Changes in non-core regions stabilise plastocyanin from the thermophilic cyanobacterium Phormidium laminosum

ARTICLE *in* EUROPEAN JOURNAL OF BIOCHEMISTRY · NOVEMBER 2009

Impact Factor: 2.54 · DOI: 10.1007/s00775-009-0605-6 · Source: PubMed

5 AUTHORS, INCLUDING:

Simone Raugei
Pacific Northwest National Laboratory
96 PUBLICATIONS **1,742** CITATIONS
SEE PROFILE

Miguel A. De la Rosa
Universidad de Sevilla
178 PUBLICATIONS **2,605** CITATIONS
SEE PROFILE

Antonio Díaz-Quintana
Universidad de Sevilla
70 PUBLICATIONS **986** CITATIONS
SEE PROFILE

Paolo Carloni
Forschungszentrum Jülich
320 PUBLICATIONS **6,079** CITATIONS
SEE PROFILE

J Biol Inorg Chem (2010) 15:329–338
DOI 10.1007/s00775-009-0605-6

ORIGINAL PAPER

Changes in non-core regions stabilise plastocyanin from the thermophilic cyanobacterium *Phormidium laminosum*

Francisco J. Muñoz-López · Simone Raugei · Miguel A. De la Rosa · Antonio J. Díaz-Quintana · Paolo Carloni

Received: 1 July 2009 / Accepted: 19 October 2009 / Published online: 14 November 2009

Abstract We report a theoretical investigation on the different stabilities of two plastocyanins. The first one belongs to the thermophilic cyanobacterium *Phormidium laminosum* and the second one belongs to its mesophilic relative *Synechocystis* sp. These proteins share the same topology and secondary-structure elements; however, the melting temperatures of their oxidised species differ by approximately 15 K. Long-time-scale molecular dynamics simulations, performed at different temperatures, show that the thermophilic protein optimises a set of intramolecular interactions (interstrand hydrogen bonding, salt bridging and hydrophobic clustering) within the region that comprises the strands β_5 and β_6, loop L_5 and the helix. This region exhibits most of the differences in the primary sequence between the two proteins and, in addition, it is involved in the interaction with known physiological partners. Further work is in progress to unveil the specific structural features responsible for the different thermal stability of the two proteins.

Electronic supplementary material The online version of this article (doi:10.1007/s00775-009-0605-6) contains supplementary material, which is available to authorized users.

F. J. Muñoz-López · M. A. De la Rosa · A. J. Díaz-Quintana (✉)
Instituto de Bioquímica Vegetal y Fotosíntesis,
Universidad de Sevilla y C.S.I.C.,
Avda. Américo Vespucio, 49,
41092 Seville, Spain
e-mail: qzaida@us.es

F. J. Muñoz-López · S. Raugei (✉) · P. Carloni
SISSA/ISAS and INFM-DEMOCRITOS,
Modelling Center for Research in Atomistic Simulation,
via Beirut 2,
34014 Trieste, Italy
e-mail: raugei@sissa.it

Keywords Blue copper protein · Molecular dynamics · Plastocyanin · Protein stability · Thermophilic cyanobacteria

Abbreviations

MD	Molecular dynamics
MPc	Plastocyanin from the mesophilic cyanobacterium *Synechocystis* sp. PCC 6803
PDB	Protein Data Bank
RMSD	Root mean square deviation
SAS	Solvent-accessible surface
TPc	Plastocyanin from the thermophilic cyanobacterium *Phormidium laminosum*

Introduction

The understanding of thermal stability in proteins has been the subject of thorough research because of its incipient applications in biotechnology and the interest in unveiling the fundamentals of protein folding. Many studies have revealed that the stability of a given protein depends on the balance among many different factors [1–18], such as (1) the presence of electrostatic and hydrogen-bond networks [1], (2) the degree of inner packing [2, 3], (3) deletion or reduction of loops [4], (4) the entropy gap between folded and unfolded states [5], (5) interactions involving α-helix dipoles [6], (6) optimisation of the hydrophobic interactions at the protein core [7, 8], (7) minimisation of the solvent-accessible surface [9] and (8) preference or discrimination for specific types of amino acids [10, 11]. High-throughput genomics [12], structural data mining [13, 14] and structural genomics studies [15] comparing

different organisms have underlined the putative role of salt bridges and the protein compactness in protein thermostability. The strategy followed by nature in each case, however, seems to be heterogeneous [1–18].

Proteins from thermophilic organisms withstand high temperatures in all their folding stages despite their showing the same structural elements as their mesophilic homologues [17]. This is the case of the plastocyanin (Pc) from the thermophilic cyanobacterium *Phormidium laminosum* (TPc) and that from its less thermophilic (from here on referred to as mesophilic) counterpart *Synechocystis* sp. PCC 6803 (MPc) [19, 20].

Plastocyanins are small blue copper proteins [21] that act as mobile electron carriers in the thylakoid lumen of cyanobacteria, algae and plants [22, 23]. The interactions with the physiological partners involve two sites: a hydrophobic patch containing the active-site pocket (north patch) and a surface region (east patch), which is rich in charged residues. The active site of plastocyanin consists of a single copper atom bonded to two His nitrogen atoms, one Cys sulphur atom and a more weakly bound Met sulphur atom, yielding a coordination sphere in a distorted tetrahedral geometry [24, 25] that is responsible for its particular spectroscopic properties among copper compounds. Both TPc and MPc contain two β-sheets and a α-helix arranged in a Greek-key topology (Fig. 1).

The unfolding of TPc and MPc has been proved to be irreversible under both aerobic and anaerobic conditions [19], following the Lumry–Eyring scheme [26] like other plastocyanins or cupredoxins [27–30]. The irreversibility of this process is due to an exothermic step subsequent to the endothermic peptide unfolding [27, 30].

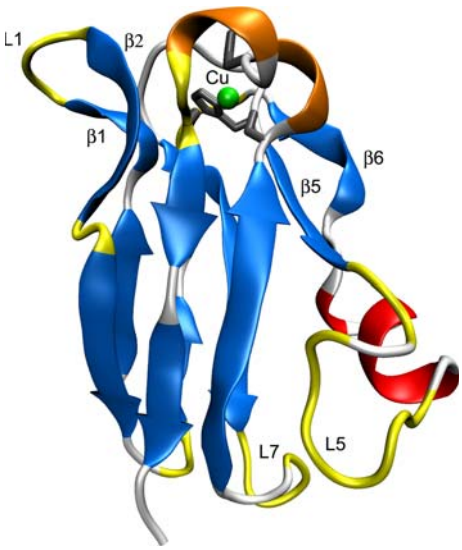

Fig. 1 Robertson diagram of the three-dimensional structure of *Phormidium* plastocyanin (Protein Data Bank entry 1BAW). *Blue* β-strands, *red* α-helix, *orange* 3_{10} helix, *yellow* turn, *white* loop. The copper atom is in *green* and its ligands are in *grey*. Labels mark some important secondary structure elements: the first, second, fifth and sixth strands (*β1*, *β2*, *β5* and *β6*) and the first and fifth turns (*L1* and *L5*). The image was created with the program VMD [62]

To measure the stability of both proteins, the melting temperatures (T_m) of unfolding at different pH and ionic strength values were determined by Feio et al. [19, 20]. In any condition tested, TPc showed a higher T_m than MPc. T_m of TPc and MPc oxidised species differed by approximately 15 K and T_m of the reduced species differed by approximately 5 K. Both proteins showed a maximum thermal resistance near their physiological pH, though that of MPc features a mild ionic strength dependency that is not found in the case of TPc. The copper atom stabilises both proteins [30–33] as cofactors usually do, thereby both TPc and MPc apoproteins show lower T_m values than their corresponding holoproteins. The sequence identity of the two proteins is 63%, with the hydrophobic core of the β-barrel being highly conserved. However, contrary to what could be expected [4], TPc has more residues than MPc, as TPc shows two residue insertions at loops.

To rationalise the structural features that cause the experimentally observed different thermal resistances, we performed molecular dynamics (MD) simulations at different temperatures of both oxidised plastocyanins.

Methods

Copper centre parameterisation

The force field employed for the standard residues was AMBER99 [34], complemented with the parameters developed by Comba and Remenyi [35] for the oxidised copper centre except for the atomic RESP [36] charges, which were calculated on a larger model to better reproduce the active-site electrostatic potential. To this end, the model of Comba and Remenyi, which is formed by the metal centre and the ligand side chains, was extended to include the ligands backbone atoms, and every ligand residue was capped with amine and acetyl groups. The amine moiety contains the amine nitrogen of the next residue in sequence of the protein. The acetyl moiety instead comprises the α-carbons, carbonyl carbons and the carbonyl oxygen of the preceding residue. Heavy atom coordinates were taken from the X-ray structure of TPc corresponding to Protein Data Bank (PDB) entry 1BAW [37] and the NMR model of MPc, PDB entry 1J5C (model number 19) [38]. The initial coordinates were subjected to geometry optimisation calculations before the definitive charge set was obtained. Before optimisation, the positions of hydrogen atoms were initially relaxed by the semiempirical method AM1 [39], as implemented in Gaussian 03 [40]. For this purpose only, copper, not included in the

standard AM1 parameterisation, was replaced by zinc. The new set of atomic point charges is reported in Table S1.

Molecular dynamics simulations

MD simulations were performed on both MPc and TPc. The starting coordinates for the simulations were taken from the X-ray structure of TPc (PDB entry 1BAW [37]), and the NMR structure of MPc (PDB entry 1J5C [38]). Among the 35 NMR models available for MPc, we chose model 19, whose copper centre geometry was more consistent with the adopted parameterisation of the copper centre, as regards distances between the ligands and the metal ion, to avoid clashes between them in the starting structure. An additional set of initial coordinates for MPc was obtained from a MPc triple mutant [41] by transforming its sequence back to the wild-type sequence.

The proteins were solvated in a cubic box containing TIP3P water molecules [42] (4,599 and 4,935 molecules for TPc and MPc, respectively). Sodium ions were added to achieve electroneutrality. The AMBER 9 [43] molecular simulation package was used for minimisation and MD equilibration of the system, whereas GROMACS 3.3.1 [44, 45] was used for the final equilibration, production phase and trajectory analysis. The positions of protein side chains and water molecules were separately optimised up to a residual energy gradient lower than 0.01 kcal mol^{-1} $Å^{-1}$. The solvent molecules were then randomised by a 50-ps-long MD simulation at 298 K and constant volume, keeping the protein fixed. Next, the whole system was gradually heated up from 0 to 298 K in the NPT ensemble using Berendsen's thermostat and barostat [46] with time constants of 0.5 and 1.0 ps, respectively. Then, after an additional period of equilibration, trajectories were collected for at least 11 ns (Table 1). Then, starting from an equilibrated configuration at 298 K, we carried out successive simulations at higher temperatures. For this purpose, the temperature was monotonically increased in discrete steps. In particular, simulations were performed at the following temperatures: (1) room temperature, (2) the temperature at which the physiological activity is interrupted in vivo (323 and 333 K for MPc and TPc, respectively [47]), (3) the in vitro melting temperatures (338 and 354 K [19, 20]) and (4) a temperature far above the melting point to facilitate protein unfolding on the time scale spanned by the MD simulations (498 K). A summary of all simulations performed can be found in Table 1. In addition, a linear heating up of both proteins from 298 to 498 K in 10 ns was also performed. In all the simulations, the integration time step was 1.5 fs. Electrostatics was treated with the fast particle mesh Ewald method and a cut-off of 0.9 nm. The cut-off for Lennard-Jones interactions was 1.4 nm. Finally, C–H bond lengths were kept fixed.

Table 1 Summary of the molecular dynamics simulations performed

TPc			MPc		
Temperature (K)	t_{eq} (ns)	t_{tot} (ns)	Temperature (K)	t_{eq} (ns)	t_{tot} (ns)
298	2.5	20	298	5	11
333	1	11	323	1.5	11.5
338	2	13	338	1.8	11
354	2.5	11	354	1.5	13
498	2.8	14	498	1	11

TPc plastocyanin from the thermophilic cyanobacterium *Phormidium laminosum*, *MPc* plastocyanin from the mesophilic cyanobacterium *Synechocystis* sp. PCC 6803, t_{eq} equilibration time (defined as the time at which the root mean square deviation has reached a plateau), t_{tot} total simulation time

To test if the equilibration protocol affected the results of MD calculations, we performed two additional simulations of MPc (Figs. S1, S2). In the first one, we applied restraints to the backbone atoms. In the other, we raised the temperature of the solvent during the equilibration steps. The convergence of the conformational sampling was further assessed by principal component analysis, using g_covar and g_anaeig modules of the GROMACS 3.2 package [44, 45]. Protein motions were first characterised in terms of the eigenvectors of the covariance matrix $c_{ij,\alpha\beta} = \langle \delta x_{i\alpha} \cdot \delta x_{j\beta} \rangle$, where α and β denote the Cartesian components and the brackets indicate the time average, obtained from the instantaneous displacement vector $\delta \mathbf{x}_i$ (after removal of rigid body motions) of the ith C_α atom from the reference (time-averaged) position [48, 49]. Then, an estimate of the degree of convergence of sampling was obtained from the subspace overlap between eigenvectors obtained from different time intervals [48]:

$$\text{Overlap}(v, w) = \frac{1}{n} \cdot \sum_{i=1}^{n} \sum_{j=1}^{m} (v_i \cdot w_j)^2$$

where v and w are the two orthonormal vector sets to be compared. Given that each set is made of orthonormal vectors, the overlap is 1 when the two sets are identical and 0 when all the vectors of v are orthonormal to those of w.

Contact orders, which are the average relative sequence separation of native contacts [50], were calculated using the atomic coordinates of α-carbons and a cut-off distance of 6 Å. Statistical and topological contact network analyses [51] were carried out using a cut-off of 4.5-Å distance between heavy atoms, and a three-residue-distance cut-off [15] along the sequence was set to eliminate trivial—short range—contacts from analysis. The degree of a node is defined as

$$k_i = \sum_{j=1}^{n} a_{ij}$$

where a_{ij} are the elements of the adjacency matrix, which take the value 1 if an edge connects the nodes i and j and 0 otherwise. To calculate average node clustering coefficients, every link between nodes was weighted according to the number of pairs of contacting heavy atoms of the two residues involved, according to the following equation [51]:

$$C_i^w = \frac{1}{(k_i - 1)\sum_{j=1}^{n} a_{ij} \cdot w_{ij}} \cdot \sum_{j,h} a_{ij} \cdot a_{ih} \cdot a_{jh} \cdot \left(\frac{w_{ij} + w_{ih}}{2}\right)$$

where w_{ij} is the relative weight of a given link.

The number of salt bridges was calculated in terms of the time-averaged number of contacts between Asp or Glu carboxylic oxygen atoms and Lys amine or Arg guanidinium nitrogen atoms, within a 4-Å cut-off distance [1].

Results and discussion

In this work, we compare a thermophilic plastocyanin (TPc) with a less thermophilic counterpart (MPc) to identify structural features bearing upon their different thermal resistances. Specifically, we focus on the oxidised species of both plastocyanins, wherein the highest dissimilarity in thermal resistance is found [20].

Primary sequence analysis

Although the two proteins have a substantial sequence identity (63%), a detailed analysis (Fig. 2) reveals that most differences are concentrated in two sequence stretches. One of them, henceforth termed "region A", has a much lower similarity (43%) (Fig. 2). This sequence stretch contains the strands β_5 and β_6,, the loop L_5 and the α-helix that lies between them (Fig. 1).

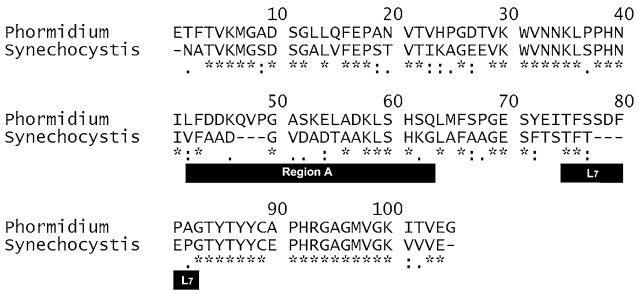

Fig. 2 Sequence alignment of plastocyanins from *Phormidium* and *Synechocystis*. Residues that are identical in both proteins are marked with *asterisks*, those that share similar polarity, charge or aromaticity are tagged with *colons* and residues showing a weak analogy are labelled with a *single dot*. *Black bars* locate sequence stretches of specific interest in this work. The alignment was performed with the program ClustalW2 [63] and edited with the program BioEdit [64]

Structural genomics analyses [15] suggest that thermophilic proteins tend to increase the compactness of regions that are already compact in mesophilic proteins while keeping the full functionality of flexible regions. This is achieved, for instance, by reducing or eliminating loops [4]. In contrast, compared with its mesophilic orthologue, TPc shows two insertions at loop regions. The first locates at region A with respect to the corresponding area in MPc. The second insertion is at loop L_7, opposite region A.

Differences in the primary sequence are also found in strands β_1 and β_2. Among them, the presence of a Phe residue in TPc instead of Ala at position 3 could play an important role in thermal resistance, as has been noted already [20]. However, an aromatic residue in this position is found in other plastocyanins from non-thermophilic species such as members of *Nostoc*. Thus, this difference might not be enough to explain the higher thermal stability of TPc. In summary, sequence alignments suggest that region A may be responsible for the different stabilities of the two proteins and, in the ensuing analysis, particular attention will be paid to it.

Molecular dynamics simulations and conformational sampling convergence

To investigate the thermal stability of the proteins, we ran MD simulations at temperatures ranging from 298 to 498 K. As pointed out in previous studies [20], oxidised species show large differences in their melting points, whereas the reduced species do not. In addition, despite a Cu^{I} force field parameterisation existing [52], it is not consistent with the parameters developed for Cu^{II} nor for the AMBER force field. Hence, we decided to study the oxidised species, where the differences in thermostability are more significant. We then calculated several properties that might be related to the thermal stability of the protein.

MD simulations were performed at 298 K, using the X-ray structure of TPc [37], the NMR structure of MPc [38] and the MPc structure obtained from the X-ray coordinates of a triple mutant [41] as starting coordinates. Different protocols and parameters were also tested in parallel calculations, which converged to similar conformational ensembles and provided nearly identical structural properties (see, for example, Figs. S1–S3).

Principal component analysis (see "Methods") was also carried out to assess the reliability of the calculations. The eight top eigenvectors of the covariance matrix were calculated from the Cartesian displacements from the average position. At room temperature, the overlaps of the eight highest-amplitude eigenvectors from the first half with the full set of eigenvectors from the second half of the

trajectory (Fig. S4) are 0.46 for TPc and 0.54 for MPc, indicating that the simulation time is long enough to obtain a significant sampling of the system [49].

Figure 3 shows the root mean square deviations (RMSDs) of the trajectories for the different simulations. It is worth noting that all simulations show similar behaviour and stability, except those at 498 K, which are less stable. The RMSDs at 298 K with respect to the initial coordinates, calculated on backbone atoms of TPc and MPc, are 0.97 and 1.89 Å, respectively, thereby indicating that the structure of the two proteins did not change substantially along the trajectories (see also Table S2). When calculated on all the atoms of TPc, the RMSDs are similar to those reported in previous plastocyanin simulations (1.8 Å) [51] and therefore are indicative of reproducibility. However, MPc shows a high all-atom RMSD (2.6 Å; Table S2), which might be due to uncertainties in the NMR model regarding the copper coordination sphere. Indeed, the RMSD relative to turns and loops of the copper environment, where the higher uncertainties are located [38], is 4.1 Å, whereas for the rest of the protein it is 2.0 Å. Other parameters, such as the root mean square fluctuations and the gyration radius (Table S2) show small fluctuations, suggesting that stability of the simulations was obtained except for the case of 498 K, where the proteins move considerably, as could be expected.

Figure 3 reveals substantial differences between TPc and MPc at 498 K. MPc shows two sharp jumps in the RMSD values, followed by stable periods, whereas TPc shows a continuous drift of its structure during the calculations. A careful analysis (Fig. S5) indicates that sudden rearrangements in region A are responsible for the RMSD

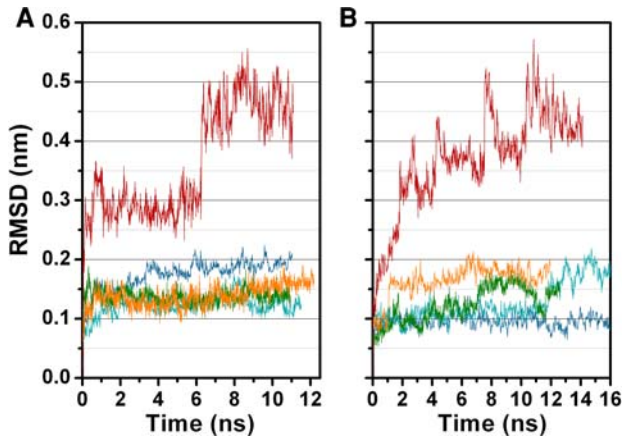

Fig. 3 Root mean square deviation (*RMSD*) with respect to the experimental structure: *A* plastocyanin from the mesophilic cyanobacterium *Synechocystis* sp. PCC 6803 (MPc); *B* plastocyanin from the thermophilic cyanobacterium *Phormidium laminosum* (TPc). Runs at 298 K runs are in *blue*, runs at 323–333 K are in *cyan*, runs at 338 K are in *green*, runs at 354 K are in *orange* and runs at 498 K are in *red*

discontinuities of MPc (Fig. 4). On the other hand, at 498 K TPc undergoes significant and continuous changes all over the structure (Fig. 4). It is worth noting that in MPc region A suffers a first rearrangement and, later, it flips upwards, thereby uncovering the nearby flank of the β-barrel. On the other hand, region A of TPc is not able to flip despite it undergoing substantial changes as it maintains its interactions with loop L_7. As we will show later, these interactions involve hydrophobic residues at both neighbouring regions, which are not conserved in MPc.

Proteins' flexibility

The root mean square fluctuation per residue features maxima located in turns and loops connecting β-strands. In agreement with previous MD studies [20], the largest fluctuations are always found (at all temperatures; not shown) in the sequence stretches defined by residues 10–12 and 50–52 for TPc and 10–12 and 46–48 for MPc, which correspond to loops L_1 and L_5 (Fig. 5). Loop L_1 is part of the north face of the protein, close to the copper atom, whereas loop L_5 belongs to region A. Both locate at the rims of the β-sandwich, where residues show less hydrogen bonding with amino acids far apart in the protein sequence than those in the middle of the two β-sheets (data not shown). Our calculations show that region A is largely affected by temperature. Even at room temperature, this is the part of the protein with the highest mobility

According to the literature, a higher secondary-structure content might contribute to an increase in thermal stability [53]. Figure 6 shows secondary-structure timelines of TPc and MPc trajectories. As already reported by Feio et al. [20], strands β_1 and β_2 are slightly less stable in MPc than in TPc at room temperature. Strands β_5 and β_6 are also less stable in MPc than in TPc (see also Table S3). The difference between the previous and the present simulations can be attributed to the longer time scale used in this work.

At 338 K, TPc loses most of these strands and becomes more similar to MPc at room temperature (Fig. 6). However, this change represents a minor perturbation of the overall topology. In fact, the RMSDs calculated on backbone atoms of the average structures of the trajectories at 298 and 338 K are 0.52 and 1.02 Å for TPc and MPc, respectively.

As expected, changes in secondary structure are more substantial at 498 K than at lower temperatures. However, β-sheets are much more affected by high temperatures in TPc than in MPc (Fig. 6). At room temperature, the β-sheet content is larger in TPc; however, the opposite happens at 498 K (see Table S3 for data regarding β_5 and β_6), probably because of the stronger interaction between the β-barrel and region A through L_7. This point will be discussed later.

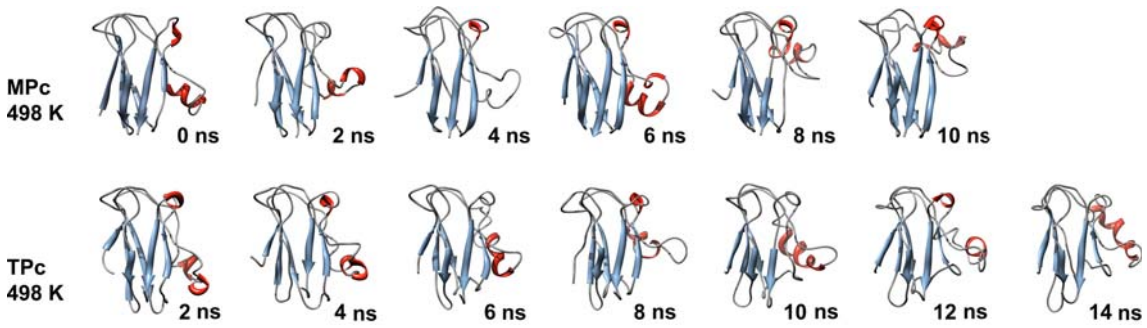

Fig. 4 Structural changes in MPc and TPc along the 498 K trajectories. Robertson diagrams represent snapshots at different times from molecular dynamics calculations at 498 K of MPc (*upper*) and TPc (*lower*). Helix regions are represented in *red* and β-strands are in *cyan*. The images were generated with the UCSF Chimera package [65]. Proteins are oriented with their electrostatic patches (region A) pointing to the *right*

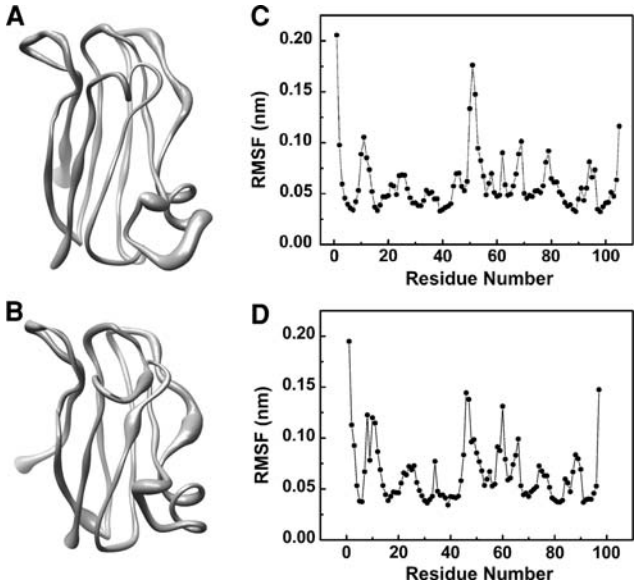

Fig. 5 Calculated root mean square fluctuations (*RMSF*) at 298 K. **a**, **c** The regions with higher mobility of TPc. **b**, **d** The data corresponding to MPc. Proteins are oriented with their electrostatic patches (region A) pointing to the *right*

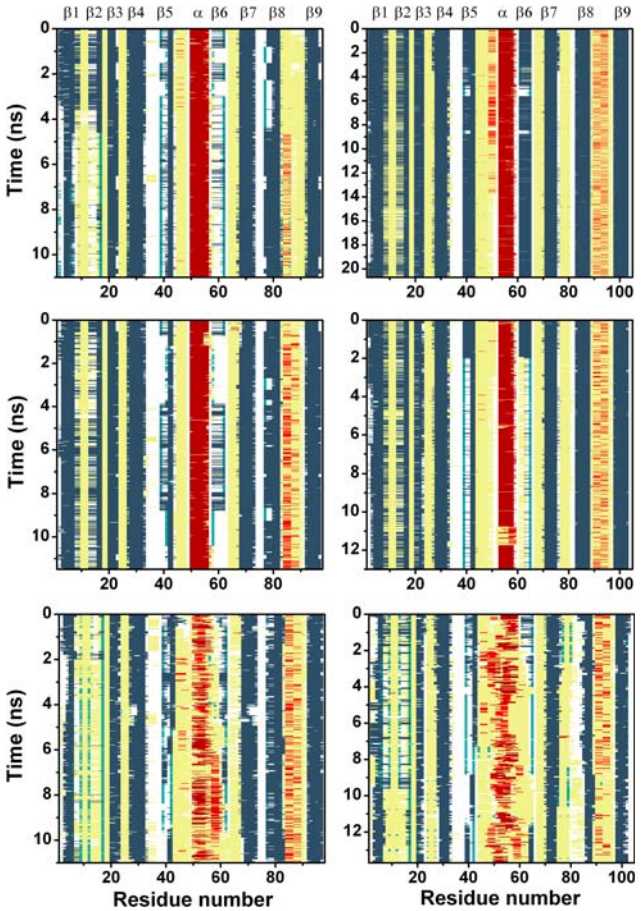

Fig. 6 Time evolution of secondary structures of the proteins. The *left column* and the *right column* show timelines corresponding to trajectories of MPc and TPc, respectively: *top row* 298 K, *middle row* 338 K, *bottom row* 498 K. *Blue* extended β-strand, *cyan* β-bridge, *red* 3_{10} helix, *dark red* α-helix, *orange* π-helix, *yellow* turn, *white* coil. Calculations were performed with the program VMD [62]

Hydrogen bonding

Hydrogen bonding moulds the secondary-structure elements, and it is generally assumed to stabilise protein structures [13, 14, 56]; however, its particular role in thermostability remains controversial [15]. In the present case, the two proteins are characterised by a similar number of intramolecular hydrogen bonds. In fact, in TPc the total number of bonds (normalised to the total number of residues) is only 1.1 times larger than in MPc (0.67 ± 0.04 and 0.59 ± 0.04, respectively).

As observed for other properties, differences in hydrogen bonding between the two proteins are larger within region A. Indeed, at room temperature, TPc features 4.8 ± 0.8 hydrogen bonds between strands β_5 and β_6, whereas MPc only displays 3.2 ± 0.8 hydrogen bonds in this region (Table 2, Fig. 7). Such a difference could be attributed to the unusual presence of a Gly (G59) within strand β_6 of MPc, which causes a mismatch in the hydrogen-bonding pattern between the two strands. At higher

Table 2 Occupancies of hydrogen bonds between strands β_5 and β_6 of TPc and MPc at 298 K

TPc		MPc	
Residues	Occupancy	Occupancy	Residues
HN_{H39}–O_{M65}	0.86 ± 0.35	0.92 ± 0.27	HN_{H39}–O_{A62}
O_{H39}–HN_{M65}	0.85 ± 0.36	0.72 ± 0.45	O_{H39}–HN_{A62}
HD_{N40}–O_{S62}	0.88 ± 0.32	0.34 ± 0.47	HD_{N40}–O_{G60}
O_{I41}–HN_{H61}	0.98 ± 0.12	0.27 ± 0.44	O_{I41}–HN_{H58}
HN_{I41}–O_{H61}	0.33 ± 0.47	–	HN_{I41}–O_{H58}
HN_{F43}–O_{S60}	0.89 ± 0.32	0.97 ± 0.16	HN_{F43}–O_{S57}
Total	4.78 ± 0.87	3.22 ± 0.77	Total

Errors are given as standard deviations; 1 represents full occupancy

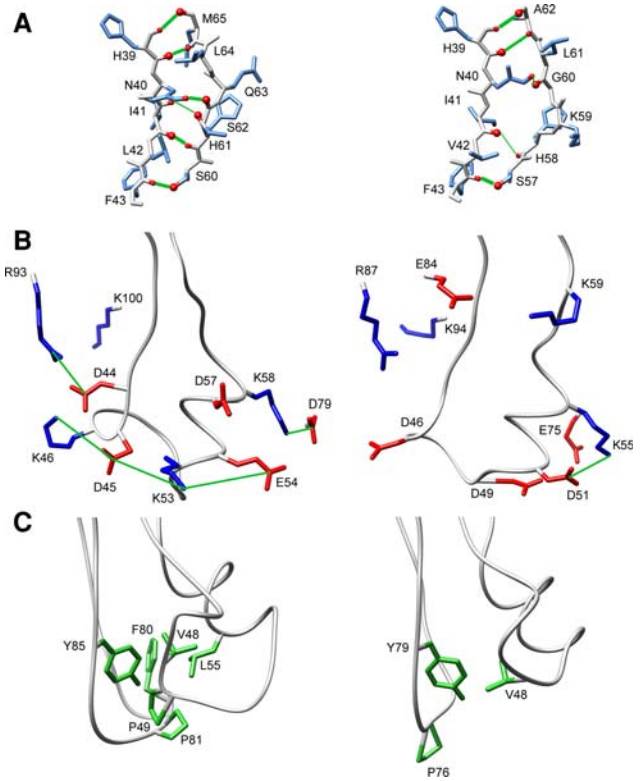

Fig. 7 Different interactions within region A of TPc (*left*) and MPc (*right*) at 298 K. **a** Hydrogen bonds between the fifth and sixth strands. The backbone trace is in *grey*, side chains are in *blue*, hydrogen and oxygen atoms involved in the hydrogen bonds are in *red* and hydrogen bonds are represented by *continuous green lines* (line width proportional to the hydrogen-bond occupancy (see Table 2). **b** Attractive ionic interactions with occupancies higher than 60% at 498 K. Residues of other regions close to region A are also shown. The backbone trace is in *grey*, negative side chains are represented as *red sticks* and positive side chains are represented as *blue sticks*. **c** Hydrophobic cluster formed by residues located in the loops between the fifth and sixth strands and the seventh and eighth strands. The backbone trace is in *grey* and the side chains of hydrophobic residues are represented as *green sticks*. The UCSF Chimera package was used for these images [65]

temperatures, the hydrogen-bond differences between the two proteins decrease, probably owing to the larger loss of secondary structure in TPc.

Contact orders

Because the two proteins differ in the stability of their β-barrels at high temperatures, we tested if the arrangement of contacts within the two proteins followed the same pattern. Contact orders account for the average differences in the sequence position between contacting residues, and are empirically related to folding rates [53] and thermal stability [15]. Hence, we calculated the contact orders for the structures of TPc and MPc available in the Protein Data Bank [37, 41] and for the MD trajectories, using a 6-Å cut-off from heavy atoms [15]. In the case of MD trajectories, the contact orders of the two experimental structures were 16.2 for TPc and 15.5 for MPc. Such differences can be attributed to the two three-residue insertions in TPc. Considering the values relative to the number of residues in each sequence, the differences between the two proteins are negligible. This outcome contrasts with the general trend in the literature, where a larger contact order in thermophilic proteins is reported [15].

In addition to contact orders, recent structural genomics studies [15] suggest that the stability of thermophilic organisms also relies on how intramolecular contacts are arranged in the native protein. This is usually measured by node connectivities and weighted clustering coefficients. The first accounts for the average number of links or contacts per node. Considering the protein residues as nodes of the interaction network, it indicates how many residues, on average, contact each protein amino acid. The latter assesses the cohesiveness of a network by measuring the connection probability between two nodes (i.e. residues) given that both are linked to a third one (see "Methods" for further information). The two parameters are weighted according to the frequency of contacts along the trajectories and the number of heavy atoms within the cut-off in each contact.

In contrast to structural genomics analyses [15], there are more contacts per residue (larger connectivities) in the starting experimental structure of MPc than in that of TPc: 2,669 and 1,558, respectively. However, the clustering coefficients are larger in TPc, in agreement with the literature [15]. Hence, at least in specific regions, contacts between residues are arranged in a tighter way in TPc than in MPc. These differences largely fade away when the data are averaged along the MD trajectory (Table S4) at room temperature. After averaging, TPc features a higher connectivity in the stretches where insertions are located: residues 45–48 and residues 76–81 (L_5 and L_7, respectively). The insertion at loop L_7 increases the compactness

of its connections without affecting its overall connectivity. This insertion decreases the number of contacts within the β-barrel, but increases the number of contacts with region A. Moreover, although these insertions are not long enough to affect the secondary structure [61], they do increase the contact order and, accordingly, they should at least affect the folding rate.

The arrangements of residue contacts within the two proteins change when temperature increases. The number of contacts decreases in both proteins (Table S4). In TPc, the contact order decreases monotonically with temperature, indicating that native contacts involving residues far in the sequence are preferentially lost. This trend is not seen in MPc, except at 498 K. On the other hand, at high temperatures each residue is in contact with more residues than at room temperature, but in a much looser way: the node connectivity increases, but the total number of contacts decreases. Again, this effect is more pronounced in TPc.

Solvent-accessible area

The ratio between the solvent-accessible areas (SAS) of non-polar and polar hydrogen atoms serves as an indication of thermal stability by valuating the strength of the hydrophobic effect within the protein [54]. It reaches a maximum in the unfolded state, where most of the atoms should be exposed to the solvent, whereas it is lowest in the folded state, in which most hydrophobic residues are buried in the protein core. The two proteins display a similar ratio between non-polar and polar hydrogen atoms (13.3 for TPc and 13.2 for MPc), despite TPc having a slightly lower amount (32%) of non-polar residues than its mesophilic orthologue (39%). The average ratio between non-polar SAS and polar SAS is shown in Table 3. At all temperatures, TPc shows a lower non-polar to polar SAS ratio than MPc. At 498 K, the differences are rather small, but it is worth noting that despite the substantial changes in the β-barrel of TPc, it still shows better solvation properties than MPc. Thus, a larger change of heat capacity in TPc upon unfolding might be expected because of the exposure of buried non-polar groups of the protein that become accessible to water [55]. Remarkably, these differences in the SAS are more pronounced (Table 3) in region A.

Table 3 Average ratios of non-polar and polar solvent-accessible surfaces for the entire proteins and the residues within strands β_5 and β_6 (from residues 39–65 in TPc and from residues 39–62 in MPc)

Temperature (K)	TPc		MPc	
	All residues	Residues 39–65	All residues	Residues 39–62
298	1.11	1.04	1.35	1.44
338	1.16	1.05	1.33	1.49
354	1.16	1.05	1.37	1.35
498	1.32	1.31	1.38	1.53

Hydrophobic interactions

Hydrophobic interactions may play a role in the different behaviour shown by the two proteins at high temperatures. TPc contains two hydrophobic residues in region A (V48 and P49) that, along with L55, F80, P81 and Y85, compose a six-residue hydrophobic bunch. This cluster joins region A to loop L_7 and the near edge of the β-barrel. MPc lacks three of these residues (P49, I55 and F80) and, consequently, does not possess important stabilising hydrophobic interactions within region A (Fig. 7). In addition, this information supports a role of region A in the TPc stability. We remark that other hydrophobic cores, however, are highly conserved between the two proteins.

Salt bridging

Another factor related to the stability of proteins is the presence of salt bridges on their surface. At room temperature, salt bridges do not correlate with protein stability because of the desolvation energetic cost [56, 57]. At high temperatures, however, they may become important because the water dielectric constant is lower than at room temperature [58].

The largest density of charged residues for both proteins is found in region A. In fact, this is the area responsible for electrostatic interactions of plastocyanin with its physiological partners [59]. As can be seen from Table 4 and Figs. 7 and 8, region A in TPc establishes a larger number of salt bridges than in MPc. This finding supports a crucial role of this region in determining the higher thermostability of TPc with respect to MPc. Interestingly, at 498 K, the number of salt bridge interactions seem to be greater, which could be related to the conformational changes shown in Fig. 4, also being favoured by an easier desolvation process at this temperature, as pointed out before.

Conclusions

We have compared molecular simulations of a thermophilic plastocyanin with a less thermophilic counterpart to identify structural features playing a role in their different thermal resistances. Selected structural properties related to thermal stability were compared by performing MD simulations at different temperatures, from 298 to 495 K.

As a general outcome of our calculations, TPc shows some common attributes in thermophilic proteins [1, 16]: a higher number of hydrogen bonds and salt bridges and a

Table 4 Salt bridge occupancies (percent) of TPc and MPc at different temperatures

TPc					MPc				
Bridges	Temperature (K)				Temperature (K)				Bridges
	298	338	354	498	298	338	354	498	
D44–K46	7	1	4	25	62	61	61	58	D51–K55
D45–K46	41	39	38	28	0	0	14	0	D46–K94
K53–D57	0	0	3	23	4	3	4	0	E84–K59
E54–K58	5	2	6	58	5	40	45	36	E84–R87
K53–E54	12	2	1	34	0	0	0	30	K55–E75
D57–K58	0	0	0	22					
D45–K53	10	17	8	0					
D44–R93	25	29	32	6					
K58–D79	45	43	46	63					
D44–D45	5	6	6	3					
E54–D57	0	0	0	11					
E54–D79	0	0	0	13					

The interaction between the carboxylic oxygens of Glu or Asp and the side-chain nitrogens of Lys or Arg is considered as a salt bridge when these are found within the Coulomb radius cut-off of 4 Å

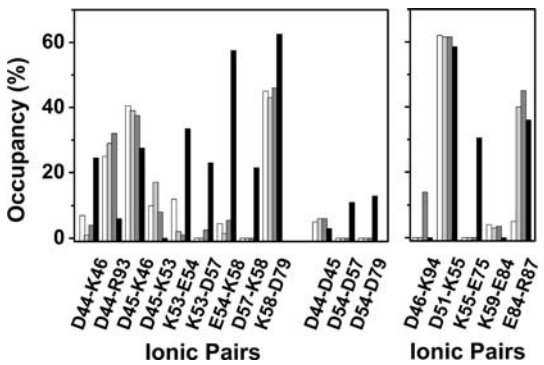

Fig. 8 Occupancies of electrostatic contacts of TPc (*left panel*) and MPc (*right panel*) at four different temperatures (298 K, *white*; 338 K, *grey*; 354 K, *light grey*; 498 K, *black*). Attractive interactions are grouped on the *left* and repulsive interactions are grouped on the *right*. See also Table 4

better ratio of non-polar and polar SAS, specially at room temperature. However, in agreement with with previous works [1, 14, 16–18], no specific structural properties are responsible for its greater thermal stability. This supports the proposal [60] that a combination of structural factors favours the thermal stability rather than a single key facet. Nevertheless, we reported evidence that specific, neighbour regions in TPc, namely region A and the L_7 loop at the β-barrel, play a fundamental role in its enhanced stability. However, the two plastocyanins studied do not follow some typical thermal resistance patterns. Indeed, (1) the TPc primary sequence is longer than that of MPc owing to the loop insertions, which contrasts with the general rule, and (2) the proportion of charged and uncharged polar residues is the same in both proteins.

We conclude by noting that the proposed role of region A is consistent with some biological facts. Indeed, the relevance of this region seems to be smaller in TPc than in other plastocyanins concerning the long-range interactions between plastocyanins and their physiological partners [22, 59]. Indeed, this part of the protein does not contact cytochrome *f* in the case of TPc, whereas it does contain it when other plastocyanins bind to the haem protein [22]. Thus, the evolution of this region could be affected by two different driving forces: functionality and stability of the copper protein—attractive electrostatic interactions between region A of TPc and cytochrome *f* are impaired to allow charged residues in this region to set up stabilising intramolecular surface salt bridges. Further work is in progress to unveil the specific structural features responsible for the different thermal stabilities of the two proteins.

Acknowledgments The authors thank J. Angulo for critical reading of the manuscript. This work was supported by research grants from the Spanish Ministry of Science and Innovation (BFU2006-01361), the Andalusian Government (BIO-198 and P06-CVI-01713) and Project HPC-EUROPA++ (RII3-CT-2003-506079), with the support of the European Union Research Infrastructure Action under the Sixth Frame Programme "Structuring the European Research Area" programme. F.J.M.-L. is the recipient of an FPI fellowship (BES-2005-10404) from the Spanish Ministry of Science and Innovation.

References

1. Vogt G, Woell S, Argos P (1997) J Mol Biol 269:631–643
2. Argos P, Rossman MG, Grau UM, Zuber H, Frank G, Tratschin JD (1979) Biochemistry 18:5698–5703
3. Russell RJ, Ferguson JM, Hough DW, Danson MJ, Taylor GL (1997) Biochemistry 36:9983–9994
4. Thompson MJ, Eisenberg D (1999) J Mol Biol 290:595–604
5. Matthews BW, Nicholson H, Becktel WJ (1987) Proc Natl Acad Sci USA 84:6663–6667
6. Nicholson H, Becktel WJ, Matthews BW (1988) Nature 336:651–656
7. Yutani K, Ogasahara K, Tsujita T, Sugino Y (1987) Proc Natl Acad Sci USA 84:4441–4444
8. Vlassi M, Cesareni G, Kokkinidis M (1999) J Mol Biol 285:817–827
9. Wigley DB, Clarke AR, Dunn CR, Barstow DA, Atkinson T, Chia WN, Muirhead H, Holbrook JJ (1987) Biochim Biophys Acta Protein Struct Mol Enzymol 916:145–148
10. Yokota K, Satou K, Ohki S (2006) Sci Technol Adv Mater 7:255–262
11. Saelensminde G, Halskau O, Helland R, Willassen NP, Jonassen I (2007) Extremophiles 11:585–596
12. Suhre K, Claverie JM (2003) J Biol Chem 278:17198–17202
13. Querol E, Perez-Pons JA, Mozo-Villarías A (1996) Protein Eng 9:265–271
14. Sadeghi M, Naderi-Manesh H, Zarrabi M, Ranjbar B (2006) Biophys Chem 119:256–270

15. Robinson-Rechavi M, Alibes A, Godzik A (2006) J Mol Biol 356:547–557
16. Kumar S, Nussinov R (2001) Cell Mol Life Sci 58:1216–1233
17. Jaenicke R, Bohm G (1998) Curr Opin Struct Biol 8:738–748
18. Greaves RB, Warwicker J (2007) BMC Struct Biol 7:18
19. Feio MJ, Navarro JA, Teixeira MS, Harrison D, Karlsson BG, De la Rosa MA (2004) Biochemistry 43:14784–14791
20. Feio MJ, Diaz-Quintana A, Navarro JA, De la Rosa MA (2006) Biochemistry 45:4900–4906
21. Freeman HC, Guss JM (2001) In: Messerschmidt A, Huber R, Wieghardt K, Poulos T (eds) Handbook of metalloproteins. Wiley, New York, pp 1153–1169
22. Díaz-Quintana A, Hervás M, Navarro JA, De la Rosa MA (2008) In: Fromme P (ed) Photosynthetic protein complexes: a structural approach. Wiley-VCH, Weinheim, pp 181–200
23. Hervás M, Navarro JA, De la Rosa MA (2003) Acc Chem Res 36:798–805
24. Adman ET (1991) Adv Protein Chem 42:145–197
25. Sykes AG (1991) Struct Bond 75:175–224
26. Lumry R, Eyring H (1954) J Phys Chem 58:110–120
27. Sandberg A, Leckner J, Shi Y, Schwarz FP, Karlsson BG (2002) Biochemistry 41:1060–1069
28. Sandberg A, Harrison DJ, Karlsson BG (2003) Biochemistry 42:10301–10310
29. Milardi D, La Rosa C, Grasso D, Guzzi R, Sportelli L, Fini C (1998) Eur Biophys J 27:273–282
30. Wittung-Stafshede P (2004) Inorg Chem 43:7926–7933
31. Stirpe A, Sportelli L, Guzzi R (2006) Biopolymers 83:487–497
32. Pozdnyakova I, Guidry J, Wittung-Stafshede P (2001) Arch Biochem Biophys 390:146–148
33. Deu E, Kirsch JF (2007) Biochemistry 46:5819–5829
34. Wang JM, Cieplak P, Kollman PA (2000) J Comput Chem 21:1049–1074
35. Comba P, Remenyi R (2002) J Comput Chem 23:697–705
36. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) J Phys Chem 97:10269–10280
37. Bond CS, Bendall DS, Freeman HC, Guss JM, Howe CJ, Wagner MJ, Wilce MCJ (1999) Acta Crystallogr Sect D Biol Crystallogr 55:414–421
38. Bertini I, Ciurli S, Dikiy A, Fernandez CO, Luchinat C, Safarov N, Shumilin S, Vila AJ (2001) J Am Chem Soc 123:2405–2413
39. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909
40. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03. Gaussian, Wallingford
41. Romero A, De la Cerda B, Varela PF, Navarro JA, Hervás M, De la Rosa MA (1998) J Mol Biol 275:327–336
42. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
43. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayyik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER 9. University of California, San Francisco
44. Berendsen HJC, Vanderspoel D, Vandrunen R (1995) Comput Phys Commun 91:43–56
45. Lindahl E, Hess B, van der Spoel D (2001) J Mol Model 7:306–317
46. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (1984) J Chem Phys 81:3684–3690
47. Durán RV, Hervás S, De la Rosa MA, Navarro JA (2005) Biochem Biophys Res Commun 334:170–175
48. Arcangeli C, Bizzarri AR, Cannistraro S (2001) Biophys Chem 90:45–56
49. deGroot BL, vanAalten DMF, Amadei A, Berendsen HJC (1996) Biophys J 71:1707–1713
50. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A (2004) Proc Nat Acad Sci USA 101:3747–3752
51. Plaxco KW, Simons KT, Baker D (1998) J Mol Biol 277:985–994
52. Deeth RJ (2007) Inorg Chem 46:4492–4503
53. Chakravarty S, Varadarajan R (2002) Biochemistry 41:8152–8161
54. Szilagyi A, Zavodszky P (2000) Structure 8:493–504
55. Murphy KP, Gill SJ (1991) J Mol Biol 222:699–709
56. Xiao L, Honig B (1999) J Mol Biol 289:1435–1444
57. Waldburger CD, Schildbach JF, Sauer RT (1995) Nat Struct Biol 2:122–128
58. Archer DG, Wang PM (1990) J Phys Chem Ref Data 19:371–411
59. Díaz-Moreno I, Díaz-Quintana A, De la Rosa MA, Crowley PB, Ubbink M (2005) Biochemistry 44:3176–3183
60. Matthews BW, Weaver LH, Kester WR (1974) J Biol Chem 249:8030–8044
61. Makarov DE, Plaxco KW (2003) Protein Sci 12:17–26
62. Humphrey W, Dalke A, Schulten K (2003) J Mol Graph 14:33–38
63. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Bioinformatics 23:2947–2948
64. Hall T (1999) Nucl Acids Symp Ser 41:95–98
65. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) J Comput Chem 25:1605–1612