Protein Engineering vol.11 no.10 pp.881–890, 1998

Effects of counter-ions and volume on the simulated dynamics of solvated proteins. Application to the activation domain of procarboxypeptidase B

Marc A.Martí-Renom, José M.Mas, Baldomero Oliva, Enrique Querol and Francesc X.Avilés[1]

Institut de Biologia Fonamental and Departament de Bioquímica, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

[1]To whom correspondence should be addressed.
E-mail: fx.aviles@blues.uab.es

Molecular dynamics (MD) simulations of the globular activation domain of porcine procarboxypeptidase B (ADBp) and its isolated α-helix 1 were performed in order to understand the effects of adding salts and using periodic boundary conditions (this being reflected in the box size) along the simulations. α-Helix 1 was chosen because it is the most charged element of the secondary structure within ADBp. Different types of MD simulations with the GROMOS package were performed, studying either the whole activation domain or the isolated α-helix 1 with different water box sizes and counter-ionic shells. The analyses of the trajectories show that simulations of solvated proteins are highly sensitive to the presence of counter-ions and less sensitive to the volume of the water box. The differences in protein potential energies, r.m.s. deviations and radius of gyration between the simulations with and without counter-ions demonstrate that during such studies secondary structures of proteins are more stable when their charges are carefully neutralized. This stresses the need for such a procedure when analysing significantly charged proteins. The results also showed that the enlargement of the water box helps in the stabilization of the system.
Keywords: molecular dynamics/procarboxypeptidase B/pro-segment/salt effects/water box

Introduction

The conformational stability of a protein results from a large array of local and non-local interactions. Their individual contributions to protein stability have been studied experimentally (Fersht and Serrano, 1993; Matthews, 1993, 1995; Villegas *et al*., 1995b; Walter *et al*., 1995; Kohn *et al*., 1997; Nölting G. *et al*., 1997) and theoretically (van Gunsteren and Mark, 1992a; Bodkin and Goodfellow, 1995; Honig and Yang, 1995). In particular, the stability of helical regions of proteins has recently been analysed (Fersht and Serrano, 1993; Muñoz and Serrano, 1997). The statistical and experimental analyses of protein α-helices have revealed that the first turn (N-terminal) is preferentially negatively charged in order to neutralize the helix dipole (Hol *et al*., 1981; Harper and Rose, 1993; Negrete *et al*., 1998). However, this preference in the N-terminal end could not be precisely counterbalanced by a preference for positively charged residues in the C-terminal end (Walter *et al*., 1995). In addition, buffer conditions (salt and pH) can have dramatic effects upon the contribution of ion pairs to stability (Kohn *et al*., 1997). The occurrence of stabilizing or destabilizing electrostatic effects in proteins can be tested by determining the effects of salt concentration upon protein stability. This is based on the common assumption that a high salt concentration screens electrostatic interactions (Kohn *et al*., 1997).

While experimental knowledge is essential to understand the effects of counter-ions on the structure and dynamic properties of proteins in solution, theoretical studies and computer simulations are necessary approaches to supplement the experimental data. One of the most demanding systems for such studies is the highly charged globular proteins lacking disulphide bridges. Among the latter, the pro-segments of metallo-carboxypeptidases can constitute good models for the above studies because of both their monodomain nature and their easy folding, which lead to an intramolecular chaperone function being assigned to them (Villegas *et al*., 1995a,b; Reverter *et al*., 1998).

Proteases and their zymogens can be generally classified according to the prominent functional groups at their active site as Ser, Asp, Cys, Zn proteases and others. Pancreatic (pro)carboxypeptidases belong to the Zn-containing class of proteases. They constitute a protein family whose best known representatives are the pancreatic digestive (pro)carboxypeptidases A and B (Coll *et al*., 1991; Guasch *et al*., 1992; Avilés *et al*., 1993). The activation of procarboxypeptidases occurs in a two-step cascade at the duodenum. First, enterokinase converts the pancreatic trypsinogen to trypsin and, second, trypsin activates procarboxypeptidases by releasing a long (about 95 amino acids) activation segment, also termed pro-segment (Vendrell *et al*., 1990; Avilés *et al*., 1993). These proenzymes and their activation pieces have been used as models for folding studies (Conejero-Lara *et al*., 1991; Villegas *et al*., 1995a,b). We have performed simulation studies using the activation domain of porcine procarboxypeptidase B (ADBp) which corresponds to the 81 residue N-terminal globular part of the pro-segment. The activation domain is attached to the carboxypeptidase B moiety by a 15 residue-connecting segment. Structurally, the activation domain consists of a four-stranded antiparallel β-sheet with two α-helices and one 3_{10} helix packed over its external face, in an antiparallel α/antiparallel β topology, as shown by X-ray crystallography and NMR studies (Coll *et al*., 1991; Vendrell *et al*., 1991). Between the two layers there is a hydrophobic core with 20 non-polar amino acid side-chain and 71% of all residues in the globular part participate in well defined secondary structures. The structure presents two internal salt bridges ($10GluO\varepsilon_2$-$55ArgNH_1$ and $55ArgNH_2$–$36AspO\delta_1$). There are no disulphide bridges in this domain. The topology is very simple and presents short loops, except for the loop connecting β_2 and β_3, which is the contact region over the active site of the enzyme. Figure 1 shows the three-dimensional structure of such an activation domain.

Molecular dynamics (MD) simulation is an approach, still in development (Elber, 1996), for studying the structure, stability and folding determinants of proteins. The production of large and stable trajectories using MD is necessary to obtain

© Oxford University Press

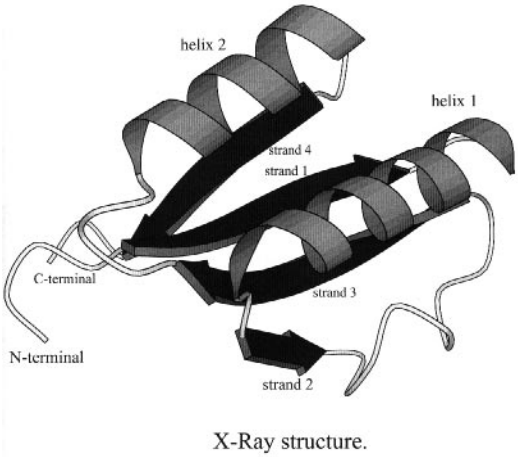

Fig. 1. Three-dimensional structure representation of the activation domain from porcine procarboxypeptidase B (ADBp). The picture shows the β-sheet strands (β_1 from residues 11 to 17, β_2 from 35 to 39, β_3 from 50 to 56 and β_4 from 74 to 79) and the α-helices (α_1 from 19 to 32 and α_2 from 61 to 70). The helix 3_{10} corresponds to residues 42–46. The protein core has been defined as the conjunction of all regular secondary structures.

interpretable results for such studies (Karplus and Petsko, 1990; van Gunsteren and Mark, 1992b; Caflisch and Karplus, 1995; Schaefer *et al*., 1997). In particular, the maintenance of structural stability in proteins is important for free-energy MD calculations, in which the identification of the effects of small changes in the protein is investigated (Daura *et al*., 1996a). Accordingly, the potentially unrealistic behaviour of the MD simulations over longer time-scales becomes a crucial problem (Fox and Kollman, 1996).

In order to analyse the factors which allow us to simulate a more realistic behaviour, we studied here the role of the counter-ions and the dimensions of the box of water molecules surrounding the protein upon its stabilization using the current methodology for performing MD analysis of solvated proteins. The activation domain from procarboxypeptidase B and its first α-helix were chosen as application models. These structures have charged residues placed in α-helix 1 (Asp19, Glu20, Asp22, Glu25 and Glu28), α-helix 2 (Glu65, Asp66 and Glu69), β_1 strand (Lys11 and Arg14), β_2 strand (Asp36 and Lys39), β_3 strand (Asp53 and Arg55), β_4 strand (Glu76) and loop regions (Glu8, Glu10, Glu18, Arg33, Asp41, Lys47, Lys57, Glu59, Asp60 and Glu72). This is a system with a stability in solution dependent on the environmental ionic conditions (Conejero-Lara *et al*., 1991), probably owing to relationships among charges in its three-dimensional structure (Vendrell *et al*., 1991). In this study we neutralized these charges using counter-ions. The effects of the above factors on the stability of the protein during the MD simulations could be generalized.

Methods

The GROMOS package (van Gunsteren and Berendsen, 1991) was used to perform solvated simulations using SPC/E water molecules (Berendsen, *et al*., 1987) and 37C4 parameters. Some changes in the solute–solvent (SPC water) parameters previously suggested by our group (Daura *et al*., 1996b) were incorporated. Two different protein cases, the ADBp and the α-helix from residues Asp19 to Arg33 of ADBp, were tested. The crystallographic coordinates of ADBp, obtained at 2.3 Å resolution (Coll *et al*., 1991), were used to construct the initial structures of both systems for MD analysis, while the NMR structure (Vendrell *et al*., 1991) was used as the reference structure in some analyses. For the ADBp system, three MD simulations (N300, NI300 and NIB300) were carried out and two for the α-helix 1 system (H300 and HI300). α-Helix 1 was chosen because it is the richest segment in charged residues of the ADBp secondary structures.

In simulations N300, NI300 and NIB300, the initial structure was minimized with 2000 steps of steepest descent (Levitt and Lifson, 1969) without SHAKE (van Gunsteren and Berendsen, 1977) and 300 steps with SHAKE, all of them in vacuum. The protein potential energy decreased, reaching a negative value. The minimized protein was solvated by a rectangular box of water molecules with the PROBOX program (van Gunsteren and Berendsen, 1977). The dimensions of the rectangular boxes were 40.2×44.7×54.4 Å for the reference simulation N300, increased to 41.6×45.1×54.4 Å and 45.6×49.1×58.4 Å for NI300 and NIB300 simulations, respectively, corresponding to increments of 4.41 and 33.76% in volume. Simulation N300 is representative of a non-counterbalanced protein charged system. For simulations NI300 and NIB300 the charges of ADBp were neutralized by seven Cl^- ions and 18 Na^+ ions. The water molecules closest to the charges in the protein structure were replaced by the counter-ions. Each total system was energetically minimized using 500 steps of steepest descent and positional constraints over the protein structure; 1.75 ns of simulation at constant temperature (300 K), constant pressure (1 atm) and neutral pH were run for each system under periodic boundary conditions

For H300 and HI300 simulations of α-helix 1 from ADBp, the initial structure was minimized in vacuum with 1000 steps of steepest descent without SHAKE. The helix was solvated by placing it in the centre of a rectangular box of water molecules. The dimensions of the boxes were 44.0×44.0×44.0 Å for both simulations. For simulation HI300, the charges of the system were neutralized by five Na^+ counter-ions placed nearby the charged residues of the helix. For simulation H300, the system was not neutralized. Each total system was energetically minimized using 300 steps of steepest descent and positional constraints on the helix conformation; 1 ns of MD simulation at constant temperature (300 K) and pressure (1 atm) and neutral pH (uncharged titratable groups) were run for the two systems under periodic boundary conditions.

Table I shows some of the technical aspects of the simulations performed. The different systems contain 2765–3984 water molecules per protein molecule, in boxes of various sizes ranging from 84 712 to 131 927 $Å^3$. Therefore, the ion concentration ranged from 0.300 to 0.088 M and the protein concentration from 0.020 to 0.013 M. The ionic strength varied from 0.238 to 1.200 M if both the counter-ions and protein-charged groups are taken into account. The charged state of acid and basic residues was fixed to those corresponding to an initial neutral pH (7.0). A 12 Å cut-off was used in the analysis for the electrostatic interactions.

Six physical properties of the system throughout the resulting simulations were analysed. Two energetically dependent properties (potential energy and hydrogen bonding net) and three structural properties (r.m.s., temperature *B* factors and radius of gyration) were used to describe the behaviour of the systems. The difference in the protein potential energy with respect to the charged system (simulations N300 and H300) was used to determine the relative stability of the protein; on the other hand, the hydrogen bond net was used to characterize the

Table I. Technical aspects of the simulations performed

Simulation	Time (ns)	No. of water molecules	Volume of water box (Å^3)	No. of ions	Ion concentration (M)	Ionic strength[a] (M)
N300	1.75	2878	98866	–	–	1.017
NI300	1.75	2853	99668	18 Na^+/7 Cl^-	0.300/0.117	1.200
NIB300	1.75	3984	131297	18 Na^+/7 Cl^-	0.226/0.088	0.883
H300	1.00	2770	84712	–	–	0.238
HI300	1.00	2765	85301	5 Na^+	0.097	0.286

[a]Ionic strength calculated using the protein with a net charge of –11e for the ADBp and –5e for isolated α-helix 1.

secondary structure stability and distortions. The root mean square deviation (r.m.s.d.) was analysed in order to characterize the conformational changes of the protein. The MD *B*-factors derived from the matrix of the atomic fluctuations were calculated using the equation $B\text{-factor} = 8\pi^2||\text{r.m.s.}||^2/3$. The analysis of the changes in the radius of gyration provided an additional measure of the structural global changes of the protein structure.

The stabilization of the ions in the different solvated systems was analysed by means of their diffusion coefficients. The diffusion coefficient (D) for a given ion species (Cl^- or Na^+) was calculated using the Einstein equation (McQuarrie, 1976):

$$\langle |r_t - r_0|^2 \rangle = 6Dt$$

and the time series of D for each ion species (extracted in windows of 500 ps) through the simulation was calculated. It was shown that at the end of the NI300 and NIB300 simulations (1.75 ns) the D values become fairly stabilized and the distribution stationary. Remarkable are the differences presented by the D values at the end of both simulations NI300 (0.5 and 1.25 Å^2/ps for Cl^- and Na^+) and NIB300 (2.5 and 1.7 Å^2/ps), the ions moving faster when the box of water is larger. Nevertheless, all values are of the same magnitude as those obtained by Lee (1996) in simulations performed in bulk water.

Results

Potential energy and general properties

The different profiles of the protein potential energy between N300 and NI300 simulations (that is, without and with counter-ions) demonstrate that the protein is energetically more stable after 500 ps for the latter simulation than for the former (see Figure 2a). Therefore, the presence of counter-ions has improved the stabilization of the protein. Even better results were found in the NIB300 simulation, in which the protein shows lower values of energy than in the N300 simulation throughout all of the trajectory. This indicates that the enlargement of the water box stabilizes the system. Taking into account that the simulations were carried out under periodic boundary conditions, the increment of the water box reflects a larger dilution of the system. Hence the system is probably closer to the experimental conditions in which this protein has been previously analysed (Conejero-Lara *et al.*, 1991; Vendrell *et al.*, 1991) when ionic strength neutralizes the charged residues of the protein under infinite solvent conditions.

The effect of counter-ions and of ionic strength for a simpler system was also tested. The energetic analysis of the simulations carried out with the isolated α-helix 1 of ADBp (15 residues length) showed a relatively similar behaviour when compared with the whole protein. Figure 2b presents the protein potential energy differences between the two H300 and HI300 simulations (without and with counter-ions). The protein fragment alone was energetically more stabilized for the HI300 simulation than for the H300 during the first 500 ps. Over longer times, the simulation behaviour was, apparently, moderately reversed. However, it should be mentioned that at long simulation times the α-helix 1 three-dimensional structure was severely affected, in both the absence and presence of counter-ions, as evidenced by the large r.m.s.d. attained (Figure 3e; see next subsection), a fact which complicates the energy analysis. These results suggest that the presence of ions in the HI300 simulation contributes to the initial stabilization of the α-helix 1, although further studies are necessary to make interpretations for larger time-scales.

It should be mentioned that in all of the systems tested the electrostatic energy becomes stabilized after 50 ps, showing a very small variation of the values during the simulation: from 0.78% for N300 (the largest) to 0.58% for H300 (the smallest) around the average values. We did not calculate the pK_as of the different ionizable groups during simulation (Antosiewicz *et al.*, 1994). On the other hand, the volume of the water boxes experienced very small changes during simulations: from ±451 to ±468 Å^3 for the different water boxes, corresponding to an approximate average value of 0.45% of the overall volume.

Structural properties of the protein system

The comparison of MD ensembles with the experimental native structure depends upon whether the protein is considered as a whole or in regions. The ensemble of simulation <N300> between 1650 and 1750 ps showed the better agreement with respect to both X-ray and NMR experimental structures (see Table II), although ensemble <NIB300> is slightly better over the whole time-scale (Figure 3a). The r.m.s.d. results for discrete regions of the protein are also shown in Table II. These regions correspond to the different main regular secondary structures and to the protein core (that is, to the conjunction of regular secondary structures). The smallest deviation for α-helix 1 was obtained for the <NIB300> ensemble (the one with counter-ions and the larger water box). In contrast, for α-helix 2 the smallest deviations were found in the <NI300> ensemble. Simulation N300 showed the largest deviations for both α-helices. For the β-sheet structure, ensemble <NIB300> also showed a better agreement than <N300> with experimental structures.

Figure 3 displays the time series of r.m.s.d. of the protein core, α-helices and β-sheet C_α atoms. The results from the three different simulations done with ADBp (see Figure 3a–d) showed that the system did not reach a structural equilibrium for the core of the protein during the entire first 1500 ps. On the other hand, α-helix 1 reached a complete equilibrium in

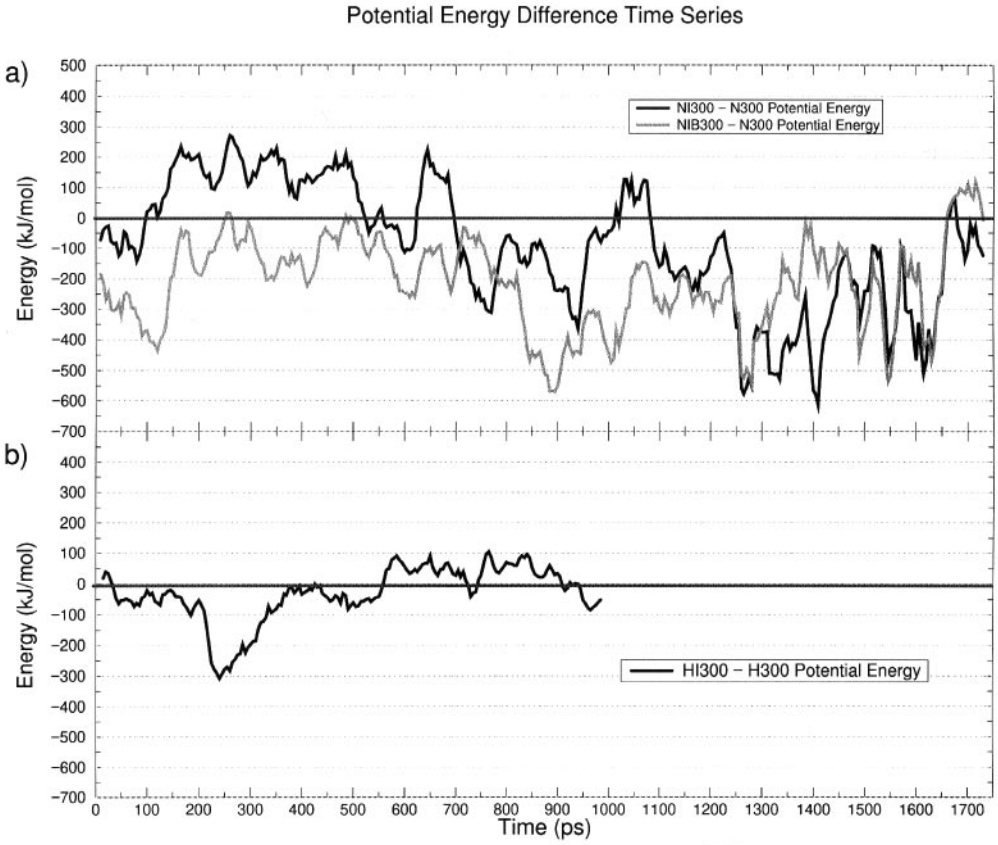

Fig. 2. Protein potential energy–time series during MD simulations: (**a**) for ADBp, simulations N300, NI300 and NIB300; and (**b**) for α-helix 1, simulations H300 and HI300.

about 100 ps in simulations NI300 and NIB300, whereas in simulation N300 it was stable during the first 600 ps and showed a larger increase of r.m.s.d. after this time, losing its initial conformation during the rest of the simulation. α-Helix 2 became equilibrated only for simulation NI300 within the first 300 ps, whereas for simulations N300 and NIB300 this helix was structurally less maintained, reaching structural equilibrium after 1000 ps. The β-sheet does not reach equilibrium in any simulation before the first 1500 ps, showing the largest deviations from the reference conformation.

From the above results it is evident that the whole structure of ADBp has not been well maintained for all simulations. This is probably due to the fact that ADBp is a highly charged protein, with a clear excess of negative charges, so that a large number of counter-ions and an appropiate treatment of the electrostatics are necessary to compensate the system. Consequently, the stabilization of the whole protein can be difficult. Nevertheless, as shown in Figure 3a, the average value of r.m.s.d. for the isolated secondary structures (<SS-r.m.s.d.>, dotted lines) is significantly lower than the r.m.s.d. calculated over the whole core (regions defined as in Figure 1). Figure 3a also shows that simulations NI300 and NIB300 present lower <SS-r.m.s.d.> than does simulation N300, indicating that the secondary structures are better maintained when the system is neutralized by the counter-ions.

On the whole, the results obtained for all simulations showed that α-helix 1 is the most stable part of the entire protein, particularly in the presence of counter-ions and a large water box. This helix is the most charged secondary structure of ADBp. However, for simulations H300 and HI300, that is, in the isolated state, α-helix 1 reached a structural equilibrium within the first 300 ps for H300 simulation and within 700 ps for HI300 simulation. Nevertheless, simulation HI300 (with counter-ions) showed a smaller r.m.s.d. than simulation H300 during the first 550 ps (Figure 3e). The r.m.s.d. for the isolated α-helix 1 structure during the last 100 ps of each simulation, with respect to the crystal structure of the whole proenzyme or the NMR structure of ADBp, are reported in Table II. The results indicate that, during the simulation, the conformation of α-helix 1 suffered important alterations with respect to the original structure in simulations H300 and HI300 (values of backbone atoms over 3 Å). Although simulation HI300 maintained the original conformation during the first 550 ps, the whole structure during the 900–1000 ps period presented a larger r.m.s.d. than did simulation H300. Therefore, the presence of counter-ions did not further stabilize the α-helix 1 structure when analysed alone.

Figure 4 shows the temperature *B*-factors for the side-chain as function of residue number. The results were obtained from an average structure of the last 100 ps of every simulation. The graph indicates that the presence of counter-ions decreases the value of the fluctuation in some charged residues of the protein. This effect was particularly observed at the C-terminal end of the protein, where five charged residues are situated, and also at the amino-terminus. The 'end-effects' are therefore minimized. On the other hand, a further decrease in the temperature *B*-factor of those side-chain atoms is observed when both the counter-ions are present and the water box is increased. A similar trend is observed for the isolated α-helix 1; thus, in HI300 simulations, the counter-ions clearly reduce

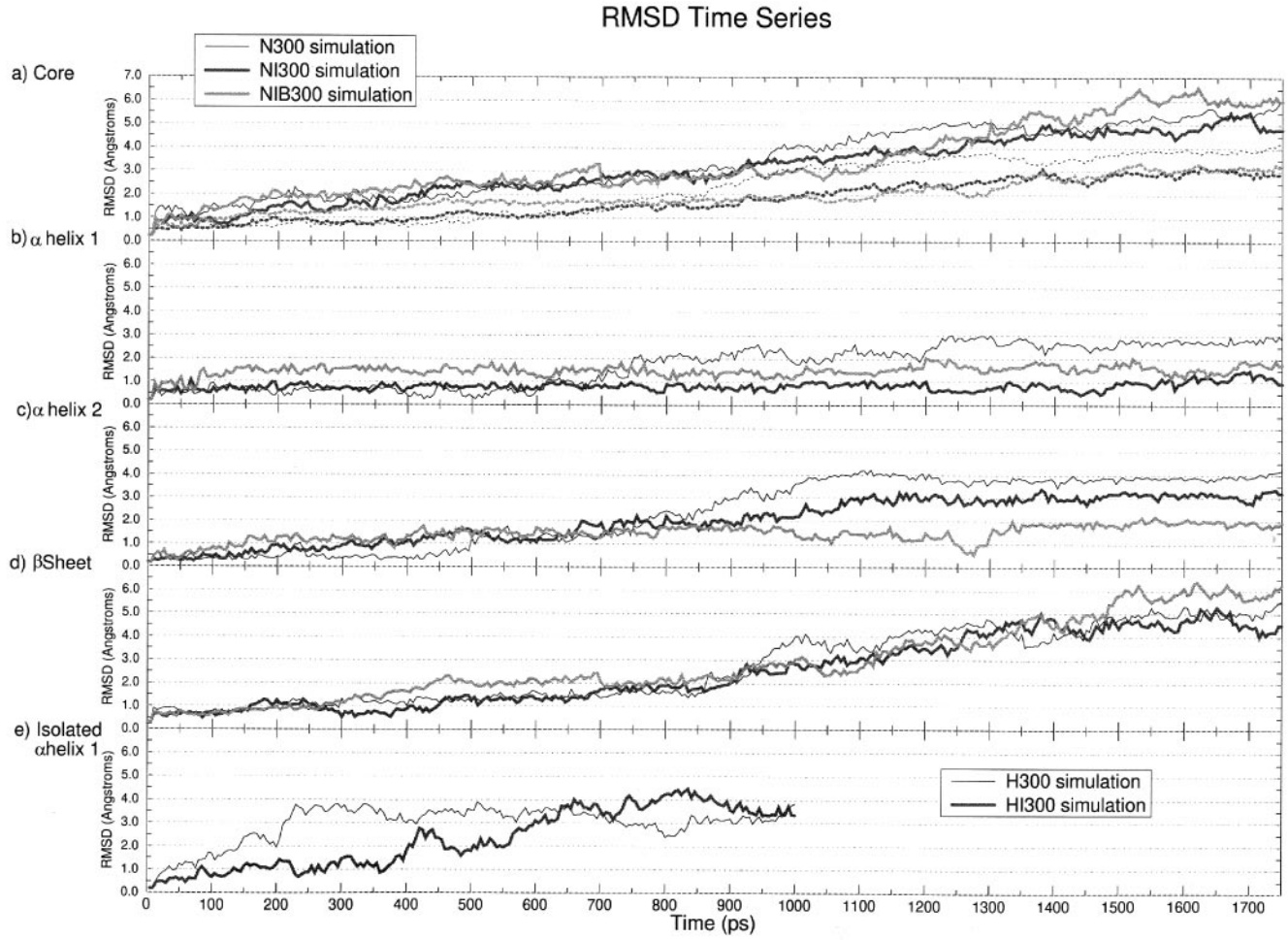

Fig. 3. R.m.s.d.–time series during the MD simulations of ADBp (N300, NI300 and NIB300) and α-helix 1 (H300 and NI300). The graph shows the r.m.s. deviation values in Å for Cα atoms of (**a**) protein core, (**b**) α-helix 1, (**c**) α-helix 2, (**d**) the β-sheet and (**e**) the isolated α-helix 1. Values indicated by a dotted line shown in panel (a) represent the average r.m.s. deviations from the individual secondary structures, α-helix 1, α-helix 2 and β-sheet (<SS-r.m.s.d.>).

Table II. Average and standard deviation of r.m.s.d. values

	All residues		Core residues[a]		α-Helix 1 residues[a]		α-Helix 2 residues[a]		β-Sheet residues[a]	
	All atoms	Backbone	All atoms	Backbone	All atoms	Backbone	All atoms	Backbone	All atoms	Backbone
X-ray vs <NMR>	2.88	1.73	2.37	1.66	1.67	1.66	1.61	0.61	2.49	1.49
<N300> vs X-ray	7.89 ± 0.17	6.72 ± 0.19	6.19 ± 0.17	5.40 ± 0.16	3.72 ± 0.19	2.70 ± 0.11	5.18 ± 0.07	3.66 ± 0.10	5.47 ± 0.21	4.96 ± 0.19
<NI300> vs X-ray	8.68 ± 0.12	7.80 ± 0.14	6.79 ± 0.11	5.91 ± 0.12	2.47 ± 0.18	1.58 ± 0.18	3.26 ± 0.10	1.68 ± 0.09	6.76 ± 0.21	5.86 ± 0.17
<NIB300> vs X-ray	8.09 ± 0.19	7.08 ± 0.21	6.32 ± 0.16	6.06 ± 0.27	2.26 ± 0.20	1.14 ± 0.16	5.04 ± 0.19	2.84 ± 0.15	6.03 ± 0.34	4.60 ± 0.36
<N300> vs <NMR>	6.60 ± 0.15	5.52 ± 0.18	5.73 ± 0.16	4.77 ± 0.18	3.71 ± 0.21	2.51 ± 0.14	5.43 ± 0.05	3.46 ± 0.09	5.37 ± 0.25	4.63 ± 0.21
<NI300> vs <NMR>	8.03 ± 0.10	7.05 ± 0.14	6.38 ± 0.13	5.35 ± 0.11	3.31 ± 0.12	1.96 ± 0.16	3.03 ± 0.08	1.63 ± 0.10	6.45 ± 0.20	5.49 ± 0.16
<NIB300> vs <NMR>	7.50 ± 0.12	6.34 ± 0.20	6.05 ± 0.15	4.49 ± 0.28	3.25 ± 0.22	1.46 ± 0.18	5.00 ± 0.29	3.01 ± 0.18	5.61 ± 0.32	4.10 ± 0.34
<H300> vs X-ray					4.71 ± 0.31	3.10 ± 0.23				
<HI300> vs X-ray					4.60 ± 0.28	3.43 ± 0.15				
<H300> vs <NMR>					4.61 ± 0.21	3.11 ± 0.18				
<HI300> vs <NMR>					4.83 ± 0.16	3.41 ± 0.13				

R.m.s.d. values in angstroms over the last 100 ps of every simulation with respect to the reference structures (X-ray and NMR).
[a]See caption of Figure 1 for region definitions.

the values of the temperature *B*-factors, indicating smaller fluctuations of the charged side-chains.

The radius of gyration through the dynamics gives information on the overall shape behaviour of the protein (or on specific regions). Values of radius of gyration larger than the originals imply an expansion of the structure. The radius of gyration values through MD simulations for the whole protein are reported in Figure 5a. The graph shows that the protein core presents a similar expansion in simulations N300, NI300 and NIB300. The protein shows a continuous expansion not reaching a structural equilibrium. The radius of gyration of α-helix 1 (within ADBp) in simulations N300, NI300 and NIB300

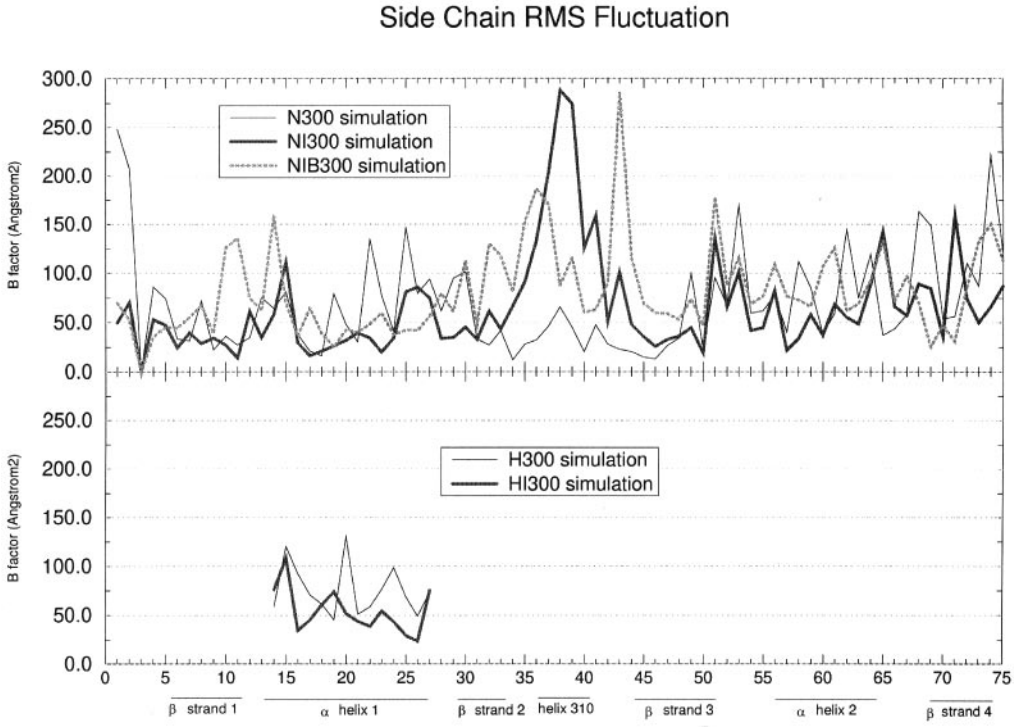

Fig. 4. Representation of the *B*-factors of side-chain residues derived from the MD simulations of ADBp and its α-helix 1. The results are from an averaged structure from the last 100 ps of each simulation. The inset shows the values from simulations done with α-helix 1 alone.

shows that the structure maintains the original shape for the simulations NI300 and NIB300, undergoing a significant expansion after 550 ps in simulation N300 (see Figure 5b). Figure 5c reports the results for the α-helix 2 (in ADBp), where the largest deviation of the radius of gyration is found in the second half of simulation NIB300. Here the differences observed between simulations NI300 and NIB300 are larger than those observed for the same protein region for the r.m.s.d. values, indicating, on the whole, that the distortion of the initial structure suffered by α-helix 2 is larger in NIB300 than in NI300. This is visualized in the ribbon structures displayed in Figure 6c and d. On the other hand, for the β-sheet, the three simulations present values of the radius of gyration similar to the initial values for the first 900 ps and a subsequent 30% increase in the 900–1750 ps time span (Figure 5d).

Finally, in both the H300 and HI300 simulations, the isolated α-helix 1 similarly retains the original structural shape during the first 400 ps. After this time, the helix in simulation HI300 expands (a 15% increase in radius of gyration), whereas in simulation H300 it remains stable until about 750 ps, when the structure begins to expand (Figure 5e).

Protein–protein hydrogen bonds

Table III lists the backbone hydrogen bonds found in the X-ray structure and the frequency of their occurrence in the simulations (when >40%). Simulations N300, NI300 and NIB300 showed that the C-tail end of ADBp loses all hydrogen bonds. The most internal hydrogen bonds of the β-sheet in the crystal structure remained in the β_1 and β_3 strands. In contrast, the β_2 and β_4 strands have lost the major part of their original hydrogen bonds. Table III also shows that the hydrogen bonds defined on α-helix 1 and α-helix 2 were primarily maintained for the NI300 and NIB300 simulations but not for the N300 simulation. Simulation NIB300 showed the largest percentages of conservation of original hydrogen bonds, in general, but the results were better for NI300 than for NIB300 in the case of α-helix 2.

Simulation HI300 presented the largest percentages of original hydrogen bonds for the isolated α-helix 1, particularly in its central region. From these results we can deduce that in this simulation the general conformation of the α-helix was structurally well defined, better than for simulation H300. It is noteworthy that the hydrogen bonds for α-helix 1 in all five simulations (that is, alone and within the whole protein) are differently maintained. The conformation of α-helix 1 within the whole protein (ADBp) was stabilized in its N-terminal region (from Asp19 to Glu28) for all of the three simulations (N300, NI300 and NIB300), particularly in the presence of counter-ions and a large box (NIB300 simulation). However, the α-helix alone (simulations H300 and HI300) was mainly stabilized in its central region (from Ser24 to Thr32), this probably being due to the unfolding of the tail regions because of the lack of adjacent structures. Figure 6 shows for a sketched view of the final structures (averaged over the last 100 ps) of all simulations.

Discussion

Computer simulation of protein structure, dynamics and stability is very useful as a characterization tool and to help in protein redesign. Among MD simulations, one of the challenges is the simulation of the effect of the solvent on protein behaviour (Honig and Yang, 1995). In previous work, we have shown that an adequate representation of solvation effects is required and proposed changes in the solute–solvent parameters (Daura *et al*., 1996b). In such a case, a moderately charged and compact protein, stabilized by disulphides, was used as a model. However, the problem is much more difficult when a highly charged and flexible protein (i.e. lacking disulphides) is used, because an accurate treatment of electro-

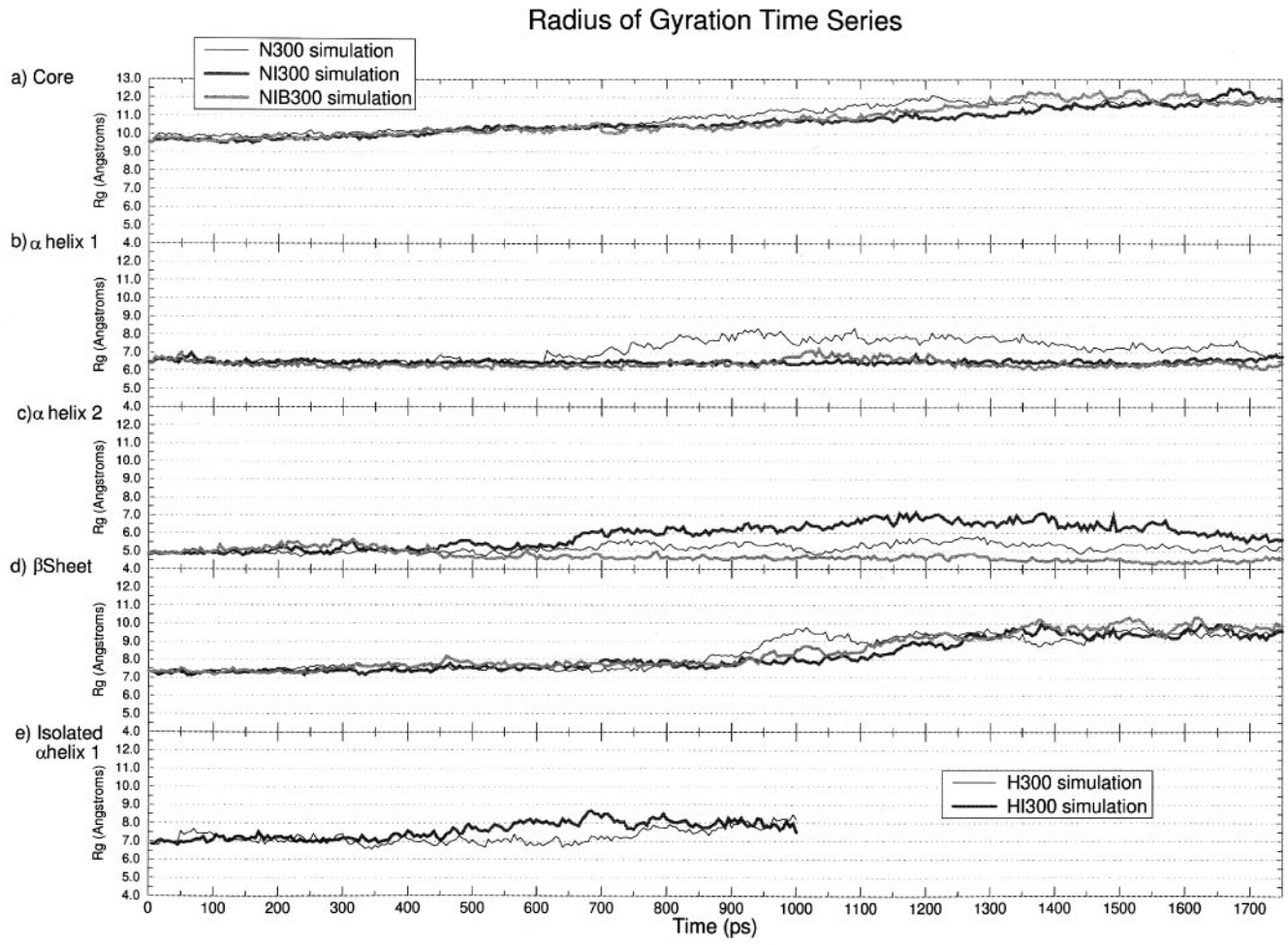

Fig. 5. Radius of gyration–time series during the MD simulations of ADBp (N300, NI300 and NIB300) and α-helix 1 (H300 and NI300). The graph shows the radius of gyration values (Å) for Cα atoms of (**a**) protein core, (**b**) α-helix 1, (**c**) α-helix 2, (**d**) the β-sheet and (**e**) the isolated α-helix 1.

static interactions is essential. This is reflected in the present work, where we have shown that MD analyses of the ADBp protein and its α-helix 1, two highly charged and flexible protein pieces, clearly indicate that the absence/presence of counter-ions and the selected periodic boundary conditions (reflected in the box size) significantly affect the secondary structures and stability of simulated proteins.

The problem of computationally representing highly charged biological macromolecules is exemplified in the treatment given to DNA by different authors. It is well known that DNA conformation stability is modulated by the surrounding environment and is dependent upon the neutralization of its charged groups. Recently, it has been shown that the careful treatment of its electrostatic interactions and addition of counter-ions allow a much better reproduction of the properties of DNA by molecular dynamics simulations (Cheatham and Kollman, 1996; Yang and Pettitt, 1996; Tapia and Velázquez, 1997). In the case of proteins, different electrostatic/solvation models and methodological approaches have been proposed to cope with the above problems (York *et al.*, 1993; Luty *et al.*, 1994; Cheatham *et al.*, 1995; Gilson, 1995; Fox and Kollman, 1996; Hunenberger and van Gunsteren, 1997; Tironi *et al.*, 1997), such as the particle–particle particle–mesh, Ewald and reaction field among others, but the much higher structural diversity and flexibility of proteins make the selection of the proper one difficult. Just an apparently simple problem, such as the appropriate cut-off value for electrostatic interactions, generated great controversy (Gilson, 1995; Moult, 1997).

Experimentally, it has been shown elsewhere that the increase in the concentration of added salts leads to significant variations in the stability of proteins (Kohn *et al.*, 1997), usually to stabilization for non-chaotropic salts, probably owing to changes in the apparent hydrophobic effect and charge screening. A similar conclusion is reached from the present MD simulations: a comparison of r.m.s.d. from the crystallographic ADBp structure suggests that both the use of counter-ions and of a larger water box (simulation NI300) significantly contribute to the maintenance of native secondary structures. The presence of counter-ions helps the stabilization of the internal hydrogen bonds of the α-helices (as seen in the hydrogen bond analysis). This stabilization is correlated with the decrease in fluctuation of the charged side-chain when a counter-ion neutralizes the residue charge (as shown in the *B*-factor analyses). Although the core of the protein was structurally better maintained in simulation NI300, the behaviour was not uniform over the whole structure, the distinct protein secondary structures being maintained differently in the three simulations.

It is also remarkable that a larger water box facilitates the maintenance of the native secondary structures, although the extent of this effect is not so clear and general (for the different

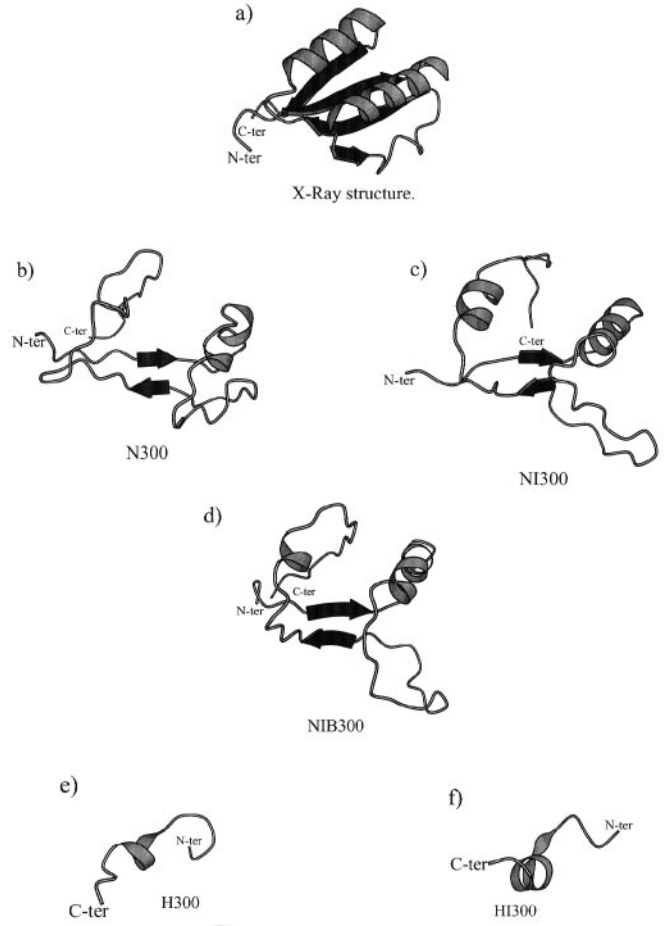

Fig. 6. Representation of the three-dimensional structures taken by ADBp and α-helix 1 in different MD simulations: (**a**) reference X-ray structure; (**b**)–(**f**) average structures from the last 100 ps in simulations N300, NI300, NIB300, H300 and HI300, respectively.

regions of the protein) as in the case of counter-ions. Whether this observation is due to a higher dilution of the protein system (the simulations were performed under periodic boundary conditions, so the size of the water box was important) or to other factors is difficult to judge. It is worth mentioning that the 'formal' concentration of the protein in the water boxes is around 100–150 mg/ml, 2–3 orders of magnitude higher than the concentrations used in experimental analysis of the stability of this protein (or related ones) in solution (Conejero-Lara *et al*., 1991; Villegas *et al*., 1995a,b). We do not know whether this could affect the properties of the protein. Simulations at a much greater 'formal' dilution for large proteins are unfeasible nowadays.

A partial unfolding of ADBp during the simulations was observed when water molecules interacted with the protein core, competing with the original protein–protein hydrogen bonds. A similar overall behaviour has been observed during unfolding simulations done with barnase (Caflisch and Karplus, 1995) and potato carboxypeptidase inhibitor (Martí-Renom *et al.*, 1998). The authors monitored the hydrogen bond formation and breaking during the simulations, concluding that water molecules play a role during the destabilization of the protein by substituting the original donor of the protein–protein hydrogen bonds.

A direct effect of water penetration and partial unfolding is the continuous increment of the radius of gyration of the protein core throughout the simulation (about +20% at the end). The partial unfolding was most clearly observed in simulation N300, which gave rise to a severe loss of helical and β structures. The β sheet (four β-strands) is one of the most severely affected, the central β_1 and β_3 strands being the best maintained. To what extent might its unfolding be related to an imbalance of the electrostatic interactions? Analysis of the initial three-dimensional structure of ADBp indicates that two potential salt bridges are formed between Glu10 and Arg55 (connecting loop 1 with β_3 strand) and between Asp36 and Arg55 (connecting β_2 strand with β_3 strand). On the other hand, two potential repulsive ion pairs can be established between Arg14 and Lys 39 (β_1 and β_2) and Asp53 and Glu56 (β_3 and β_4). It could happen that the penetration of water leads to an increasing imbalance of the electrostatic interactions, which potentiates further unfolding of the β structure and of the whole core. The fact that the potential repulsive interactions are among the internal and the external β strands could facilitate the separation of the latter. In this context, it is worth mentioning recent theoretical work in which evidence has been found that most salt bridges by themselves weakly destabilize the native conformation of proteins (Hendscht and Tidor, 1994). Also remarkable is the observation in the present work that the evolution of the β-sheet structure of ADBp during MD simulations, as judged by the r.m.s.d. and radius of gyration parameters, is not greatly affected by the absence/presence of counter-ions.

In contrast, the evolution of the α-helices is much more dependent of the different MD conditions explored in this work. Interestingly, the slightly less charged helix, α-helix 2, was the most affected and the major part of its original hydrogen bonds were lost, except in simulation NI300. On other hand, α-helix 1, within ADBp, was fairly stable in the simulations with counter-ions, their presence being clearly beneficial. The higher stability of α-helix 1 could also be related to a better neutralization of its dipole because of the occurrence of proper N- and C-cap residues, besides other potential factors such as more efficient hydrophobic and van der Waals internal interactions (Honig and Yang, 1995).

At this stage, it is worth making comparative studies with a highly homologous structure, the activation domain of human procarboxypeptidase A2. This domain has a very similar three-dimensional structure to ADBp for both helices and its environment, as we have shown by X-ray diffraction analyses (Coll *et al.*, 1991; García-Sáez *et al.*, 1997). α-Helices 1 and 2 of such an activation domain from human procarboxypeptidase A2 have been used by Villegas *et al.* (1995b) to study the stabilization of proteins through the redesign of local interactions. The authors demonstrated, using circular dichroism and nuclear magnetic resonance studies on different side-directed mutants, that isolated α-helix 1 in solution better maintains the secondary structure than α-helix 2. This fits with the system studied here, in which α-helix 1 of ADBp has a better helical propensity than α-helix 2 (21.1 and 2.4%, respectively) when calculated using the AGADIR program (Muñoz and Serrano, 1997). The propensity is different if the N- or the C-terminal region of each helix is taken; for α-helix 1, AGADIR gives a 48% propensity for the former and 20% for the latter, which also agree with the results of our simulations. Perhaps the intrinsic helical propensity is as important as is the accurate charge neutralization in the maintenance of protein helices during MD simulations.

Similar conclusions to those achieved for helices within the

Table III. Intramolecular hydrogen bonds in the X-ray structure and in the simulation structures

Donor	Acceptor	Structure	X-ray occurrence[a]	N300 occurrence (%)	NI300 occurrence (%)	NIB300 occurrence (%)	H300 occurrence (%)	HI300 occurrence (%)
10GLU	6PHE	Loop	+	–	–	–		
11LYS	56VAL	β-Sheet 1	–	–	55.9	–		
12VAL	79ILE	β-Sheet 1	+	–	–	–		
13PHE	54PHE	β-Sheet 1	+	83.8	88.7	67.5		
14ARG	76GLU	β-Sheet 1	+	–	–	54.5		
15VAL	52VAL	β-Sheet 1	+	90.5	79.7	86.5		
16ASN	74GLN	β-Sheet 1	+	–	66.2	71.8		
17VAL	50SER	β-Sheet 1	+	83.2	60.6	52.0		
23ILE	19ASP	α-Helix 1	+	–	86.7	91.8	72.9	48.0
24SER	20GLU	α-Helix 1	+	74.8	84.8	87.7	–	61.0
25GLU	21ASN	α-Helix 1	+	73.6	82.8	89.2	–	–
26LEU	22ASP	α-Helix 1	+	44.4	76.0	89.4	–	–
27HIS	23ILE	α-Helix 1	+	64.0	78.0	96.3	–	–
28GLU	24SER	α-Helix 1	+	55.1	81.0	92.4	–	41.0
29LEU	24SER	α-Helix 1	–	–	–	–	51.5	–
29LEU	25GLU	α-Helix 1	+	56.0	86.5	91.3	–	41.2
30ALA	25GLU	α-Helix 1	–	–	–	–	45.8	–
30ALA	26LEU	α-Helix 1	+	68.6	–	94.0	–	82.8
31SER	27HIS	α-Helix 1	–	61.5	72.2	77.8	55.3	70.8
32THR	28GLU	α-Helix 1	–	59.4	–	78.3	43.4	47.6
32THR	29LEU	α-Helix 1	+	–	–	–	–	–
33ARG	28GLU	α-Helix 1	–	–	81.8	–	–	–
33ARG	29LEU	α-Helix 1	+	85.1	–	81.0	–	–
35ILE	29LEU	α-Helix 1	–	–	77.2	–	–	–
36ASP	55ARG	β-Sheet 2	+	–	–	–		
38TRP	53ASP	β-Sheet 2	+	79.7	54.6	–		
39LYS	53ASP	β-Sheet 2	+	–	–	–		
45GLN	42SER	3_{10}-Helix	+	–	–	–		
46ILE	43VAL	3_{10}-Helix	+	–	–	48.4		
49HIS	18GLU	β-Sheet 3	–	57.0	–	41.1		
50SER	17VAL	β-Sheet 3	+	80.3	–	74.6		
52VAL	15VAL	β-Sheet 3	+	90.0	90.4	93.8		
53ASP	39LYS	β-Sheet 3	+	–	–	44.5		
54PHE	13PHE	β-Sheet 3	+	94.2	92.4	88.1		
55ARG	36ASP	β-Sheet 3	+	–	–	–		
56VAL	11LYS	β-Sheet 3	+	81.0	88.4	60.0		
60ASP	57LYS	α-Helix 2	+	41.8	–	–		
61ILE	58ALA	α-Helix 2	+	–	–	–		
64VAL	60ASP	α-Helix 2	+	–	–	–		
64VAL	61ILE	α-Helix 2	–	–	–	40.2		
65GLU	61ILE	α-Helix 2	+	75.7	65.4	–		
66ASP	62LEU	α-Helix 2	+	89.3	–	–		
67PHE	62LEU	α-Helix 2	–	–	40.5	–		
67PHE	63ALA	α-Helix 2	+	63.4	–	–		
68LEU	64VAL	α-Helix 2	+	55.0	–	54.4		
69GLU	65GLU	α-Helix 2	+	45.1	58.0	52.3		
70GLN	66ASP	α-Helix 2	+	–	75.3	40.3		
71ASN	67PHE	α-Helix 2	+	40.9	72.7	44.5		
72GLU	69GLU	α-Helix 2	+	–	–	–		
75LEU	68LEU	α-Helix 2	+	–	–	–		
76GLU	14ARG	β-Sheet 4	+	–	–	51.7		
79ILE	12VAL	β-Sheet 4	+	–	–	–		

[a]Presence (+) or absence (–) of a particular hydrogen bond. Only those hydrogen bonds appearing in at least 40% of the analysed conformations are shown. The values were taken using a cut-off distance of 2.5 Å and cut-off angle of 135° between donor and acceptor.

whole protein domain can be extracted from simulations carried out with the isolated α-helix 1. Although the overall structure of the helix was not well preserved in either simulation, H300 and HI300, simulation HI300 better kept the original hydrogen bonds that held the helix. In this simulation α-helix 1 started to lose its conformation in its C-terminal end; this is the less charged part of this helix (thus being less affected by the counter-ions) and has a lower propensity to form an α-helix (see above). Our analysis also indicates that the α-helix 1 structure is better preserved in MD simulations when inside ADBp than when in the isolated state, a fact which agrees with the occurrence of the nucleation/condensation cooperative effects upon folding determinants in proteins (Honig and Yang, 1995; Nölting *et al.*, 1997).

In summary, although the overall structure of ADBp and its isolated α-helix 1 have not been well maintained during the simulations, the introduction of counter-ions in the system clearly stabilizes the α-helix secondary structures, particularly when they are highly charged. Our analyses indicate that the simulated system is closer to the experimental conditions and better reproduce them when the charges of the protein are carefully neutralized and the box of water tends to be of a larger

size. However, alternative methods to represent electrostatic interactions, such as particle–particle particle mesh, Ewald or reaction fields, among others, are probably required to reproduce accurately the structure and dynamics of highly charged proteins.

Acknowledgements

The authors thank Dr V.Villegas for helpful discussions on the manuscript. This work was supported by grants BIO95-0848 and BIO97-0511 from the CICYT (Ministerio de Educación y Ciencia, Spain) and by the Centre de Referència en Biotecnologia de la Generalitat de Catalunya. The support by C4-CESCA is also gratefully acknowledged. M.A. Martí-Renom is the recipient a fellowship from the Universitat Autònoma de Barcelona (FI-DGR/UAB).

References

Antosiewicz,J. McCammon,J.A. and Gilson,M.K. (1994) *J. Mol. Biol.*, **238**, 415–436.

Avilés,F.X., Vendrell,J., Guasch,A., Coll,M. and Huber,R. (1993) *Eur. J. Biochem.*, **211**, 381–389.

Berendsen,H.J., Griguera,J.R. and Straatsma,T.P. (1987) *J. Phys. Chem.*, **91**, 6269–8271.

Bodkin,M.J. and Goodfellow,J.M. (1995) *Protein Sci.*, **4**, 603–612.

Caflisch,A. and Karplus,M. (1995) *J. Mol. Biol.*, **252**, 672–708.

Cheatham,T.E. and Kollman,P.A. (1996) *J. Mol. Biol.*, **259**, 434–444.

Cheatham,T.E., Miller,J.L., Fox,T., Darden,T.A. and Kollman,P.A. (1995) *J. Am. Chem. Soc.*, **117**, 4193–4194.

Coll,M., Guasch,A., Avilés,F.X. and Huber,R. (1991) *EMBO J.*, **10**, 1–9.

Conejero-Lara,F., Sanchez-Ruiz,J.M., Mateo,P.L., Burgos,F.J., Vendrell,J. and Avilés,F.X. (1991) *Eur. J. Biochem.*, **200**, 663–670.

Daura,X., Hünenberg,P.H., Mark,A.E., Querol,E., Avilés,F.X. and van Gunsteren,W.F. (1996a) *J. Am. Chem. Soc.*, **118**, 6285–6294.

Daura,X., Oliva,B., Querol,E., Avilés,F.X. and Tapia,O. (1996b) *Proteins: Struct. Funct. Genet.*, **25**, 89–103.

Elber,R. (1996) *Curr. Opin. Struct. Biol.*, **6**, 232–235.

Fersht,A.R. and Serrano,L. (1993) *Curr. Opin. Struct. Biol.*, **3**, 75–83.

Fox,T. and Kollman,P.A. (1996) *Proteins*, **25**, 315–334.

García-Sáez,I., Reverter,D., Vendrell,J., Avilés,F.X. and Coll,M. (1997) *EMBO J.*, **16**, 6906–6913.

Gilson M.K. (1995) *Curr. Opin. Struct. Biol.*, **5**, 216–223.

Guasch,A., Coll,M., Avilés,F.X. and Huber,R. (1992) *J. Mol. Biol.*, **224**, 141–157.

Harper,E.T. and Rose,G.D. (1993) *Biochemistry*, **32**, 7605–7609.

Hendsch,Z.S. and Tidor,B. (1994) *Protein Sci.*, **3**, 211–226.

Hol,W.G., Halie,L.M. and Sander,C. (1981) *Nature*, **294**, 532–536.

Honig,B. and Yang,A.-S. (1995) *Adv. Protein Chem.*, **46**, 27–58.

Hunenberger,P.H. and van Gunsteren,W.F. (1997) In van Gunsteren,W.F., Weiner,P.K. and Wilkinson,A.J. (eds), *Computer Simulation of Biomolecular Systems. Theoretical and Experimental Applications*. ESCOM, Leiden, 3–82.

Karplus,M. and Petsko,G.A. (1990) *Nature*, **347**, 631–639.

Kohn,W.D., Kay,C.M. and Hodges,R.S. (1997) *J. Mol. Biol.*, **267**, 1039–1052.

Lee,S.H. (1996) *J. Phys. Chem.*, **100**, 1420–1425.

Levitt,M. and Lifson,S. (1969) *J. Mol. Biol.*, **46**, 269–279.

Luty,B.A., Davis,M.E., Tironi,I.G. and van Gunsteren,W.F. (1994) *Mol. Simul.*, **14**, 11–20.

McQuarrie,D.A. (1976) *Statistical Mechanics*. Harper and Row, New York, Ch. 21.

Martí-Renom,M.A., Stote,R.H., Querol,E., Aviles,F.X. and Karplus,M. (1998) *J. Mol. Biol.*, in press.

Matthews,B.W. (1993) *Annu. Rev. Biochem.*, **62**, 139–160.

Matthews,B.W. (1995) *Adv. Protein Chem.*, **46**, 249–278.

Moult,J. (1997) *Curr. Opin. Struct. Biol.*, **7**, 194–199.

Muñoz,V and Serrano,L. (1997) *Biopolymers*, **41**, 495–509.

Negrete,J.A., Viñuales,Y. and Palau,J. (1998) *Protein Sci.*, **7**, 1368–1379.

Nölting,B., Golbik,R., Neira,J.L., Soler-Gonzalez,A.S., Schreiberg,G. and Ferscht,A. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 826–830.

Reverter,D., Ventura,S., Villegas,V., Vendrell,J. and Aviles,F.X. (1998) *J. Biol. Chem.*, **273**, 3535–3541.

Schaefer,M., Sommer,M. and Karplus,M. (1997) *J. Phys. Chem.*, **101**, 1663–1683.

Tapia,O. and Velázquez,I. (1997) *J. Am. Chem. Soc.*, **119**, 5934–5938.

Tironi,I.G., Luty,B.A. and van Gunsteren,W.F. (1997) *J. Chem. Phys.*, **106**, 6068–6075

van Gunsteren,W.F. and Berendsen,H.J.C. (1977) *Mol. Phys.*, **34**, 1311–1327.

van Gunsteren,W.F. and Berendsen,H.J.C. (1991) *Groningen Molecular Simulations (GROMOS)*. Biomos, Nijenborgh.

van Gunsteren,W.F. and Mark,A.E. (1992a) *J. Mol. Biol.*, **227**, 389–395.

van Gunsteren,W.F. and Mark,A.E. (1992b) *Eur. J. Biochem.*, **204**, 947–961.

Vendrell,J., Cuchillo,C.M. and Aviles,F.X. (1990) *J. Biol. Chem.*, **265**, 6949–6953.

Vendrell,J., Billeter,M., Wider,G., Aviles,F.X. and Wüthrich,K. (1991) *EMBO J.*, **10**, 11–15.

Villegas,V., Azuaga,A., Catasus,L., Reverter,D., Mateo,P.L., Aviles,F.X. and Serrano,L. (1995a) *Biochemistry*, **34**, 15105–15110.

Villegas,V., Viguera,A.R., Aviles,F.X. and Serrano,L. (1995b) *Fold Des.*, **1**, 29–34.

Walter,S., Hubner,B., Hahn,U. and Schmid,F.X. (1995) *J. Mol. Biol.*, **252**, 133–143.

Yang,L. and Pettitt,M. (1996) *J. Phys. Chem.*, **100**, 2564–2566.

York,D.M., Darden,T.A. and Pedersen,L.G. (1993) *J. Phys. Chem.*, **99**, 8345–8348.

Received November 11, 1997; revised May 5, 1998; accepted June 8, 1998