ELSEVIER

Biophysical Chemistry 69 (1997) 185–198

Biophysical Chemistry

Long-term molecular dynamics simulation of copper plastocyanin in water

A. Ciocchetti [a,b], A.R. Bizzarri [a,b], S. Cannistraro [a,b,*]

[a] *Unità INFM, Dipartimento di Fisica dell'Università, Perugia I-06100, Italy*
[b] *Dipartimento di Scienze Ambientali, Università della Tuscia, Viterbo I-01100, Italy*

Received 10 February 1997; revised 12 June 1997; accepted 12 June 1997

Abstract

A long molecular dynamics simulation (1.1 ns) of fully hydrated plastocyanin has been performed and analysed to relate protein dynamics to structural elements and functional properties. The solvated structure is described in detail by the analysis of H-bond network. During all the simulation, the crystal H-bond network is maintained in the β-sheet regions, while several H-bonds are broken or formed on the external surface of the protein. To evaluate whether such changes could be due to conformational rearrangements or to solvent competition, we have examined the average number of H-bonds between protein atoms and water molecules, and the root mean square deviations from crystal structure as a function of protein residues. Protein mobility and flexibility have been examined by positional and dihedral angle rms fluctuations. Finally, cross-correlation maps have revealed the existence of correlated motions among residues connected by hydrogen bonds.

Keywords: Cross-correlations; Hydrogen bonding; Molecular dynamics; Plastocyanin

1. Introduction

A full understanding of the wide range of functions exerted by proteins in living organism requires a complete description of their structure and internal dynamics.

Several experimental and theoretical investigations have been performed to obtain information in this direction. Due to spatial and temporal features that characterize each technique, a combination of different approaches is generally required to achieve effective results. An accurate description, with atomic resolution, of both the protein and the surrounding water is accessible by molecular dynamics (MD) simulations. A description of protein and solvent dynamics can be obtained, at present, on time scales ranging from femtoseconds to some nanoseconds. Data obtained from MD simulations have been used for comparison with experimental results, revealing good reliability and opening interesting perspectives for complementing neutron, NMR and X-ray spectroscopy data [1–3].

Recently, our attention has been drawn by plastocyanin (Pc), a blue copper protein that is involved in the electron-transport mechanism from Cyt *f* to reaction center chlorophyll (P700) in the chloroplasts of higher plants, green algae and cyanobacteria. Its

* Corresponding author. E-mail: cannistraro@perugia.infn.it

0301-4622/97/$17.00
PII S0301-4622(97)00089-6

99 amino acids are arranged to form two β-sheets that surround a well-defined hydrophobic core, where a single copper atom is liganded by the side chains of two histidines, a cysteine and a methionine.

Pc molecule (Fig. 1) is an interesting model for our research for many reasons. Different experimental approaches can be used to elucidate the structural and the dynamical properties of Pc; in particular, the paramagnetic behaviour of Pc, investigated by EPR spectroscopy on which our interest is also committed [5–7], offers the opportunity to analyze in detail the electronic properties of the copper ion, the conformational substates and the spectral density. Moreover,

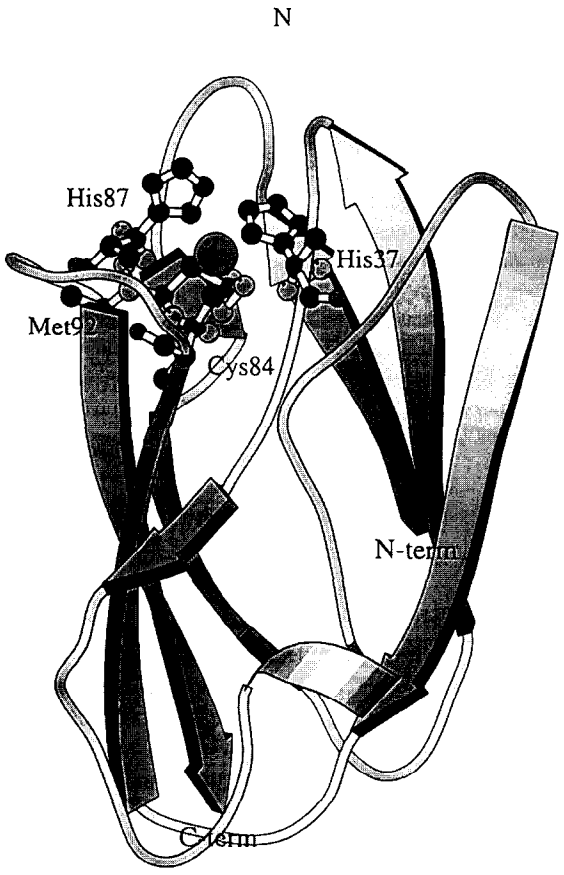

Fig. 1. Model of Pc crystal structure, showing 8 β-strands and a short α-helix. The copper atom is shown at the top, surrounded by the four ligands. The northern (N) and southern (S) part of the protein to which we refer in the text, are indicated; eastern part faces the reader, western part is on the opposite side. This drawing was generated using Molscript [4].

Pc (10541 dalton) is small enough to be explored with the methods of MD even in conditions of full hydration and over long times; it is one of the best characterized protein in the electron transfer chain [8–10], which means that many experimental data are available on both its structure and function; the globular and compact β-structure is maintained over time, allowing a comparison to more flexible proteins.

We have discussed MD simulations of Pc as a function of hydration in some previous articles [11,12], where we described the dependence of structural and dynamical properties of Pc on the hydration levels. The results reported therein are in agreement with the idea that water acts as a 'mobility catalyzer' for biomolecules [13], affecting protein motion through modification of electrostatic and van der Waals interactions. On the other hand, the protein macromolecule is able to strongly affect the structural and the dynamical properties of the hydration water. An anomalous behaviour, particularly evident at low hydration, was put into evidence by MD for water diffusion close to the protein surface [14,15].

In this work we study the dynamical behaviour of Pc in water in relation to both the protein structure and function as described by experimental works reported in literature. In particular, we have focused our attention on the intraprotein H-bond network to understand the relative importance of secondary structure elements and the effect of solvent on them. We have also examined data concerning the residues located in the copper site, in the acidic and hydrophobic patches [10]. The copper site, responsible for the electron-transfer process of the protein, has a distorted tetrahedral geometry that causes the high value of the redox potential and the fine tuning of electron transfer reactions. The acidic and hydrophobic patches are located on the external surface of Pc and surround the remote and adjacent sites for electron transferring from and to the copper atom, respectively [10,16]: such regions are conserved in the eukaryotic organisms and are supposed to have a role in the binding of the Pc reaction partners.

To perform accurate analyses, we run a long MD simulation by using the GROMOS force field: the examined time interval (1.1 ns) allowed us to separate a well-defined initial equilibration time from the following sampling time, which is long enough to

provide a good statistic for the structural and dynamical variables taken into consideration. Moreover, we used some new parameters for improving the computation of the potential function: in particular, we modified the cutoff parameters as suggested by Steinbach and Brooks [17] and also modified the GROMOS force field, as suggested by Mark et al. [18]. The use of Mark modification is also recommended by Daura et al. [19] who performed several MD simulations using different parameters sets, and observed that this change is able to best preserve the protein native structure.

2. Computational methods

Initial coordinates of Pc including 110 crystallization waters were taken from the X-ray Pc crystal structure [20] at 0.133 nm resolution (1PLC entry of Brookhaven Protein Data Bank). The MD trajectories of Pc were generated by an integration step of 0.002 ps, using the GROMOS87 program package [21] with the GROMOS force field modified according to Ref. [18]. With the exception of copper ligands (His37, Cys84, His87, Met92), all the ionizable residues are assumed to be in their fully charged state, according to the value of pH = 6 as reported in the crystallographic work [20]. The copper ion is coordinated to the side chain nitrogens of His37 and His87 and to the side chain sulfurs of Cys84 and Met92. Since the GROMOS force field does not include parameters for amino acids liganded to metal ions, some changes of the potential should be done. In one of our previous simulations, the copper–ligand interaction was described by a nonbonded approach based on an electrostatic model [11]. However, a deviation from the X-ray data for the copper–ligand arrangements was detected during the MD simulation; such an effect being particularly significant for the Cu–S–Met 92 distance. Recently, absorption spectroscopy investigations showed a general covalent nature for the copper site [10]; a particular strong covalent nature was put into evidence for the Cu–S–Cys84. On such a ground, a covalent bond between the copper and each ligand was introduced to preserve the X-ray structure. The charge of Cys84 was set to −0.5 e, while His37, His87 and Met92 are considered neutral; the copper atom was given a charge of 0.6 e. The total protein charge is then −7.9 e. It should be, however, considered that the bonded description might introduce some harmonic terms in the bond-stretching potential [22].

The protein molecule was centered in a truncated octahedron obtained from a cube of edge 6.20257 nm filled with bulk SPC/E waters [23]. A minimum distance of 1.06 nm between the solute atoms and the box edges was maintained. Any water molecule placed to a distance from any protein atom smaller than 0.23 nm was removed to give a final fully hydrated system containing 3514 water molecules corresponding to 6.001 g of water per grams of protein. To avoid edge effects and to better describe the condition of full hydration, periodic boundary conditions were applied. In agreement with the work by Steinbach and Brooks [17], the cutoff radii of 0.8 nm for the nonbonded interactions and of 1.4 nm for the long-range charged interactions were used. During 105 steps of energy minimization with the steepest descent method, a harmonic position restraining a force constant equal to 9000 kJ mol^{-1} nm^{-2} was used to minimize the root mean square deviations (RMSD) from the X-ray structure. Longer energy minimization, also coupled with the conjugate gradient method, have been performed, without a real improvement of the minimized structure.

Initial atomic velocities were assigned from a Maxwellian distribution corresponding to 250 K [24]. Any residual translational and rotational motion of the center of mass was removed from the initial velocities. The temperature was kept fixed at 250 K for the first 6 ps, then it was increased of 5 K every 4 ps to reach the value of 300 K which was maintained throughout the following simulation. The temperatures of the protein and of the solvent were separately coupled to an external bath with relaxation times of 0.1 ps. The pressure was kept constant by a coupling to a pressure bath at 1.013 bar with a relaxation time of 0.5 ps. A decreasing positional restraining force, with a constant going from 9000 kJ mol^{-1} nm^{-2} to 50 kJ mol^{-1} nm^{-2} was also applied during the first 40 ps.

The MD simulation consisted of 100 ps for equilibration, followed by 1000 ps of data collection. Configurations of all trajectories and energy were saved every 0.1 ps. The neighbour pair list was updated every 10 steps. The Shake constraint algo-

rithm [25] was used throughout the simulation to fix the internal geometry of water molecules, and to keep bond lengths of protein rigorously fixed at their equilibrium values.

The structural and dynamical properties of Pc have been computed on the basis of two different averaging time: 100–600 ps and 100–1100 ps. Since the results show very few qualitative differences, only data obtained from the nanosecond trajectory are reported.

In computing the RMSD, the gyration radius and the rms-fluctuations (RMSF), the overall translational and rotational protein diffusions have been removed by superimposing the backbone of each configuration onto the backbone of the Pc structure using a mass-weighted least-squares fitting algorithm [26].

3. Results and discussion

3.1. Stability of the simulation

A collection of properties (see Table 1) as a function of time was monitored to assess the stability of the simulation, and to check that the protein has properly equilibrated. The time evolution of the kinetic energy and of the various components of the potential energy were separately analysed. After a transient period of about 35 ps, both the kinetic and the potential total energies are almost stable (data not shown). An oscillatory behaviour with a long period (about 200 ps) is superimposed on the short time oscillations for both the protein–protein and the protein–solvent Coulomb energies (data not shown).

Table 1
Selected properties of Pc calculated from MD trajectories[a]

100–1100 ps	Mean	sd	Min	Max	Drift
All RMSD (nm)	0.174	0.011	0.139	0.201	0.020
C^{α} RMSD (nm)	0.141	0.012	0.104	0.169	0.020
R_g (nm)	1.235	0.005	1.221	1.251	0.007
E_{pot} (MJ/mol)	−172.27	0.48	−170.41	−174.10	−0.31
E_{kin} (MJ/mol)	29.87	0.24	28.98	30.90	0.03

[a]All values are calculated during the 100–1100 ps time interval. sd: standard deviation; Min: minimal value; Max: maximal value; RMSD: rms deviation; R_g: radius of gyration; E_{pot}: total potential energy; E_{kin}: total kinetic energy.

These oscillations, observed also in the simulation of BPTI with the same force field [27], show an opposite phase in the protein–protein and protein–solvent energies; such a behaviour being indicative of a slow exchange between the protein and the surrounding solvent.

The gyration radius, analysed as a function of time, reveals an initial fast increase (which is consistent with the solvation of charged and polar residues on the protein external surface), a slow decrease after 100 ps, and a fluctuation around a stable value after 450 ps.

The RMSD of the Pc atoms as a function of simulation time are plotted in Fig. 2. An initial rise is followed by small fluctuations between 0.15 nm and 0.20 nm, indicating that protein structure is practically stable after 100 ps. During the time interval of 100–1100 ps, only a small drift of the RMSD is registered. Thus, the first 100 ps are included in the equilibration phase, whereas the period from 100 to 1100 ps is used in the analysis. The overall RMSD from crystal structure is 0.174 ($\pm$0.011) nm calculated for all protein atoms and 0.141 ($\pm$0.012) nm calculated for C^{α} atoms (see Table 1). Such values are in agreement with other protein simulations, and suggest the existence of a few differences between the original and the solvated structure.

3.2. Protein dynamics: conformational changes, mobility and flexibility

A detailed analysis of the RMSD from crystal structure of the time-averaged simulation structure as function of backbone (henceforth, all references to backbone refer to N, C^{α}, C and O atoms) atoms is shown in Fig. 3. The graph reveals 4 major peaks, corresponding to residues 8–9, 48–54, 59–61 and 89–91.

Regions showing large differences from the initial structure could be involved in conformational changes, which may result from local relaxation in water of residues located on the external surface, and may be likely related to functional role. All the residues showing high values of RMSD are indeed located on the external surface of Pc, where they are not uniformly distributed. Residues 52–54 are located in the α-helix region, residues 8–9 and 48–50

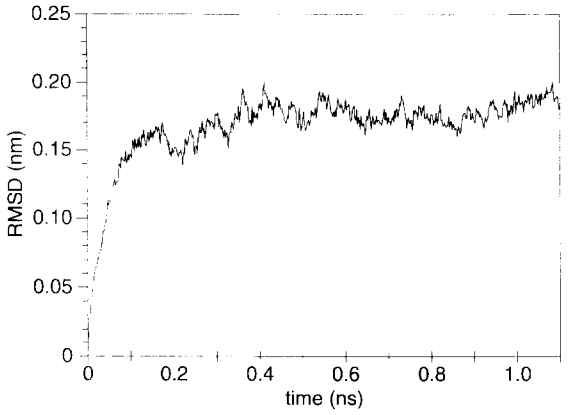

Fig. 2. Time evolution of all atoms RMSD from the X-ray structure for the simulated structure of Pc.

are located in turns on the protein top and bottom, residues 59–61 form the second strand of the acidic patch, residues 89–90 take part in the formation of the hydrophobic patch.

Asn99 also shows a high RMSD value, probably as a consequence of its terminal position in the protein.

The extent of the protein motions seen in the simulation has been evaluated by the RMSF of atomic positions in the trajectory, according to:

$$\Delta R = \left(\langle r_i^2 \rangle - \langle r_i \rangle^2\right)^{1/2} = \left[\frac{1}{N_{at}} \sum_{i=1}^{N_{at}} \left\langle \left\{ (\Delta x_i)^2 + (\Delta y_i)^2 + (\Delta z_i)^2 \right\} \right\rangle\right]^{\frac{1}{2}} \quad (1)$$

where N_{at} is the total number of atoms, Δx_i, Δy_i, Δz_i are the differences between the instantaneous and averaged atomic coordinates for the ith atom, and the brackets $\langle \ldots \rangle$ represent a time average.

A plot of these RMSF values versus backbone atoms is shown in Fig. 4. As expected, residues that show high values of RMSF are mainly found on the external surface, especially in turns and α-helix; low values of fluctuations are, on the contrary, observed for residues forming β-strands. For comparison, the positional RMSF derived from crystallographic B-factors according to the expression:

$$\left(\langle r_i^2 \rangle - \langle r_i \rangle^2\right) = 3B_i/8\pi^2 \quad (2)$$

have been shown in the same figure.

The two plots have similar trends and show main peaks at the same positions: the crystallographic values being smaller than the simulated ones for the majority of residues. The few segments where crystallographic RMSF are found to be larger, correspond to regions exhibiting low mobility. The values derived from crystallographic B-factors are indicative of atomic motion (thermal fluctuations within a single harmonic potential well), of conformational disorder (molecules existing in different conformations in the crystal) and lattice disorder (crystal heterogeneity); on the contrary, the values calculated in the simulation are connected almost with internal motion of the protein itself [28]. Therefore, the difference value between the simulated and crystallographic RMSF in regions where the former assume minimum values may be considered as indicative of lattice disorder contribution [29].

The largest quantitative differences between the two RMSF plots are instead found in the areas of higher mobility: namely residues 7–10, 23–24, 47–54, 72–75.

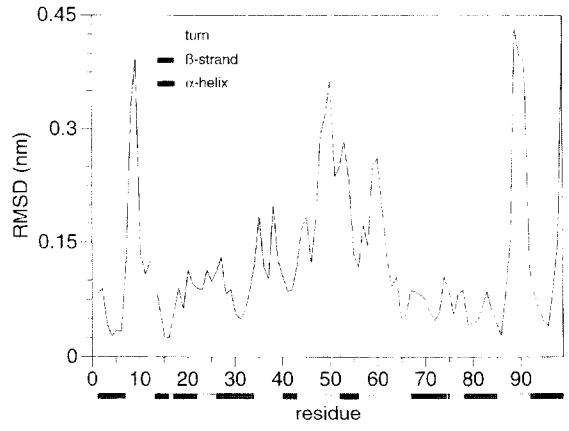

Fig. 3. RMSD between the crystal structure and the average conformation from MD simulation averaged over the 100–1100 ps time interval and plotted versus backbone atoms (N, C^{α}, C and O). Secondary structure elements are also shown.

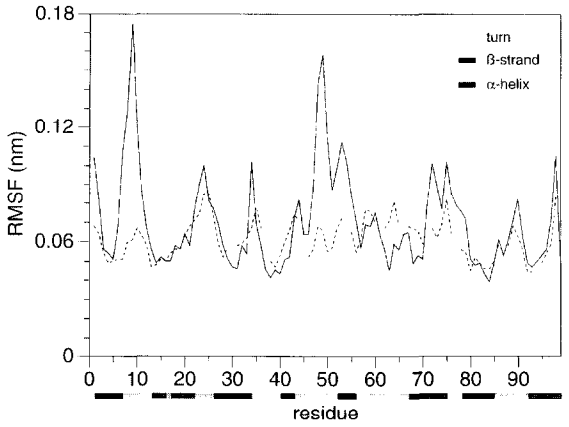

Fig. 4. Positional RMSF of Pc plotted versus the backbone atoms (N, C^{α}, C and O): calculated from the simulated structure averaged over the 100–1100 ps time interval (solid line) and obtained from crystal structure *B* factors (broken line). Secondary structure elements are also shown.

The turn corresponding to residues 7–10 is located on the northern side and is the most mobile region of the protein: together with turn 47–50 located in the southern side, they represent the mobile top and bottom of Pc, respectively. On the external surface, close to residues 47–50 are located the amino acids 23–24, 51–54 and 72–75.

Residues 1, 34 and 99 also show high values of RMSF. The mobility of the initial (1) and terminal (99) amino acids is probably due to end-effect; residue 34 shows instead an anomalous behaviour, since there is not any other adjacent residue with high mobility.

To better characterize the flexibility of some regions in Pc, we have also calculated the RMSF of the ϕ and Ψ dihedral angles during the simulation (data not shown). Few residues show high value of dihedral angle RMSF: they are located on the external surface of Pc, mostly in turns, and show medium-high values of positional RMSF also.

Positional and dihedral angles RMSF both show low values in the residues involved in the acidic and the hydrophobic patches and in the copper site, even when they are located in turns, where the mobility is usually higher. The only exception is represented by residue 34, located in the hydrophobic patch. Since the described regions are responsible for binding the reaction partners and for processing the electron transfer with them, these data may be interesting for understanding functional behaviours of solvated Pc.

3.3. Hydrogen bonds analysis

A geometric criterion was used to define the formation of a H-bond during the simulation: the hydrogen to acceptor distance has to be shorter than 0.32 nm and the donor–acceptor angle has to be larger than 120°. The frequency of intraprotein H-bonds has been computed and compared with data obtained from crystallography [20], which revealed 86 intramolecular H-bonds: 53 between polypeptide backbone groups and 33 involving side-chains. The results are reported in detail in Table 2 for H-bonds involving backbone atoms and in Table 3 for H-bonds involving side-chain atoms. Moreover, for better clarity in the discussion of structural and functional aspects, we have shown in Fig. 5 a diagram of the Pc residues, the H-bonds present in the crystal and/or in the simulated structure, the secondary structure elements, the acidic patch, the hydrophobic patch and the copper ion. We have considered as maintained the crystallographic H-bonds, which are present in the simulation with a time percentage greater than 0.25. Furthermore, H-bond interactions between residues have been considered as maintained even if the atom acting as donor or acceptor changes in the simulation: for instance, the crystallographic H-bond $85O^{\gamma}$–59O, which is replaced by H-bond $85O^{\gamma}$–$59O^{\eta 1}$, is considered as maintained in the simulation.

With regard to backbone H-bonds, a comparison of our results with crystallographic data reveals that 47 H-bond interactions are present also in the simulation with a percentage greater than 0.25 and two with percentage smaller than 0.25; four are lost and six new ones are formed (see Table 2). The close similarity between the simulated and the crystal H-bond network, found for backbone atoms, is particularly evident in the two β-sheets forming the Pc scaffold (Fig. 5): each pair of strands is indeed stabilized by two or more H-bonds which are maintained during the simulation. No H-bond is observed among strands belonging to different sheets, in agreement with the β-sandwich structure description of Pc [20].

In the α-helix region, backbone H-bonds are

maintained during the simulation; the H-bond between residue 55 and 51 (H-bond 55–51), which was described as 'weak' from NMR data [30] is recorded as well; on the contrary, there is no significant evidence for the H-bond 56–52 which was thought to stabilize the helix [30].

An analysis of the H-bond pattern has been reported in a previous paper [12], revealing its depen-

Table 2
Intramolecular hydrogen bonds involving backbone atoms[a]

Donor residue	Acceptor residue	Distance (Å) X-ray	Distance (Å) MD	Frequency MD
Ile1	Lys26	2.7	3.0	73.6
Val3	Val28	2.9	2.9	92.1
Leu4	Val15	2.9	2.9	95.7
Leu5	Lys30	2.8	2.9	99.8
Gly6	Leu4	–	3.2	55.5
Gly6	Ala13	2.9	3.0	91.7
Ala7	Ser11	2.9	2.9	85.9
Gly10	Ala7	2.9	3.0	66.7
Ala13	Ser11	3.1	3.0	56.3
Phe14	Gly91	–	3.1	96.6
Val15	Leu4	3.0	3.1	97.6
Phe19	Lys95	3.3	3.1	96.2
Ile21	Thr97	2.9	3.0	99.0
Gly24	Leu74	2.8	2.8	97.1
Glu25	Ser22	3.0	3.1	94.3
Ile27	Val72	2.9	2.9	67.8
Val28	Ile1	2.7	3.0	89.0
Phe29	Phe70	2.8	2.9	98.0
Lys30	Val3	2.7	2.9	97.8
Asn31	Glu68	2.8	3.0	99.3
Asn32	Leu5	2.8	3.1	90.0
Ala33	Leu5	3.1	2.9	97.8
His37	Leu63	3.0	3.0	98.5
Val40	Tyr83	2.9	2.9	95.5
Asp42	Ser81	3.0	3.0	97.9
Ser45	Asp42	2.9	3.1	25.6
Ile46	Asp42	–	3.0	98.5
Val50	Pro47	3.2	3.5	10.6
Lys54	Asp51	3.1	3.0	59.3
Ile55	Asp51	–	3.1	91.6
Ile55	Ala52	3.3	3.2	34.0
Ser56	Ala52	3.0	2.9	92.3
Met57	Ile39	3.0	2.9	98.5
Asp61	Ser58	3.1	3.0	94.9
Leu63	His37	2.8	2.9	99.8
Gly67	Asn31	2.9	3.0	97.5
Glu68	Ala65	3.2	3.3	81.4
Phe70	Phe29	3.0	3.0	99.4
Val72	Ile27	3.0	3.0	63.4
Leu74	Glu25	–	3.1	95.1
Gly78	Val98	3.0	2.9	90.1
Tyr80	Val96	2.8	3.1	91.9
Ser81	Ser45	3.0	4.0	0.1
Phe82	Gly94	3.0	2.9	99.5
Tyr83	Val40	2.8	3.0	99.9

Table 2 (continued)

Donor residue	Acceptor residue	Distance (Å) X-ray	Distance (Å) MD	Frequency MD
Cys84	Met92	3.0	3.2	98.5
Ser85	Asn38	–	3.0	97.4
His87	Cys84	3.2	3.1	45.5
Gln88	Cys84	2.9	2.9	90.9
Ala90	His87	3.2	3.7	–
Gly91	Gln88	2.9	3.2	–
Met92	His87	3.1	3.1	–
Met92	Ala90	–	3.1	91.3
Met92	Gly89	–	3.0	31.8
Val93	Gly91	3.0	3.0	85.7
Gly94	Phe82	3.0	3.0	91.7
Lys95	Ser17	2.9	2.9	92.4
Val96	Tyr80	2.9	2.9	99.6
Thr97	Phe19	3.0	3.1	96.0
Val98	Gly78	2.9	3.1	88.0
Asn99	Ile21	3.0	–	–

[a]Crystallographic data are obtained from Guss et al. [20]; MD data are obtained by averaging results from the time interval 100–1100 ps of the simulation.

dence on the hydration level of Pc: the best accuracy was obtained at the highest hydration level (2516 water molecules), even if a large percentage (0.36) of backbone H-bonds revealed from crystallography was not found. The current simulation therefore shows a considerable better agreement in the description of backbone H-bonds network. Such agreement may be due to several reasons: the increased number of water molecules representing more reliably the condition of full hydration, the new cutoff parameters (that permit a better treatment of interactions), and the longer time simulation (that permits better statistics).

The data reported in Table 3 show that several disagreements from the crystal results are recorded for side-chain H-bonds, probably because of the higher mobility of side-chain atoms with respect to backbone ones: 14 H-bonds are present with percentages greater than 0.25, 13 with percentages smaller than 0.25; six are lost and 22 new ones are formed.

By examining the general H-bond network shown in Fig. 5, we observe several changes between the initial and the solvated structure in the acidic patch, which is formed by residues 42–45 and 59–61. These amino acids, which are negatively charged at physiological pH [9], form two strands which likely bind Pc reaction partners [31,32], and surround Tyr83, the remote site for electron transfer from cytochrome *f*.

Rearrangements of the H-bond network are also observed in the region surrounding the copper atom and its four ligands (Fig. 5). The imidazole ring of His37 is stabilized by the ammide group of Ala33 [20]. In crystallography, the adjacent Asn38 was described as playing an important role in defining the geometry of the copper site by two H-bond interactions [20]: with Ser85, adjacent to the copper–ligand Cys84 also involved in the electron transfer process, and with Asp61. The latter was not present in the simulated structure, even if a new H-bond interaction with Glu59, also located in the acidic patch, is found. Cys84 and Met92 form a H-bond that is a part of the pattern between β-strands previously described, but is probably important for the stabilization of the methionine ligand, as well. His87 is the only copper ligand located on the external surface of the protein: two of the H-bonds revealed by crystallography are lost, so that the only interaction between His87 and Cys84 is maintained during the simulation.

H-bond interactions of the terminal residue Asn99 also show several differences from the crystal structure, probably because of end-effect.

The rupture and formation of H-bonds during the simulation may be the result of conformational changes as well as of solvent competition, since water may act as favorable donor or acceptor of hydrogen atoms. To assess the influence of water on

the intraprotein H-bond network, we have computed the average number of H-bonds between the protein and water molecules, which is given by [33]:

$$\langle N_{hb} \rangle = N_{hb}/N_{step} \quad (3)$$

where N_{hb} is the total number of H-bonds that are found during the analysed period (1000 ps) and N_{step} is the number of trajectory frames that is equal to 1×10^4. We have observed that charged residues form the largest number of H-bonds with water and that acidic amino acids preferentially act as hydrogen acceptors, while basic amino acids preferentially act

Table 3
Intramolecular hydrogen bonds involving side-chain atoms[a]

Donor residue	Donor atom	Acceptor residue	Acceptor atom	Distance (Å) X-ray	Distance (Å) MD	Frequency MD
Asp9	N	Asp9	$O^{\delta 1}$	–	3.1	47.3
Ser11	N	Asp9	$O^{\delta 1}$	3.1	3.4	21.4
Ser11	O^{γ}	Asp9	$O^{\delta 1}$	2.8	3.1	24.7
Ser22	N	Glu25	$O^{\epsilon 1}$	–	3.0	42.3
Ser22	N	Glu25	$O^{\epsilon 2}$	3.2	2.9	59.6
Ser22	O^{γ}	Glu25	$O^{\epsilon 1}$	–	2.9	53.9
Ser22	O^{γ}	Glu25	$O^{\epsilon 2}$	–	2.9	77.3
Asn31	$N^{\delta 2}$	Leu63	O	3.0	3.1	26.1
Asn31	$N^{\delta 2}$	Asn64	O	–	3.2	50.2
Asn31	$N^{\delta 2}$	Ala65	O	3.0	3.0	83.8
Asn32	$N^{\delta 2}$	Gly6	O	2.8	3.0	61.1
Asn32	$N^{\delta 2}$	Asp8	$O^{\delta 1}$	2.8	3.3	25.7
His37	$N^{\epsilon 2}$	Ala33	O	2.7	2.8	94.3
Asn38	N	Cys84	S^{γ}	3.5	–	–
Asn38	$N^{\delta 2}$	Met57	O	–	3.2	92.0
Asn38	$N^{\delta 2}$	Glu59	$O^{\epsilon 2}$	–	3.0	83.8
Asn38	$N^{\delta 2}$	Asp61	O	2.9	3.1	–
Asn38	$N^{\delta 2}$	Ser85	O^{γ}	3.0	3.8	29.9
Phe41	N	Ser56	O^{γ}	3.0	3.0	90.1
Asp44	N	Asp42	$O^{\delta 1}$	3.0	3.1	71.3
Ser45	N	Asp42	$O^{\delta 1}$	–	3.2	68.6
Ser45	O^{γ}	Asp42	$O^{\delta 2}$	3.0	3.2	13.9
Ser53	N	Asp51	$O^{\delta 2}$	2.9	3.0	59.0
Ser53	O^{γ}	Asp51	$O^{\delta 1}$	–	3.3	55.5
Ser53	O^{γ}	Asp51	$O^{\delta 2}$	2.7	2.8	92.2
Lys54	N	Asp51	$O^{\delta 2}$	–	3.4	65.2
Lys54	N^{ζ}	Asp51	$O^{\delta 1}$	2.2	3.4	–
Ser56	O^{γ}	Ala52	O	2.9	3.6	0.1
Ser58	N	Asp61	$O^{\delta 2}$	2.5	3.8	2.0
Ser58	O^{γ}	Glu60	$O^{\epsilon 2}$	2.4	3.0	10.7
Glu60	N	Ser58	O^{γ}	3.3	4.0	0.9
Glu60	N	Glu60	$O^{\epsilon 1}$	–	3.2	79.5
Glu60	N	Glu60	$O^{\epsilon 2}$	–	3.3	60.0
Asn64	N	Glu68	$O^{\epsilon 1}$	–	3.0	94.0
Asn64	N	Glu68	$O^{\epsilon 2}$	2.5	3.5	63.4
Asn64	$N^{\delta 2}$	Glu68	$O^{\epsilon 1}$	–	3.3	51.2
Ala65	N	Glu68	$O^{\epsilon 1}$	2.7	3.6	65.7
Ala65	N	Glu68	$O^{\epsilon 2}$	–	3.0	97.8
Lys66	N^{ζ}	Asn32	$O^{\delta 1}$	3.0	3.1	16.5
Lys66	N^{ζ}	Asn32	O	3.0	3.1	17.2
Thr69	$O^{\gamma 1}$	Glu71	$O^{\epsilon 2}$	2.7	3.8	0.1
Ser75	O^{γ}	Asn76	$O^{\delta 1}$	3.3	3.6	0.2
Asn76	$N^{\delta 2}$	Tyr80	O^{η}	–	3.0	49.9

Table 3 (continued)

Donor residue	Donor atom	Acceptor residue	Acceptor atom	Distance (Å) X-ray	Distance (Å) MD	Frequency MD
Tyr80	O^{η}	Asn76	O	3.3	2.8	8.1
Ser81	N	Ser45	O^{γ}	–	3.1	92.8
Tyr83	O^{η}	Glu59	$O^{\epsilon 2}$	2.8	–	–
Ser85	N	Asn38	$O^{\delta 1}$	2.8	3.7	0.4
Ser85	O^{γ}	Glu59	O	3.2	–	–
Ser85	O^{γ}	Glu59	$O^{\epsilon 1}$	3.2	3.5	46.3
Ser85	O^{γ}	Glu59	$O^{\epsilon 2}$	3.2	2.8	82.9
Thr97	$O^{\gamma 1}$	Glu79	$O^{\epsilon 2}$	–	2.8	82.6
Asn99	$N^{\delta 2}$	Gly78	O	–	3.0	47.1
Asn99	$N^{\delta 2}$	Glu79	$O^{\epsilon 1}$	–	3.2	58.5
Asn99	$N^{\delta 2}$	Glu79	$O^{\epsilon 2}$	–	3.3	40.0
Asn99	$O^{\delta 1}$	Ile21	O	3.1	–	–

[a]Crystallographic data are obtained from Guss et al. [20]; MD data are obtained by averaging results from the time interval 100–1100 ps of the simulation.

as hydrogen donors. Moreover, $\langle N_{hb} \rangle$ values averaged over residues are found to be larger for acidic residues than for basic ones; this meaning that water preferentially form H-bonds as donor. As regard to noncharged residues, polar amino acids show lower values of $\langle N_{hb} \rangle$, while nonpolar amino acids usually show very low values. These results are in agreement with those obtained from a previous simulation [12], and with the hydration numbers obtained by NMR experiments [34]. Discrepancies between simulated and experimental values are observed for residue types Asn (4.12 ± 3.97 vs. 2.0 ± 0.5) and Tyr (1.97 ± 1.3 vs. 5.5 ± 1.0). We remark that the $\langle N_{hb} \rangle$ average for Asn residues is affected by the anomalous behaviour of residue 99 that shows a considerably higher value than other Asn amino acids: the $\langle N_{hb} \rangle$ value averaged over the remaining Asn residues (2.47 ± 1.72) being, indeed, in agreement with the experimental one. The $\langle N_{hb} \rangle$ average for Tyr may be instead affected by bad statistics, since the number of these residues in the protein is only two.

Finally, we have performed a detailed analysis to evaluate the solvent effect on each residue, by comparing the number of H-bonds in the crystal structure and the intraprotein and water–protein H-bonds in the simulation. We have found that the majority of the amino acids involved in H-bonds, present in crystallography but not in the simulation, are located on the external surface of Pc where they form several H-bonds with water molecules. Therefore, these data are indicative of an effective role of water in the modulation of the spatial and temporal organization of the H-bond network in solvent-exposed regions.

3.4. Dynamical cross-correlations

The analysis of the collective character exhibited by protein motions is interesting for its relevance in biological functions [35,36]. Such analysis can be performed by building a dynamical cross-correlation map (DCCM), which is a matrix composed by elements C_{ij} [37]:

$$C_{ij} = \frac{c_{ij}}{c_{ii}^{1/2} c_{jj}^{1/2}} = \frac{\langle r r_j \rangle - \langle r_i \rangle - \langle r_j \rangle}{\left[\left(\langle r_i^2 \rangle - \langle r_i \rangle^2 \right) \left(\langle r_j^2 \rangle - \langle r_j \rangle^2 \right) \right]^{1/2}} \qquad (4)$$

C_{ij} provides the correlation coefficients for residue displacements along a straight line: positive values are indicative of motions in the same direction, while negative values are indicative of motions in opposite directions. Correlated motions may occur among neighbouring residues that form a secondary structure, as well as among distant amino acids belonging to different regions or domains [38]. Cross-correlation coefficients do not bear any information about the magnitude of the motions: therefore, it may

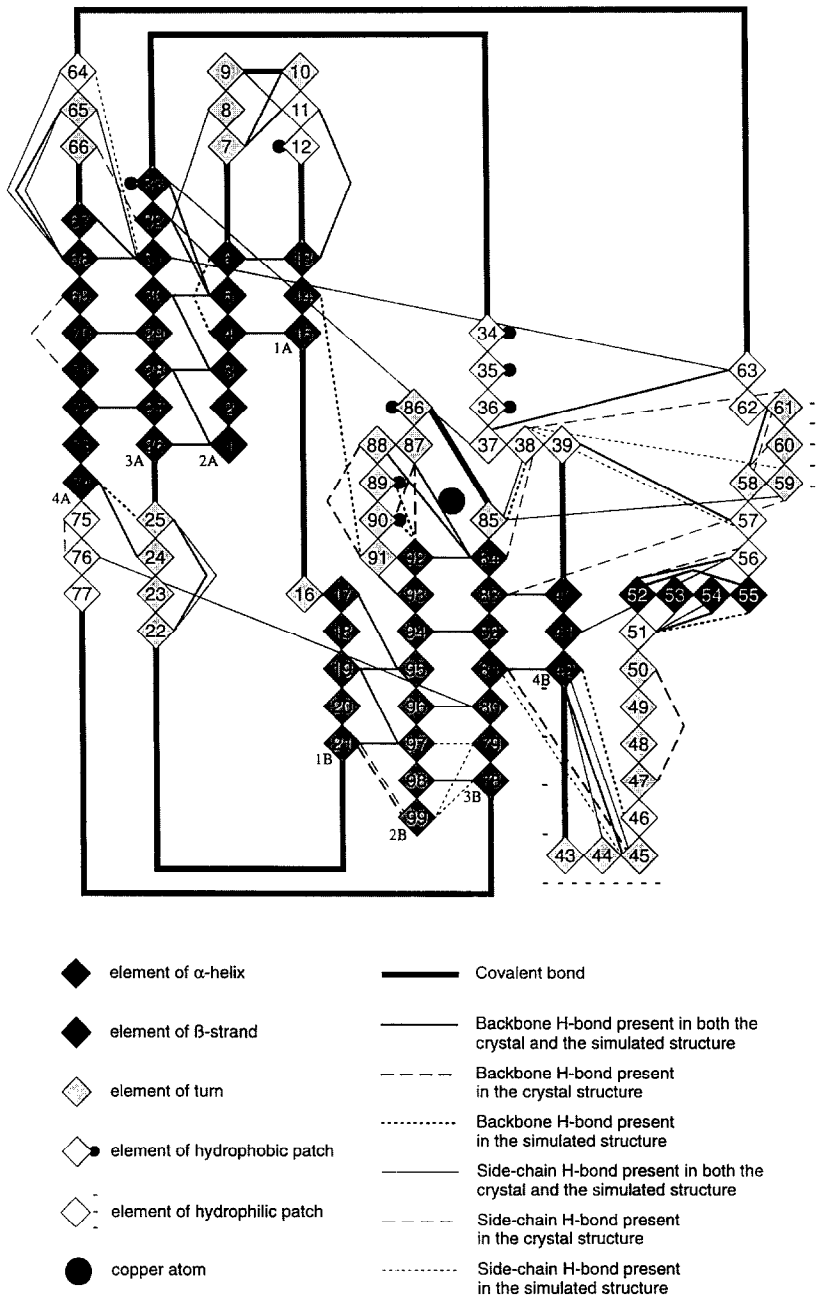

Fig. 5. Diagram of the hydrogen bonding patterns for the crystal and the simulated structure of Pc. Grey intensity represents secondary structure of amino acids, as defined by crystallographic results. The copper atom, the acidic patch and the hydrophobic patch are also indicated.

happen that both small- and large-scale collective motions are expressed by the same C_{ij} [37]. The DCCM corresponding to an average time of 1000 ps is shown in Fig. 6: the long time interval has been chosen to provide more accurate results. Positive correlations are plotted in the upper left triangle of

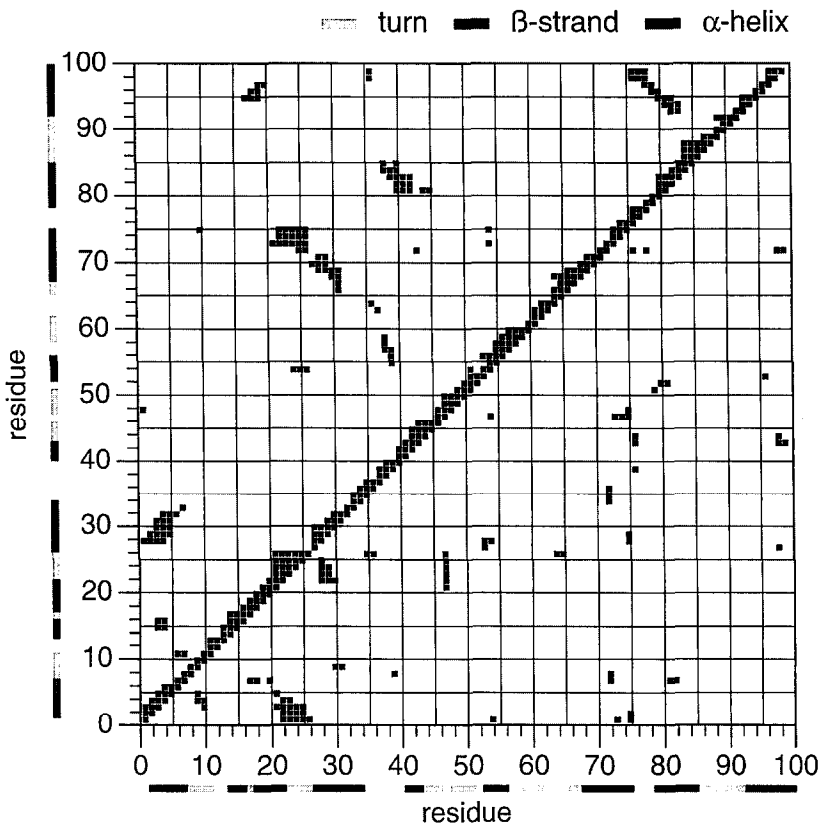

Fig. 6. Dynamical cross-correlation map for the C^{α} atom pairs of Pc averaged over the 100–1100 ps time interval. Only correlations coefficients with an absolute value greater than 0.3 are shown. Positive correlations are mapped in the upper left triangle, negative correlations in the lower right triangle. Secondary structure elements are shown along the axes.

the DCCM, while negative correlations are plotted in the lower right triangle.

Along the diagonal, correlations give rise to small plumes representing correlated fluctuations mostly between residues that are near in the primary structure and lie in the same secondary structure. The plume in position (14,17) shows a correlated motion between the strands 1A and 1B. The two strands belong to different β-sheets, and are separated by a very short turn, made up by a single residue, Pro16.

Either isolated dots or clusters are present in the matrix. Most of the clusters are present in the upper left triangle: six of them, located around positions (3,15), (4,30), (29,69), (18,95), (80,95), (41,82) represent positively correlated motions between pairs of adjacent β-strands; with the exception of the first, they have a characteristic diagonal shape, with an ascending or descending trend that is respectively correlated to the parallel or antiparallel coupling of strands in the tertiary structure. As previously reported, β-strands that are adjacent in the tertiary structure are connected by a certain number of H-bonds. It is well known that the presence of H-bond interactions gives rise to correlated motions between the involved residues; so cross-correlation data relative to β-strand residues reveal the importance of hydrogen bonding pattern for concerted motions that prevent the breakdown of secondary structure elements in the protein. It is also interesting to note that the correlation of (14,17), previously mentioned, does not correspond to any H-bond between the two strands.

The group around (24,74) is formed by residues that are near the tertiary structure of the protein: H-bond interactions are found among such residues, as well as among the amino acids that give rise to

matrix dots in positions (7,10), (37,63), (38,57), (38,59), (39,57), (38,85). Residues represented by matrix elements (44,81), (54,73), (54,75) are fairly near in tertiary structure, while residues represented by (1,48), (10,75), (25,54), (43,72), (36,98) are far apart.

Differently from positive correlations, the lower left triangle of negative correlations is characterized by the presence of several isolated dots, and few clusters. Negative correlations occur mostly among residues that are far apart in the tertiary structure of the protein: exceptions involves amino acids indicated by matrix elements (26,1), (73,1), (80–81,52), (98,27), (98–99,72), which are near or fairly near in space.

The group around (23,2) is indicative of negatively correlated motions between residues 21–26, which form a turn, and residues 1–5, which form the β-strand 2A. These residues are set along two straight lines on the western side of Pc; the resulting turn and β-strand move in opposite directions. In the central part, residues 1 and 26 are adjacent and form a H-bond: this is the only H-bond that gives rise to a negative correlation during the simulation, probably for the presence of a single bond in a peculiar position. The two groups of residues may indeed move along opposite directions to some extent, without breaking the H-bond, which is in the central part of them.

Residues 22–25 also show negative correlations with residues 28–30, which are located on the β-strand 3A. The same kind of correlated motions between turn and β-strand residues are also indicated by the map elements around (10,4), (18,7), (30,9), (64,26), (72,7), (81,7), (98,43).

We should also mention the existence of cross-correlation elements among residues 10 and 75, 21–26 and 73–75, 24–26 and 54, 54 and 73, 54 and 75 (for positive coefficients) and among residues 47 and 21–26, 54 and 47, 72 and 7–8, 73–75 and 47–48, (for negative coefficients), which involve several residues showing high RMSF values.

By using an average time of 1000 ps, fast correlated motions may be not detected, being characteristic of shorter times. For this reason, and for a comparison with the described results, we have computed data relative to DCCMs obtained with averaging times of 500, 250, 100 ps, using different starting point times (data not shown). We have found many differences among the maps, although some elements are maintained in several or all DCCMs. In particular, cross-correlations among residues that also form H-bonds can be found in all the DCCMs, whatever the length or the starting position of the sample time was chosen. On the contrary, elements which are located far apart in the tertiary structure give rise to a pattern of correlations which changes in the different maps: this result being true for both positive and negative correlations. However, some caution is in order when analysing results correspondent to short averaging time, during which RMSF might not have converged [37].

4. Conclusion

Studies of MD simulation of protein in water give information about the effect of solvation and the dynamical events that may occur in a molecule during the investigated time. During the long simulation in water, Pc maintains its globular shape and its β-sandwich structure. The β-strands forming the protein scaffold are connected by several H-bonds that are almost maintained during all the simulation time; they give rise to concerted motions among the residues that form the two β-sheets surrounding the hydrophobic core where the copper-containing active site is situated. Conversely, solvent-exposed regions are under the influence of water, which can modulate the temporal and spatial organization of the intraprotein H-bond network. The analysis of the network is of particular interest in studies on globular proteins, for their likely role in folding and in stabilizing the structure. Also, the detailed description of H-bond network achieved in the present simulation may deserve some interest in elucidating the electron transfer route in Pc, which might involve both hydrogen and covalent bonds [39].

The influence of water is put into evidence also by the analysis of RMSD from the starting structure, which mainly reveals that high values are characteristic of residues located on the external surface.

An analysis of positional RMSF, which provide information about mobility, shows high values for residues located on the external surface, particularly in the turns located on the top and the bottom of Pc.

which also show more flexibility, as indicated by dihedral angle RMSF.

The overall dynamical picture of Pc is consistent with a general stiffness of the most important functional regions: the acidic and hydrophobic patches and the copper site, which show low mobility and low flexibility, while undergoing several rearrangements in the H-bonds network as the effect of solvation.

References

[1] R.L. Hayward, H.D. Middendorf, U. Wanderlingh, J.C. Smith, J. Chem. Phys. 102 (1995) 5525.
[2] M. Philippopoulos, C. Lim, J. Mol. Biol. 254 (1995) 771.
[3] L.W. Ungar, N.F. Scherer, G.A. Voth, Biophys. J. 72 (1997) 5.
[4] P.J. Kraulis, J. Appl. Crystallogr. 24 (1991) 946.
[5] S. Cannistraro, J. Phys. (France) 51 (1990) 131.
[6] A.R. Drews, B.D. Thayer, H.J. Stapleton, G.C. Wagner, G. Giugliarelli, S. Cannistraro, Biophys. J. 57 (1990) 157.
[7] A.R. Bizzarri, S. Cannistraro, Mol. Phys. 85 (5) (1995) 913.
[8] A.G. Sykes, Chem. Soc. Rev. 14 (1985) 283.
[9] E.L. Gross, Photosynth. Res. 37 (1993) 103.
[10] M.R. Redinbo, T.O. Yeates, S. Merchant, Bioener J. Biomembr. 26 (1) (1994) 49.
[11] C.X. Wang, A.R. Bizzarri, Y.W. Xu, S. Cannistraro, Chem. Phys. 183 (1994) 155.
[12] A.R. Bizzarri, C.X. Wang, W.Z. Chen, S. Cannistraro, Chem. Phys. 201 (1995) 463.
[13] Y.N. Chirgadze, A.M. Ovsepyan, Biopolymers 11 (1972) 2179.
[14] A.R. Bizzarri, S. Cannistraro, Phys. Rev. E 53 (4) (1996) 1.
[15] C. Rocchi, A.R. Bizzarri, S. Cannistraro, Chem. Phys. 208 (1996) 95.
[16] P.M. Colman, H.C. Freeman, J.M. Guss, M. Murata, V.A. Norris, J.A.M. Ramshaw, M.P. Venkatappa, Nature 272 (1978) 9.
[17] P.J. Steinbach, B.R. Brooks, Comp. J. Chem. 15 (7) (1994) 667.
[18] A.E. Mark, S.P. van Helden, P.E. Smith, L.H.M. Janssen, W.F. van Gunsteren, J. Am. Chem. Soc. 116 (1994) 6293.
[19] X. Daura, B. Oliva, E. Querol, F.X. Avilés, O. Tapia, Proteins: Struct., Funct. and Genet. 25 (1996) 89.
[20] J.M. Guss, H.D. Bartunik, H.C. Freeman, Acta Crystallogr. B 48 (1992) 790.
[21] W.F. van Gunsteren, H.J.C. Berendsen, Groningen Molecular Simulation (GROMOS) Library Manual, Biomos, Groningen, 1987.
[22] A. Vedani, D.W. Huhta, J. Am. Chem. Soc. 112 (1990) 4759.
[23] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, J. Phys. Chem. 91 (1987) 6269.
[24] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Di Nola, J.R. Haak, J. Chem. Phys. 81 (1984) 3684.
[25] J.P. Rickaert, G. Ciccotti, H.J.C. Berendsen, J. Comput. Phys. 23 (1977) 327.
[26] W. Kabsch, Acta Crystallogr. Sect. A 32 (1976) 922.
[27] R.M. Brunne, K.D. Berndt, P. Güntert, K. Wüthrich, W.F. van Gunsteren, Proteins: Struct., Funct. and Genet. 23 (1995) 49.
[28] G.A. Petsko, D. Ringe, Annu. Rev. Biophys. Bioeng. 13 (1984) 331.
[29] E.M. Storch, V. Daggett, Biochemistry 34 (1995) 9682.
[30] J.M. Moore, C.A. Lepre, G.P. Gippert, W.J. Chazin, D.A. Case, P.E. Wright, J. Mol. Biol. 221 (1991) 533.
[31] Z. Adam, R. Malkin, Biochim. Biophys. Acta 975 (1989) 158.
[32] D.C. Pearson Jr., E.L. Gross, E.S. David, Biophys. J. 71 (1996) 64.
[33] H.J.C. Berendsen, W.F. van Gunsteren, H.R.J. Zwinderman, R.G. Geurtsen, Ann. NY Acad. Sci. 482 (1986) 268.
[34] I.D. Kuntz, J. Am. Chem. Soc. 93 (1971) 514.
[35] J.A. McCammon, S.C. Harvey, Dynamics of Proteins and Nucleic Acids, Cambridge Univ. Press, Cambridge, 1987.
[36] C.L. Brooks III, M. Karplus, B.M. Pettitt, Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Wiley, New York, 1988.
[37] P.H. Hünenberger, A.E. Mark, W.F. van Gunsteren, J. Mol. Biol. 252 (1995) 492.
[38] W.E. Harte Jr., S. Swaminathan, M.M. Mansuri, J.C. Martin, I.E. Rosenberg, D.L. Beveridge, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 8864.
[39] D.N. Beratan, J.N. Betts, J.N. Onuchic, Science 252 (1991) 1285.