of gyration. The upper limit value for the radius of gyration is taken from that obtained from the average of the 11 DG NMR conformations that were obtained experimentally without further refinements (Clore et al. 1987). This is considered as the upper limit value because distance geometry calculations with the program DISGEO (Havel and Wüthrich 1985) performed on tails, tend to elongate the structures (Billeter M., personal communication). The small NIS values for the radius of gyration are due to the contracting effect of simulations without water molecules (Oliva et al. 1991 a, b), so that these can be considered to be the lower limit values. The accuracy of the simulation with explicit water as compared with experimental NMR data is most

Table 2. Difference between the experimentally derived upper limit interproton distance restraints in solution and the interproton distances derived from the X-ray structure of PCI[a]

Atoms related	type	int[b]	X-ray[b]	Δa[b]	X-ray-RM[b]	Δ[b]	MD	Δ1	Δ2
	Short range (|i-j|≤5)				Inter-residue				
NH(i)-NH(i+1)									
C18, S19	b	s	3.5	−0.8	2.8	−0.1	3.3	−0.2	+0.5
S19, G20	a	s	4.0	−1.3	3.0	−0.3	2.7	−1.3	−0.3
G20, A21	c	s	3.4	−0.7	2.2	0.0	3.7	+0.3	+1.5
W22, F23	c	s	3.6	−0.9	2.8	−0.1	3.9	+0.3	+1.1
R32, T33	b	s	3.8	−1.1	2.9	−0.2	3.4	−0.4	+0.5
NH(i)-NH(i+2)									
W22, C24	a	w	5.9	−0.9	4.4	0.0	4.7	−1.2	+0.3
Cα H(i)-NH(i+1)									
A21, W22	a	s	3.4	−0.7	3.1	−0.4	2.6	−0.8	−0.5
Cα H(i)-Cβ H(i+3)									
D16, S19	a	w	7.6	−2.6	5.3	−0.3	4.8	−2.8	−0.5
C18, A21	b	w	6.3	−1.3	5.1	−0.1	5.2	−1.1	+0.1
Cβ H(i)-NH(i+1)									
C12, K13	b	s	3.8	−1.1	3.0	−0.3	3.5	−0.2	+0.5
K13, T14	a	s	3.4	−0.7	2.7	0.0	2.9	−0.5	+0.2
A21, W22	a	m	4.2	−0.7	3.3	0.0	3.5	−0.7	+0.2
A31, R32	b	s	4.4	−1.2	3.4	−0.2	3.9	−0.5	+0.5
T33, C34	a	s	4.6	−1.9	2.9	−0.2	3.2	−1.4	+0.3
others									
A21(Cβ H)–Q25(Cα H)	c	s	4.3	−1.6	3.0	−0.3	4.6	+0.3	+1.5
I7(Cδ H–C8(NH)	a	w	6.3	+0.8	5.5	0.0	5.5	−0.8	0.0
I7(Cδ H)–K10(Cδ H)	a	s	5.4	−2.2	3.8	0.0	4.1	−0.1	+0.3
T33(Cγ H)–C34(NH)	c	s	3.9	−0.7	2.3	0.0	4.0	0.1	+1.7
	Long range (|i-j|>5)				Inter-residue				
C8(Cα H)–A21(Cβ H)	c	m	4.9	−1.1	4.2	−0.4	5.5	+0.6	+1.3
N9(Cα H)–T33(Cα H)	d	m	4.1	−0.8	3.9	−0.6	5.8	+1.7	+1.9
N9(Cα H)–T33(Cβ H)	c	s	4.2	−1.5	3.3	−0.6	5.1	+0.9	+1.8
P11(Cα H)–T33(Cγ H)	b	s	5.6	−2.4	2.9	0.0	4.1	−1.4	+1.2
C12(NH)–T33(Cγ H)	a	m	6.3	−2.5	4.1	−0.3	4.4	−1.8	+0.3
A21(NH)–17(Cγ H)	b	m	5.1	−1.3	3.8	0.0	4.9	−0.1	+0.1
A21(NH)–17(Cδ H)	d	m	4.7	−0.9	4.1	−0.3	6.9	+2.2	+2.8
A21(Cβ H)–17(Cγ H)	b	s	4.6	−0.9	2.9	0.0	3.7	−0.8	+0.8
F23(Cδ H)–D5(Cβ H)	d	s	3.7	−1.0	3.1	−0.4	8.1	+4.4	+5.0
F23(Cδ H)–P36(Cγ H)	c	m	4.4	−1.1	3.5	−0.2	4.6	+0.2	+1.2
W28(NH)–C34(Cα H)	d	m	4.1	−0.8	3.4	−0.1	6.0	+2.0	+2.7
W28(Cδ1H)–G35(NH)	d	m	4.0	−0.7	3.3	0.0	7.3	+3.3	+4.1
T33(Cα H)–17(Cδ H)	c	w	11.2	−5.7	7.0	−1.5	11.0	−1.7	+4.0
C34(NH)–N9(Cα H)	c	m	4.3	−1.0	2.9	0.0	4.4	+0.1	+1.5
C34(NH)–P11(Cα H)	c	m	4.0	−0.7	3.7	−0.4	4.4	+0.4	+0.7

Δa: difference between the NOE upper limit distance and the X-ray structure
Δb: difference between the NOE upper limit distance and the X-ray optimized structure with distance restraints
Δ1: difference between the MD averaged structure and the X-ray structure
Δ2: difference between the MD averaged structure and the X-ray optimized structure with distance restraints
[a] The NOE intensities (int†) are classified into strong (s), medium (m), and weak (w) and are taken to correspond to distance ranges of 1.8–2.7, 1.8–3.3, and 1.8–5.0 Å, respectively. An additional 0.5 Å per methyl group is added to the upper distance limit to account for the higher intensity of methyl resonances. The types of interproton distances are classified in four types, (a, b, c and d) depending on the MD simulation (see the text)
X-ray† is the X-ray structure of the PCI in the complex with carboxypeptidase [5]
[b] As reported by Clore et al. (1987)

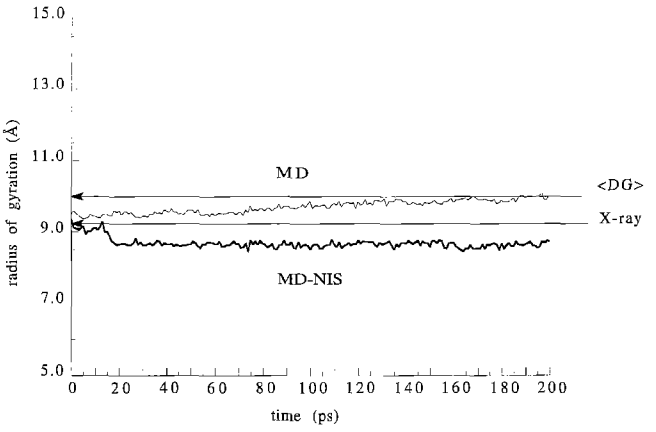

Fig. 6. Radius of gyration of PCI. Radius of gyration values along the trajectory of the whole PCI molecule in the MD simulation, in aqueous solution and in the NIS model MD simulation. Comparison with the experimental results

satisfactory in the sense that the MD-fluctuations are bounded by the two experimentally obtained gyration radii ⟨RDDG⟩ and ⟨DG⟩. X-ray values also lie between these latter two values.

Time series for selected interproton distances. It is also of interest to analyze the interproton distances in relation to the polar hydrogen of the nitrogen in the peptide bonds. Thirty-three interproton distances were analyzed and depicted as time series. The results are classified in four types:

a) the interproton distance that begins with the value of the crystal structure and moves to within the NOE upper limit distance;
b) the interatomic distance that varies between the values corresponding to the crystal structure and the NOE upper limit values;
c) the interatomic distance that remains close to the original X-ray values;
d) the interatomic distance that increases and remains greater than in the crystal structure.

Figure 7 shows four of the most illustrative cases. Among thirty-three interproton distances, ten cases of (a), eight cases of (b), ten cases of (c), and five interproton-distances of (d) were found (cf. Table 2). Thus, 85% of the interproton distances (cases "a+b+c") evolve towards an end point that is in between the crystal value and the upper limit NOE distance, and 30% of the distances (case "a") approach the latter. Why 15% of the distances (case "d") evolve away from the X-ray and upper limit values is not explained. Most of the latter involve interactions between side-chains with a main-chain hydrogen atom or two side-chain hydrogen atoms. What is clear is that the crystal also contains a large amount of water which may explain cases (c).

Cases (a) and (b) show a conformational reshaping during the 200 ps. Such dynamics behaviour also suggests that the MD simulation method is accurate enough to transform the crystal conformational structure bringing it closer to the structure found in aqueous solution.

f. Hydrogen bonds

Protein-protein hydrogen bonds in PCI. Comparisons of PCI NMR and MD data showed that the MD-simulated structure of PCI most likely represents the structure that it may have in water. To complete these analyses it is advisable to compare hydrogen bonding patterns. Table 3 shows a comparison of the H-bonding pattern found in the X-ray structure and in the MD simulation. A change in the H-bond pattern of these two structures is apparent. From this Table it follows that some hydrogen bonds are present in the complex, but not in the MD model. Others still appear, but at a reduced occurrence rate; yet others exchange their interaction partners locally without any large structural change. This is partly because in the water model the solvent molecules compete with protein groups for hydrogen bonding. It is also conceivable that the release of PCI from CPA allows for some conformational changes; such changes would facilitate alterations in the hydrogen bonding pattern.

The higher percentage of occurrence of hydrogen bonds appears to be functionally and structurally related to the surrounding of the secondary binding site and of the C-terminal tail region, from residues 26 to 35. In this respect, Gly-35 may play an important role since it appears to constitute the connection residue between the C-terminal tail and the core of PCI by establishing two hydrogens bonds with Ala-26. This is corroborated by the observation that one of the best conserved hydrogen bonds is Gly35NH-Ala26CO. This fact may be related to the inhibitory function. Such interactions are probably responsible for keeping the original position of the C-terminal tail.

Interestingly, the H-bonds corresponding to the 3_{10}-helix appear in the simulation but they are not properly described in the X-ray structure (they are not within accepted structural boundaries): 17D(N)-14T(O) and 18C(N)-15H(O) (see Table 3). This region is also one where the old set of potential function parameters used in GROMOS fails to properly represent the H-bond interactions as compared to the new set of parameters (see Discussion).

Table 3. Hydrogen bonds pattern in PCI

Donor atom	Acceptor atom	X-RAY occurrence	MD occurrence (%)
8 CYS N	5 ASP O	+	–
9 ASN N	34 CYS O	+	55
10 LYS N	7 ILE O	+	64
12 CYS N	32 ARG O	+	89
17 ASP N	14 THR O	–	72
18 CYS N	15 HIS O	–	63
23 PHE N	21 ALA O	+	44
26 ALA N	35 GLY O	+	–
28 TRP N	31 ALA O	–	55
30 SER N	28 TRP O	+	–
33 THR N	31 ALA O	–	80
34 CYS N	10 LYS O	+	94
35 GLY N	26 ALA O	+	93

+ indicates the appearance in the crystallographic structure. – Indicates that either the hydrogen bond does not appear in the crystallographic structure or that the occurrence in the MD simulation is less than the 50%

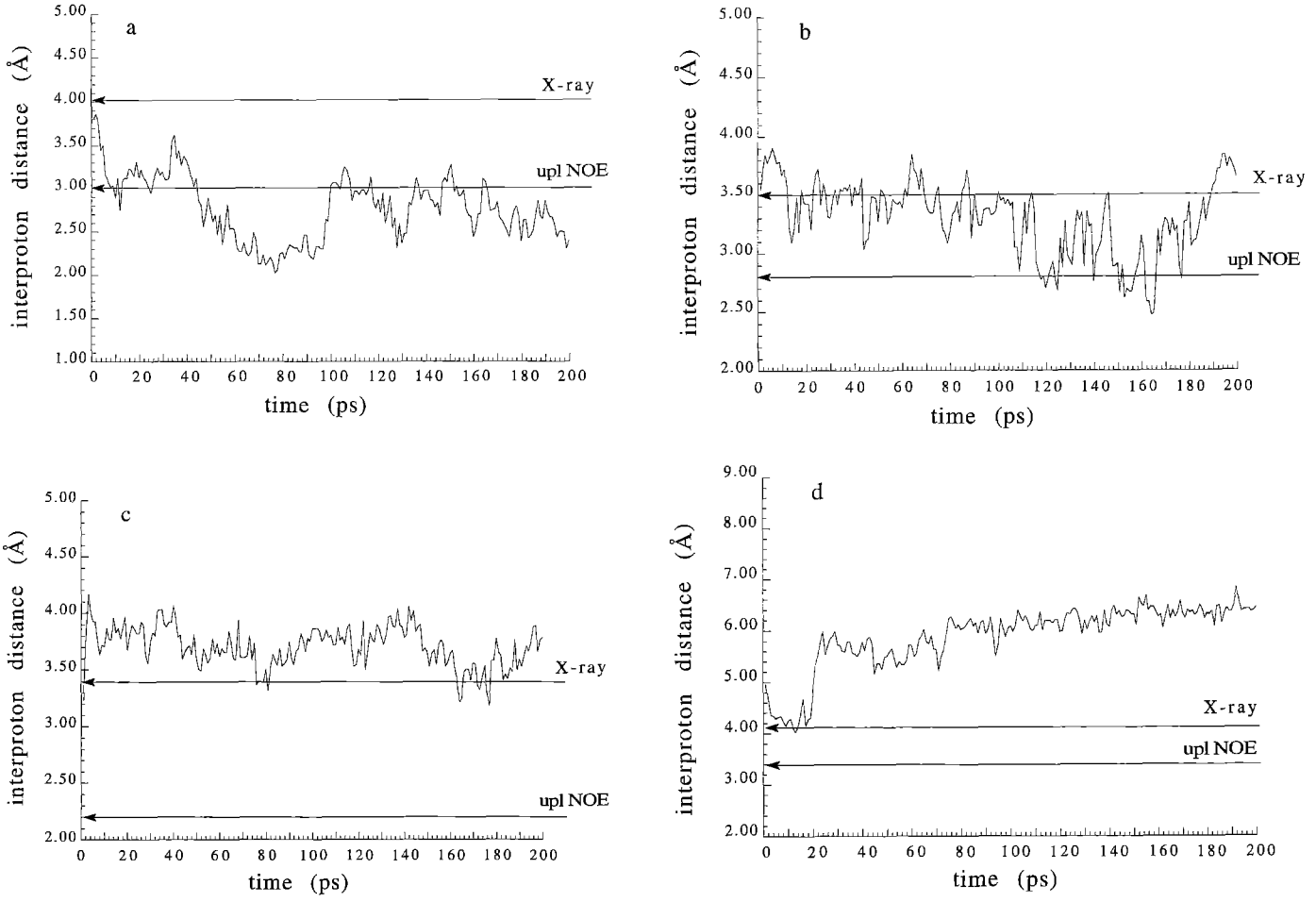

Fig. 7a–d. Time series of the interproton distances in PCI along MD simulation. Distances measured between the atoms: **a** 19 S(NH)–20 G(NH); **b** 18 C(NH)–19 S(NH); **c** 20 G(NH)–21 A (NH); and **d** 28 W(NH)–34 C(CαH). These are case examples of the interproton distances (**a**), (**b**), (**c**) and (**d**) mentioned in the text

4. Discussion

The GROMOS MD-simulations of potato carboxypeptidase inhibitor protein in water, including counterions, provide a fairly complete picture of the shape and conformational fluctuations experienced by the protein in a medium-range trajectory span (20–200 ps). The deviation of the MD-derived structural values for PCI alone from those taken from the X-ray structure of the PCI-CPA complex is small for the core region and for the C-tail and high for the N-tail. The fact that this relatively short trajectory yields an average structure and fluctuation patterns that resemble the behaviour of PCI in aqueous solution as measured by 2D-NMR suggests that the simulation is reasonably meaningful.

The structure of the PCI in water is well-defined by simulation (Table 1). The atomic r.m.s. difference of the backbone of 18 MD conformations (one each 10 ps from 20 ps to 200 ps) is about 1.05 Å, a lower value than the result obtained for the 11 conformations reported by NMR by Clore et al. (1987). The comparison of the X-ray structure with the ⟨20 ps–200 ps⟩ MD averaged structure gives similar differences for each residue to those found with the NMR (*RDDG*) structure (Clore et al. 1987) (Fig. 4). The atomic r.m.s. difference for the set of the 18 MD conformations is also similar to the set of 11 NMR conformations (Table 1). In addition, the radius of gyration (Fig. 6) and some of the interproton distances (Fig. 7 and Table 2) show that the PCI modeled by MD is in-between the crystallographic structure and the upper limits defined by NMR data.

The core of PCI is largely preserved at both the MD model and the NMR structure with respect to the X-ray structure (Fig. 4). The coincidence of high r.m.s. difference values at residue 19 for both solution structures when compared with the X-ray structure is probably due to the shift in two residues position for the 3_{10} helix when reported by NMR or by X-ray. From this point of view, and given the preservation of correct stereochemistry and H-bonding (see below), it seems that the MD-simulation is more able to define this helical structure in solution than X-ray. It is difficult to conclude whether these differences between MD- and X-ray structures at the helix are genuine or are derived from the structural constrictions to which PCI is submitted in the latter. On the other hand, comparisons between the MD model and the NMR structure at the tails confirm their mobile character, particularly the N-tail.

The hydrogen bonding pattern of the MD model shows the loss of some bonds that appear in the crystallographic structure (Table 2), even at the core. This can be attributed

to the interaction with water molecules, an important effector in a protein such as PCI with little regular secondary structure, and is in agreement with the large hydrogen exchange observed by NMR (Clore et al. 1987). In contrast, it is found that the stereochemistry of the 3_{10} helix is better defined in the MD- than in the X-ray structure. Moreover, two important hydrogen bonds for the definition of the 3_{10} helix (T14-D17 and H15-C18) appear in the MD-simulation but not in the X-ray structure. It is also remarkable that amongst the best preserved hydrogen bonds are those established between carbonyl and imino groups of Ala26-Gly35 and Lys10-Cys34. These internal hydrogen bonds may be related to the orientation of the C-tail to interact with the CPA.

The MD-simulation of the C-tail behaviour is interesting. Although this tail largely fluctuates, it does so around a relatively well-defined average conformation, which has the correct orientation allowing it to form a complex with the carboxypeptidase molecule. The latter is one of the significant characteristics derived from our work, an issue that has been further explored by the combined use of site-directed mutagenesis and MD simulation in our laboratory (Molina et al. 1994, Tapia et al. 1992). That is, the MD studies indicate that the C-tail fo PCI is much less mobile than can be expected by being outside the central globular core. Besides the reported hydrogen bonds between the residues of the C-tail and the residues of the PCI core in the crystal structure (van Gunsteren and Mark 1992), the overall analyses of the MD results point to the occurrence of the previously mentioned stable hydrogen bonds between Ala26 and Gly35 (Table 3). All these bonds probably have a significant effect on the inhibitory capacity of PCI by restricting the fluctuation pattern of the C-tail and facilitating a definite orientation to bind CPA.

Previous MD results obtained with a pseudo in vacuum (NIS) model are mostly validated by the present work (Oliva et al. 1991 a). The use of the NIS model as a rapid means of detecting mutation effects on the fluctuations and structure yielded, in that case, significant results (Tapia et al. 1992). However, the present work has also shown that a representation of the surrounding medium is required when a detailed picture of the fluctuations of PCI and an accurate representation of its behaviour in solution is sought. The NIS model must be supplemented with solvent collision effects using for instance information on solvent accessibility surface.

A more realistic representation of the PCI MD-structure and dynamics has also been obtained by the use of the recently derived new set of GROMOS parameters (see Methods). An overall comparison of the results obtained with the old and new set of parameters indicates that the latter gives structures closer to the experimental ones (as show by smaller r.m.s. differences and B factors). Besides, the new parameters identify H-bonds with a higher occurrence probability. Differences in detail will be discussed elsewhere.

In conclusion, the MD-simulation of PCI in explicit water provides a view of the structural features and behaviour of this molecule complementary to those obtained by X-ray and by NMR. The MD study is particularly useful to show that the active conformation of the C-tail is not the result of the interactions with the receptor molecule. Also to show that the core fold is very stable in water, despite its low content in regular secondary structures. Extension of similar MD work on proteins which have recently been found to share a similar fold with the PCI core – such as trypsin "squash" inhibitors, erabutoxin, wheat germ agglutinin, and neurophysin (Hom and Sander 1993) – may provide an explanation for the occurrence and function of such a fold.

Acknowledgements. The authors wish to thank Dr. M. Billeter (ETH, Zürich) for the critical revision of this manuscript. X. Daura is a fellowship recipient from the Centre de Supercomputació de Catalunya (CESCA) and the Programa de Química Fina (CIRIT), and B. Oliva and X. Daura have received fellowships from the CSIC-Swedish Institute Program and Ministerio de Asuntos Exteriores (Spain). This work has been supported by grants BIO91-0477, BIO92-0458 and IN90-0009 from the CICYT (Ministerio de Educación, Spain). The authors acknowledge support from Fundación Francisca de Roviralta and from the CESCA. O. Tapia acknowledges grants from the Swedish Research Council (NFR).

References

Åqvist J, van Gunsteren WF, Leijonmarck M, Tapia O (1985) A Molecular Dynamics Study of the Terminal Fragment of the L7/L12 Ribosomal Protein. Secondary Structure Motion in a 150 Picosecond Trajectory. J Mol Biol 183:461–477

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular Dynamics with Coupling to an External Bath. J Chem Phys 81:3684–3690

Casari G, Sippl MJ (1992) Structure Derived Hydrophobic Potential. Hydrophobic Potential Derived from X-Ray Structures of Globular Proteins is able to Identify Native Folds. J Mol Biol 224:725–723

Chiche L, Gaboriaud C, Heitz A, Mornon JP, Castro B, Kollman PA (1989) Use of Restrained Molecular Dynamics in Water to Determine Three-Dimensional Protein Structure, Prediction of the Three Dimensional structure of Ecballium elaterium Trypsin Inhibitor II. Proteins 6:405–417

Clore GM, Gronenborn AM, Nilges M, Ryan CA (1987) Three-Dimensional Structure of Potato Carboxypeptidase Inhibitor in Solution. A Study Using Nuclear Magnetic Resonance, Distance Geometry and Restrained Molecular Dynamics. Biochemistry 26:8012–8023

Gilson MK, Honig B (1988) Calculation of the Total electrostatic Energy of a Macromolecular system: Solvation Energies, Binding Energies and Conformational Analysis. Proteins 4:7–18

Hass GM, Ryan CA (1981) Carboxypeptidase Inhibitor from Potatoes: Methods Enzymol 80:778–791

Havel TF, Wüthrich K (1985) An Evaluation of the Combined Use of Nuclear Magnetic Resonance and Distance Geometry for the Determination of Protein Conformation in Solution. J Mol Biol 182:281–294

Holm L, Sander C (1993) Protein structure comparison by alignement of distance matrices. J Mol Biol 233:123–138

Karplus M, Petsko GA (1990) Molecular Dynamics Simulations in Biology. Nature 347:631–639

Kraulis PJ (1991) Molscript: A program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 24: 946–950

Molina MA, Marino C, Oliva B, Avilés FX, Querol E (1994) C-tail valine is a key residue for stabilization of complex between potato inhibitor and carboxypeptidase A. J Biol Chem 269: 21467–21472

Oliva B, Wästlund M, Nilsson O, Cardenas R, Querol E, Avilés FX, Tapia O (1991 a) On the Stability and Fluctuations of the Potato Carboxy-peptidase Inhibitor Fold, A Molecular Dynamics Study. Biochem Biophys Res Commun 176:616–621

Oliva B, Nilsson O, Wästlund M, Cardenas R, Querol E, Avilés FX, Tapia O (1991 b) Aspects of Model Building Applied to Carboxypeptidase Inhibitor Protein, A Molecular Dynamics Study of a Putative Pro 36/Gly Mutant. Biochem Biophys Res Commun 176:627-632

Rees DC, Lipscomb WN (1982) Refined Crystal Structure of the Potato Inhibitor Complex of Carboxypeptidase A at 2,5 Å of Resolution. J Mol Biol 160:475–498

Roe SM, Teeter MM (1992) Patterns for Prediction of Hydration around Polar Residues in Proteins. J Mol Biol 229:419–427

Ryan CA (1989) Proteinase Inhibitor Gene Families, Strategies for Transformation to Improve Plant Defenses Against Herbivores BioEssays 10:20–25

Straatsma TP, McCammon JA (1990) Molecular Dynamics Simulations with Interaction Potentials Including Polarization. Development of a noniterative method and application to water. Mol Sim 5:181–192

Stouten PFN, Frömmel C, Nakamura H, Sander C (1993) An effective solvation term based on atomic occupancies for use in protein simulations. Mol Sim 10:97–129

Tapia O, Oliva B, Nilsson O, Querol E, Avilés FX (1991) Molecular Dynamics as a Tool to Help Design Mutants. A Comparative Study of MD Simulated Potato Carboxypeptidase Inhibitor and Pro 36 Gly Mutant with the PCI-CPA X-Ray Structure. Mol Eng 1:249–266

van Gunsteren WF, Mark AE (1992) On the interpretation of biochemical data by molecular dynamics computer simulation. Eur J Biochem 204:947–961

van Gunsteren WF, Berendsen HJC (1990) Computer Simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry. Ang Chem 29:992–1023

van Gunsteren WF, Berendsen HJC (1987) Groningen Molecular Simulation (GROMOS) Library Manual. Nijenborgh 16, Groningen, The Netherlands (eds). BIOMOS B.V.

van Buuren AR, Marrink SJ, Berendsen HJC (1993) A molecular dynamics study of the decane/water interface. J Phys Chem 97:9206–9212

van Gunsteren WF, Berendsen HJC (1977) Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 34(5):1311–1327

Vila J, Willimas RL, Vásquez M, Sheraga HA (1991) Empirical solvation Models can be Used to Differentiate Native from Near-Native Conformations of Bovine Pancreatic Trypsin Inhibitor. Proteins 10:199–218

von Freyberg B, Braun W (1993) Minimization of Empirical Energy Functions in Proteins Including Hydrophobic Surface Area Effects. J Comp Chem 14:510–521

Wesson L, Eisenberg D (1992) Atomic Solvation Parameters Applied to Molecular Dynamics of Proteins in Solution. Protein Sci 1:227–335

Yun-Yu S, Lu W, van Gunsteren WF (1988) On the approximation of solvent effects on the conformation and dynamics of cyclosporin A by stochastic dynamics simulation techniques. Mol Sim 1:369–383