Dynamics of RNase-A and S-Protein: A Molecular Dynamics Simulation of the Transition Toward a Folding Intermediate

Simona Cotesta, Ivano Tavernelli, and Ernesto E. Di Iorio
Institut für Biochemie, Eidgenössische Technische Hochschule, ETH-Hönggerberg, HPM, 8093 Zurich, Switzerland

ABSTRACT The description at atomic level of protein folding is an ambitious goal in biophysics, particularly because of the difficulty in obtaining structural information on unfolded states. Computer simulations can contribute in achieving this goal. Here we report the results of a 10-ns comparative simulation on bovine ribonuclease A and its S-protein, obtained by removal from the native molecule of the first 20 residues, the so-called S-peptide. The atomic trajectories have been analyzed by standard procedures and by applying concepts previously developed for disordered systems. Furthermore, we used a novel approach, described in the preceding paper, to represent graphically the energy landscape of the simulated systems. Relative to RNase-A, the S-protein, while largely maintaining its structural organization, displays an increased structural flexibility, it gains ergodicity and its core loses order, thus indicating that the removal of the S-peptide from ribonuclease A triggers the transition to a folding intermediate with reduced compactness. This finding also has biochemical relevance since the S-protein is recognized as not properly folded by the machinery responsible for the control of the folding quality in the endoplasmic reticulum.

INTRODUCTION

To be fully functional, proteins have to be in a properly folded configuration in which, not only the three-dimensional organization of its atoms is well defined, but also their atomic fluctuations. The intimate relationship between structure, dynamics, and function allows the definition of the thermodynamic stability of a fold and the fine tuning of its functional behavior. There is enough experimental evidence supporting the idea of this close correlation between structural flexibility and function (Di Iorio et al., 1997), but it is not fully understood how it is attained. Unraveling the mechanisms that allow a polypeptide chain to achieve its structural, dynamic, and functional identity demands that we understand in detail how the entire folding process is driven and how its end products are controlled.

Frustration, be it energetic or geometric (Sadoc and Mosseri, 1999), is a physical feature that appears to mingle folding and structural flexibility. Energetic frustration implies the existence of competing nonbonded interactions, whereas geometric frustration comes about when local order cannot be freely propagated throughout the space. It has been suggested that the folding process is driven by a search for minimal frustration (Bryngelson et al., 1995; Wolynes and Eaton, 1999; Wolynes et al., 1996), as there is evidence for a direct correlation between frustration and structural flexibility of a fold (Tavernelli and Di Iorio, 2001).

In this context, a relevant question is whether structural flexibility and frustration phenomena are involved in the in vivo control of protein folding. A particularly interesting case is reglucosylation of the glycan moiety of glycoproteins. As briefly explained below, this is a key step in the control of the folding state of glycoproteins secreted by the endoplasmic reticulum (ER). We have, therefore, decided to compare the dynamic properties of folds that do or do not undergo reglucosylation. In particular, we have investigated bovine RNase-A and its derivative, referred to as S-protein, resulting from the removal of the first 20 residues—the S-peptide—after cleavage with subtilisin (Fig. 1). The glycosylated counterparts of these two proteins, which contain an N-linked glycan moiety at Asn-34 and are known as RNase-B and BS-protein, have been used as model system to investigate experimentally how folding is controlled in the ER (Ritter, 2002). Thanks to this quality control, only correctly folded and assembled glycoproteins can exit the ER. The process, which involves among other enzymes several molecular chaperones, has been reviewed by the Helenius group (Ellgaard et al., 1999), who named it the calnexin/calreticulin cycle. Here we outline its essential features. After its translocation from the cytosol, the nascent glycoprotein chain is cotranslationally glycosylated by the addition of an N-linked glycan to an Asn that lies in a sequence Asn-X-Ser/Thr. The glycan has a branched structure formed by two *N*-acetylglycosamine residues, nine mannose units, and three glucoses. Glucosidase I and II rapidly trim two of the three glucoses, leading to the monoglucosylated derivative. Only as such the glycoprotein can bind either calreticulin or calnexin, two lectins acting as molecular chaperones in the ER. The interaction with one of

Submitted December 17, 2002, and accepted for publication June 26, 2003.

This article is dedicated to the memory of Eraldo Antonini, brilliant scientist, prematurely deceased 20 years ago on March 19, 1983.

Address reprint requests to Ernesto E. Di Iorio, Institut für Biochemie, Eidgenössische Technische Hochschule, ETH-Hönggerberg, HPM, 8093 Zurich, Switzerland. Tel.: +41-1-6323137; Fax: +41-1-6321298; E-mail: diiorio@bc.biol.ethz.ch.

Simona Cotesta's present address is Pharmacia Italia, Computational Sciences, Department of Chemistry, Viale Pasteur 10, 20014 Nerviano, Italy.

Ivano Tavernelli's present address is Ecole Polytechnique Fédérale de Lausanne, Institut de Chimie Moléculaire et Biologique, BCH–LCBC, 1015 Lausanne, Switzerland.

0006-3495/03/10/2633/08 $2.00

FIGURE 1 The top panel depicts a cartoon representation of nicked bovine RNase-A obtained from the coordinates with PDB entry 7rsa. The dark region corresponds to the first 20 residues, referred to as S-peptide, whereas the light gray domain is the so-called S-protein (see text). RNase-A is organized in a helical domain composed by three α-helices (residues 3-13, 24-34, 50-55), one 3/10-helix (56-60), and nine β-sheets (41-48, 61-64, 71-75, 79-87, 90-91, 94-104, 105-113, 114-119, and 121-124). The molecule contains 4 S-S bonds (26-84, 40-95, 58-110, and 65-72). The model in the bottom panel corresponds to the S-protein at the end of our molecular dynamics simulation. Here the secondary structure elements are: two α-helices (24-33, 51-57) and eight β-sheets (41-46, 61-64, 71-75, 79-87, 96-104, 106-112, 115-118, and 120-123, using the RNase-A numbering). Modeling was done with MOLMOL (Koradi et al., 1996).

these chaperones and other enzymes helps the glycoprotein in attaining its fully folded structure. Upon dissociation from the macromolecular complex, the third glucose is also removed from the glycan moiety and the molecule undergoes a control by UDP-glucose:glycoprotein glucosyltransferase (UGGT), which checks whether the glycoprotein is correctly folded. If this is the case, the newly assembled molecule is allowed to enter the secretory pathway, otherwise a single glucose is added back to its glycan moiety. This is interpreted as a signal for the rebinding to calnexin or calreticulin and therefore the protein is given another chance to fold properly. After a few unsuccessful attempts, the glycoprotein is degraded. Clearly, UGGT plays a central role in this quality-control cycle acting as folding sensor.

Extensive experimental evidence shows that bovine RNase-B, as well as the nicked protein obtained by proteolytic cleavage between positions 20 and 21 without dissociation of the S-peptide, are not good substrates for UGGT. Instead, the BS-protein is recognized by the folding sensor as misfolded. Furthermore, it has also been shown that in heterodimers consisting of a folded RNase-BS-protein linked to a misfolded RNase-B via its S-peptide incubated in vitro with UGGT, only glycans linked to the misfolded domain are reglucosylated, indicating that the enzyme recognizes folding defects at the level of individual domains (Ritter and Helenius, 2000). It is also important pointing out that subtilisin treated RNase, the so-called RNase-S or nicked protein, is a very stable complex, which displays full enzymatic activity and exhibits both structural and dynamic features very similar to those of RNase-A (Chakshusmathi et al., 1999 and references therein).

We report here the results of a comparative computational investigation on bovine RNase-A and its counterpart AS-protein. They help in rationalizing the experimental data just discussed and provide new insights on the dynamic properties of folding states that share the overall three-dimensional organization of the native structure, but are characterized by a reduced compactness.

METHODS

Initial coordinates of RNase-A were taken from the Brookhaven Protein Data Bank, entry 7rsa. Molecular dynamics simulations were performed using the GROMOS96 package and the 43A1 force field (van Gunsteren et al., 1996). All acidic residues were deprotonated whereas the basic ones were fully protonated. For the histidines we selected the protonation scheme corresponding to the lowest potential energy in the starting structure, namely with histidines 28, 85, and 99 fully protonated and histidine 12 carrying a hydrogen only on its Nδ atom. A single protein molecule was centered in a truncated octahedron box, of edge 6.88 nm, leaving a minimum distance of 1.25 nm between any solute atom and the box walls. SPC/E water was added in blocks of previous equilibrated 216 molecules, resulting in a fully hydrated system containing 4854 water molecules. Bond lengths were constrained using the SHAKE algorithm (Ryckaert et al., 1977) with a tolerance of 10^{-4} nm. The integration time in the leapfrog scheme was of 2 fs. The pressure was held around 1.013 bar by the weak-coupling approach with isotropic scaling and a relaxation time of 0.5 ps. Solute and solvent were separately coupled to a temperature bath at 310 K, using a relaxation time of 0.1 ps. Nonbonded interactions were computed using the twin range method with cut-off radii of 0.8 nm and 1.4 nm, respectively, for the short- and long-range interactions. For the latter we used a coulombic term with a Poisson-Boltzmann reaction field correction (Tironi et al., 1995) characterized by a relative dielectric permittivity of 54 $kJ^{-1}mol^2nm^{-1}$ and a Debye screening length of zero.

The starting structure of the S-protein was obtained from that of RNase-A. First, we divided the protein into two submolecules (RNase-S and S-peptide). Subsequently the atoms belonging to the S-peptide (residues 1-20) were annihilated by the free-energy perturbation method (van Gunsteren et al., 1996) within 500 ps of molecular dynamics. The resulting structure was used as starting point for a simulation under the same conditions described for RNase-A and as reference configuration during the analysis of the trajectories. In this case, 4777 water molecules surrounded the protein in a truncated octahedral box with an initial edge of 6.86 nm. Other conditions were as previously described (Stella et al., 1999; Tavernelli and Di Iorio, 2001). The total simulation time for each of the two systems has been of 10 ns. Temperature, pressure, total potential energy, radius of gyration (RoG), root mean square deviations (RMSD) from the starting structure, and nonbonded interactions between solute and solvent, as well as between various segments of the protein, were continuously monitored.

To allow a direct comparison between the two proteins, the analyses

carried out on the simulated RNase atomic trajectories were performed on the S-protein atoms set, unless stated otherwise.

We followed the diffusion of the shape and size of the polypeptide chain by determining the distance between its configuration at time t and that at time zero. In addition to the Euclidean distance between atoms, this was done using the more informative D_4 distance (Pliska and Marinari, 1993), as described in the preceding paper (Tavernelli et al., 2003).

Radius of gyration, RMSD, fluctuations along the first two principal components (Amadei et al., 1996) and D_4 distance were used to obtain a two-dimensional representation of the free-energy landscape visited during the simulation, as discussed in detail in the accompanying paper (Tavernelli et al., 2003).

To monitor the spatial organization within the core of the protein, we have computed the time series of the so-called Bond Orientation Order Parameter (BOOP) and their time correlation function, employing a previously reported procedure (Tavernelli and Di Iorio, 2001) with the following modifications. For the definition of the core, we have selected all the atoms with exposed solvent surface equal to zero and belonging to hydrophobic residues present in both simulated proteins. The bond orientation order parameter was computed considering all the pairs formed by each core atom and its neighbor particles at a distance R_{bond} linked to it by a bonded or nonbonded interaction.

The solvent exposed surface was calculated using a probe radius of 0.14 nm. The time series of the Voronoi volume were calculated using the software package developed by Gerstein (Gerstein et al., 1995). The Voronoi method consists in the determination of a Voronoi polyhedron around an atom. To construct it, one defines as many lines as needed to connect the central atom to all of its neighbors within a certain cut-off distance. Subsequently, one defines planes perpendicular to such lines, positioning them at the midpoint between the connected particles. The smallest polyhedron formed by the intersection of such planes is unique and is called the Voronoi polyhedron. If one or more neighbors around an atom cannot be defined, the constructed polyhedron will remain open. This is what happens with atoms at the protein surface. Hence, the Voronoi volume is calculated only for atoms not exposed to the solvent.

We derived from the atomic trajectories the time series of the Voronoi volumes averaged over all core atoms common to both proteins.

RESULTS AND DISCUSSION

The time series of several global structural parameters computed from the atomic trajectories indicate that, for both proteins, it takes >4 ns to achieve steady-state fluctuations. We report in Figs. 2 and 3 representative data in this respect, referring respectively to RoG and to the D_4 conformational distance from the reference structure. The time series depicted in Fig. 2 show that the RoG of RNase-A, after an initial decrease likely due to an adaptation to the force field, reaches a steady state characterized by moderate and stochastic oscillations around a value of ~1.34 nm. Instead, for the S-protein, the same parameter displays large oscillations, whose distributions present two peaks centered at ~1.47 nm and 1.49 nm (Fig. 2, *insert*). For a globular protein, the radius of gyration provides quantitative information on its compactness, but not for a kidney-like shaped molecule like RNase. Nevertheless, comparing the RoG time evolution of RNase-A with that of the S-protein can pinpoint qualitative differences in the overall packing of the two molecules. The data of Fig. 2 indicate that during our simulations the S-protein tends to relax toward a looser

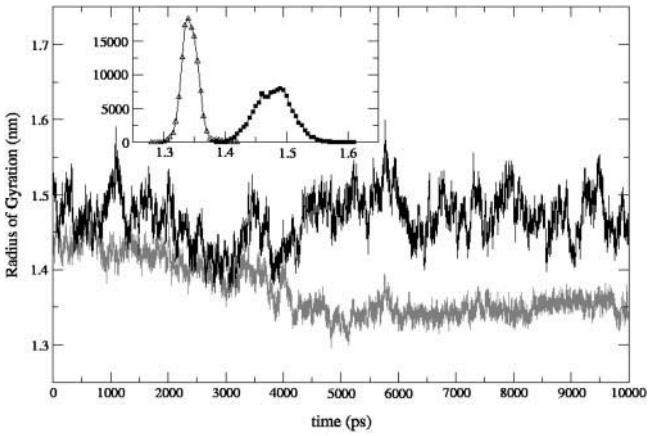

Fig. 2 Time series of the radius of gyration computed from the C_α-atoms of RNase-A (*gray* curve) and for the S-protein (*black* trace). The insert depicts the histograms computed from the time series between 4 ns and 10 ns of RNase-A (△) and of the S-protein (■). Counts are plotted versus RoG in nm. As stated in the Methods section, the analysis of the RNase-A trajectory was done neglecting the atoms of the S-peptide.

configuration whereas RNase-A keeps its original compactness. To better describe the diffusion of shape and size of the system under investigation we used the distance D_4 (Pliska and Marinari, 1993; Tavernelli et al., 2003). This parameter measures the conformational distance, based in this specific case on the intramolecular distances between all possible pairs of C_α-atoms in a configuration at time t with the corresponding distances in the starting structure. One expects D_4 to diffuse toward a maximum value and thereafter to remain practically steady if the system, after having explored the accessible conformational space, finds a minimum (global or local) in the landscape. As shown in Fig. 3, this kind of behavior applies quite well to RNase-A, but not

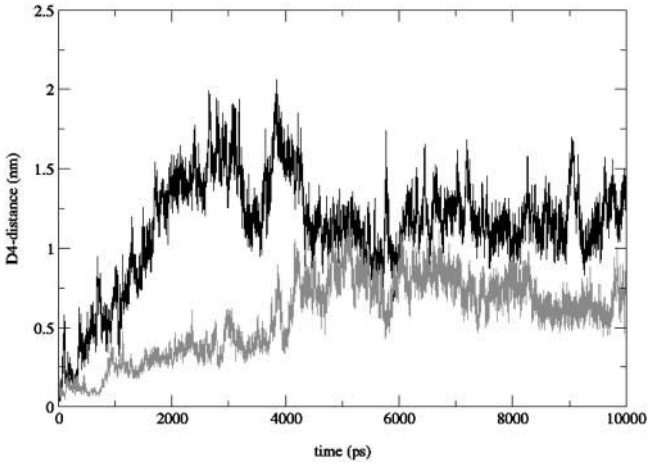

FIGURE 3 Time series of the conformational distance D_4 from the starting structure for RNase-A, neglecting the atoms of the S-peptide (*gray curve*), and for the S-protein (*black* trace).

to the S-protein. In the latter system, the D_4 distance displays first a sharp rise, immediately followed by a decline that leads to a relatively unstable profile, characterized by values consistently higher than those of RNase-A and possibly by a positive slope. This confirms that the S-protein loses compactness, as revealed by the time evolution of the RoG (Fig. 2). In addition, the data of Fig. 3 show that the S-protein has access to a larger local conformational space compared to RNase-A, most likely not completely visited during the simulation.

Fig. 4 displays the time average of the C_α-RMSD from the starting structure for RNase-A and for the S-protein (*top*), along with the difference between the two profiles (*bottom*). RMSDs are notoriously correlated with the flexibility of a molecule. As expected for a system that loses compactness, the S-protein displays larger RMSD than RNase-A, thus implying a greater structural flexibility. The profile of the RMSD difference does not correlate with the secondary structural organization of RNase-A, the largest differences being observed in the region around the glycosylation site (Asn-34 in bovine RNase-B) and for the two segments Ala-64–Gln-74 and Cys-84–Ala-96. For this reason, we have analyzed the time series of the number of residues belonging to a particular secondary structure (data not shown). As expected, RNase-A keeps its secondary structural organization throughout the simulation. The S-protein undergoes an initial adaptation to the absence of the S-peptide, which causes a temporary loss of α-helical structure, partially counterbalanced by an increase in the number of residues involved in 3/10-helix segments. However, within the first 2.5 ns of simulation, the number of residues belonging to α-helical segments returns to the starting value, hence implying that the increased structural mobility of the S-protein is not due to loss of secondary structural organization. This is confirmed by the model representation of the S-protein at the end of the simulation depicted in the bottom panel of Fig.1. The secondary structure is essentially maintained, the major difference between the starting and final structures being a relative rotation of the two lobes of the kidney-shaped protein due to a longitudinal twisting of the molecule.

From the results discussed so far we can state that, compared to RNase-A, the S-protein is structurally more flexible, both around the glycosylation site and away from it. The removal of the S-peptide from native RNase-A causes an increase of the overall volume of the molecule, without influencing significantly its secondary structural organization.

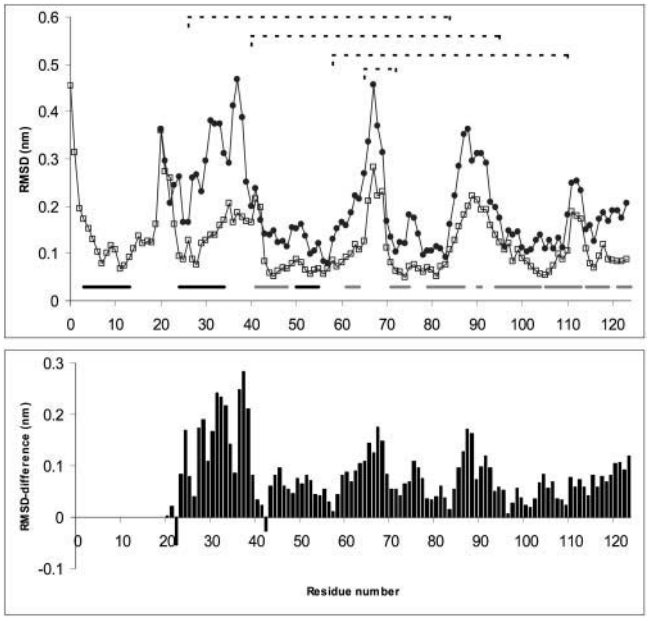

FIGURE 4 Time averages over the entire trajectory of the C_α-RMSD from the starting structure (*top* panel) for RNase-A (□) and for the S-protein (●). The difference between the two profiles is shown in the bottom panel. In the upper graph α-helical and β-sheet segments are shown respectively as black and gray continuous lines below the RMSD profiles. Disulfide bonds are indicated by the dashed lines.

Nonbonded interactions

To better define the picture just depicted, one needs a detailed analysis of the energetic changes arising from the removal of the S-peptide from RNase-A. Adding all the nonbonded interaction terms provides a good estimate of the interaction energy within a protein. For the two systems under investigation, the distributions of the total nonbonded interaction energy are all Gaussian in shape and characterized by the parameters listed in Table 1. Clearly, the S-peptide has an overall stabilizing effect quantifiable in roughly 1100 kJ/mol. RNase-A without the contributions from the S-peptide and the S-protein display comparable total nonbonded interaction energies, as expected for two systems that have the same overall three-dimensional arrangement. The moderate, but significant shift toward less negative interaction energies for the S-protein is in agreement with its reduced overall compactness discussed above.

To better characterize the nature of this change we computed from the time series of the interaction energy between all residue pairs their average values and root mean square deviations, which were used to construct two-dimensional maps. They allow us to find out whether interactions are frustrated and to what extent, an important

TABLE 1 Parameters for the distributions of the total nonbonded interaction energy computed from the entire trajectories

Protein	Peak position (kJ/mol)	Full width at half height (kJ/mol)
RNase-A	−4504	529.5
RNase-A minus S-peptide	−3669	510.0
S-protein	−3414	490.0

element when studying the dynamic properties of proteins since frustration of nonbonded interactions has been shown to correlate with structural flexibility (Tavernelli and Di Iorio, 2001). Energetic frustration arises from the presence of competing interactions, which are degenerate in energy. In principle, one could analyze the time series of all the pairs formed by a given side chain and try to find competing interactions. In practice, this type of approach is not easily applicable because in complex systems like proteins there is a large number of means to compensate for the loss or formation of an interacting pair. Two-dimensional maps of averaged nonbonded interactions are of great help to pinpoint energetic frustration.

In Fig. 5 we report this type of map for RNase-A (*lower right* triangular half) and for the S-protein (*upper right* triangular half). The plot shows that there are not many strong attractive pairs (*black* or *blue dots*). Furthermore, several strong interactions are located near the diagonal region, thus representing local contacts between nearby residues along the sequence. Interactions that stabilize the fold are between residues that are well separated along the sequence and therefore away from the diagonal of the plot. Our systems display few such interactions that in addition are, to a large extent, common to both proteins. One important exception is the pair Asp-83–Lys-104, which is very strongly attractive in the S-protein (*black dot*) and only moderately so in RNase-A (*green dot*). However, this interaction has standard deviations of 98 kJ/mol and 31 kJ/mol respectively for the S-protein and for the native molecule, thus indicating that the removal of the S-peptide not only increases its energy, but also its frustration. This is yet another sign reflecting the looser structure of the S-protein. The map of Fig. 5 also shows the existence of two very strong attractive interactions in the native protein not present in the S-protein. These interactions involve the S-peptide and certainly contribute in stabilizing the fold. The central position occupied by the α-helical segment of the S-peptide within the kidney-shaped molecule certainly locks the system such that the longitudinal twisting observed during the simulation on the S-protein cannot take place (Fig. 1).

The plot in Fig. 5 shows also that the spatial organization of nonbonded interactions for the two proteins is highly similar. Analogously, the same pattern is observed for the two molecules on two-dimensional maps displaying the relative presence of the various contact pairs during the simulation time (data not shown). This again confirms that, within our simulation times, the two systems essentially keep both their secondary structure and overall tertiary arrangement.

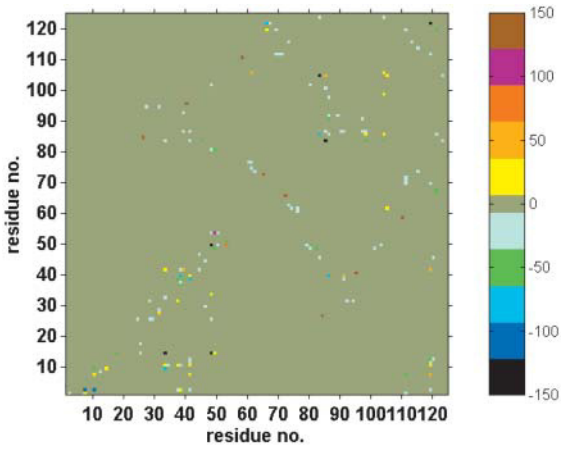

FIGURE 5 Two-dimensional map of average nonbonded interaction energies between residue pairs. The lower right and upper left triangular halves of the plot refer respectively to RNase-A and to the S-protein. For each pair, we considered separately contributions coming from the interaction between the two side chains, between backbone and side chain, and between backbone atoms. Whenever the absolute value of the interaction energy was at least once above 35 kJ/mol, we saved the corresponding time series and calculated time averages and standard deviations over the entire trajectory.

Topological analysis of the hydrophobic core

An important element for the characterization of the dynamic properties of the hydrophobic core of a protein is its geometrical organization (Tavernelli and Di Iorio, 2001). We have analyzed how the average Voronoi volume of core atoms evolves during the simulation and computed the corresponding histograms from the entire trajectory (data not shown). The S-protein displays a broader and more asymmetric distribution compared to RNase-A, thus indicating a drop in the geometrical order of the core when the S-peptide is removed. To quantify this phenomenon we analyzed the autocorrelation of the BOOP computed as reported previously (Tavernelli and Di Iorio, 2001) and

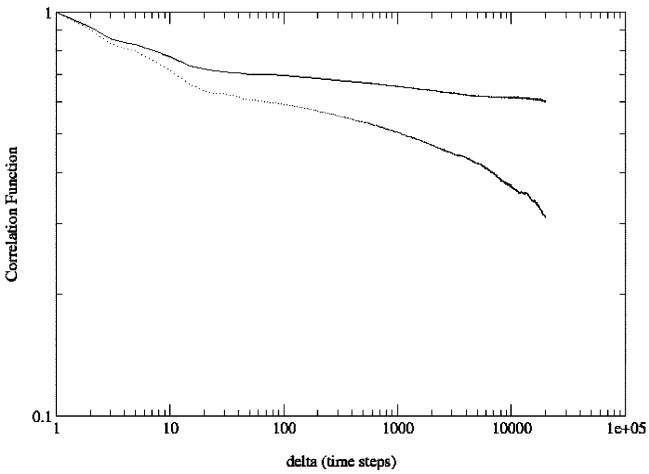

FIGURE 6 Time correlation function of the BOOP. (Tavernelli and Di Iorio, 2001) computed from the atomic trajectories of the core atoms of RNase-A (*continuous line*) and of the S-protein (*dotted line*). Only atoms common to both proteins were considered (see Methods).

considering all first neighbors of each core atom. The results are reported in Fig. 6. For RNase-A, after an initial decay, the curve approaches a level with a persistent correlation greater than zero. Instead, the trace relative to the S-protein displays an additional decay that rapidly reduces its correlation. This additional process reveals a loss of order within the core and the tendency of its atoms to organize themselves in subclusters with relatively independent dynamics. In other words, the interactions among particles in each subcluster tend to dominate over global interactions involving the whole core of the protein. This behavior indicates that the S-protein is diffusing toward a looser state, where the system gains in ergodicity, similarly to what is observed in amorphous systems and in spin glasses (Fischer and Hertz, 1991; Mezard et al., 1987). Ergodicity breaking means that time and ensemble averages do not coincide and is typical of systems with multivalley energy landscapes below a critical temperature, when trapping within local energetic minima takes place. Native proteins display this behavior, but, if the system starts diffusing toward a looser state, long-time correlations decrease rapidly while the volume of the sampled phase space increases (Tavernelli and Di Iorio, 2001). Eventually, this could lead to complete unfolding, which is the most ergodic condition for a protein.

Two-dimensional representation of the free-energy landscape

A protein under proper conditions folds into a closely packed tertiary structure. If we attempt to analyze protein folding from an energetic point of view, we cannot identify a ground, nondegenerate state as one does with simple systems such as single atoms, nuclei, or particles. The ground state of a protein, and in general of a complex system, is highly degenerate and, to account for all possible energy levels, we have to consider a highly multidimensional energy landscape. The only practicable way to represent graphically a system of this kind is to reduce its dimensionality using global parameters, for instance as described in detail in the preceding article (Tavernelli et al., 2003).

Fig. 7 depicts the free-energy contour plots computed from the total nonbonded energy of the solute and the conformational distance D_4. Instead, in Fig. 8 we report the projection of the trajectories on the same graphs. RNase-A (*top panel* of Fig. 7) displays three well-defined minima. The system remains in the leftmost minimum for the first 0.85 ns (*red trace* in the *top graph* of Fig. 8), thereafter it moves to the middle well where it stays for most of the following 3.15 ns (*yellow*), and finally jumps to the rightmost basin where it remains during the last 6 ns of the simulation (*magenta*).

Relative to that of RNase-A, the landscape of the S-protein covers a larger surface of the plot, thus showing that the system explores a greater conformational space. In addition, it displays a single, broad, and rough minimum (Fig. 7). During the first nanosecond there is a transition through substates characterized by relatively high values of the total nonbonded interaction energy (*red* and part of the *yellow trace* in the *bottom* panel of Fig. 8). This can safely be ascribed to the initial adaptation to the dissociation of the S-peptide. Thereafter, the system diffuses to the broad minimum and randomly explores it for the rest of the simulation (*second half* of *yellow* and *magenta* in the *bottom graph* of Fig. 8). The rough structure of the landscape is a sigh of frustration of the S-protein, which is reflected by the frequent jumps of the trajectory between distinct regions of the free-energy landscape (e.g., *yellow* trace).

Although different in the details and less informative, the same overall picture emerges from contour plots computed using global parameters other than the D_4 distance like RMSD, radius of gyration, or fluctuations along principal components (data not shown). The larger conformational volume visited by the S-protein implies that entropic contributions are higher in this protein than in RNase-A, in agreement with its enhanced flexibility.

Concluding remarks

Our computational investigation shows that the removal of the S-peptide from RNase-A triggers the transition to a state that, conserving its three-dimensional arrangement, displays: i), an increased structural flexibility; ii), a looser packing; iii),

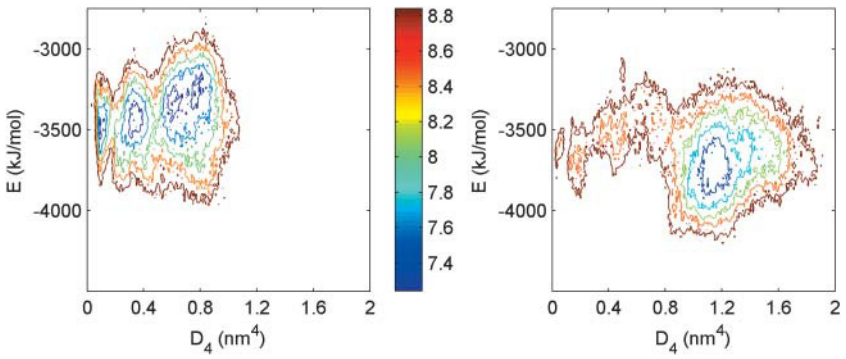

FIGURE 7 Contour plots representing the energy landscape calculated from the total nonbonded energy of the solute and the conformational distance D_4. The top and bottom panels refer respectively to RNase-A, neglecting the atoms of the S-peptide, and to the S-protein. Contour lines refer to the negative natural logarithm of the probability, coded as indicated by the color bar. More details are given in the preceding article (Tavernelli et al., 2003).

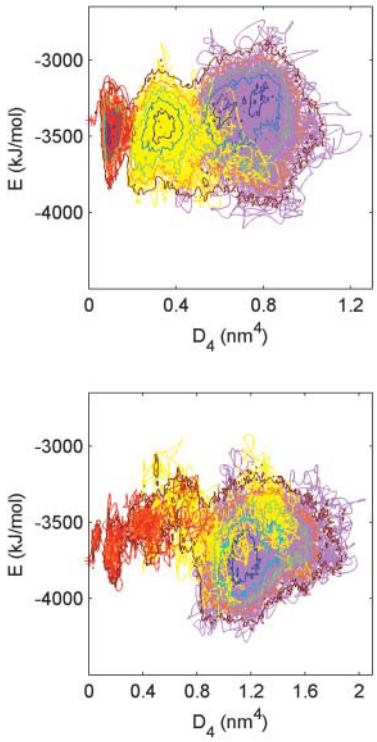

FIGURE 8 Projection of the trajectory on the free-energy contour plots depicted in Fig. 7. The top and bottom panels refer respectively to RNase-A, without considering the contribution from the S-peptide, and to the S-protein. The trajectories have been averaged in blocks of 20, corresponding to 1 ps, and splined. Distinct time frames of the trajectories are plotted using the following color code for RNase-A: red = 0–0.85 ns; yellow = 0.85–4 ns; magenta = 4–10 ns. For the S-protein: red = 0–0.75 ns; yellow = 0.75–2.15 ns; and magenta = 2.15–10 ns.

a lower coherence in the dynamics of the hydrophobic core; and iv), higher frustration and ergodicity. This is in full agreement with the finding that the rate of proton exchange strongly decreases when the S-protein forms a complex with the S-peptide (Rosa and Richards, 1981). Therefore, it is reasonable to conclude that UGGT, the molecular system responsible for the folding quality control of glycoproteins in the ER, senses the structural flexibility of its substrates. We envisage the following picture. When UGGT binds a glycoprotein, it attempts to pull the glycan moiety inside its active site, where the glucosylation reaction should be catalyzed. Only if the glycoprotein is loosely folded the glycan will have the necessary mobility for this process to be successfully completed. The finding that UGGT recognizes misfolded RNase, no matter where the sugar moiety is located along the sequence (Ritter, 2002), is in agreement with this interpretation.

Besides providing a plausible explanation for the mechanism underlying the control of protein folding in the ER, this investigation confirms the power of comparative analyses of the time autocorrelation of the BOOP (Tavernelli and Di Iorio, 2001) and of the energy landscape (Tavernelli et al., 2003). Even from short simulations relative to the timescale of protein denaturation, these approaches pinpoint features indicative of a transition to folding intermediates.

We thank A. Helenius, K. V. Pervushin, and C. Ritter for stimulating discussions.

This work was supported by the Eidgenössische Technische Hochschule, ETH-Zurich (grants 41-2517.5 and 0-50503-00).

REFERENCES

Amadei, A., A. B. M. Linssen, B. L. deGroot, D. M. F. vanAalten, and H. J. C. Berendsen. 1996. An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dyn.* 13:615–625.

Bryngelson, J. D., J. N. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins. 21:167–195.

Chakshusmathi, G., S. Ratnaparkhi-Girish, P. K. Madhu, and R. Varadarajan. 1999. Native-state hydrogen-exchange studies of a fragment complex can provide structural information about the isolated fragments. *Proc. Natl. Acad. Sci. USA.* 96:7899–7904.

Di Iorio, E. E., I. Tavernelli, and W. Yu. 1997. Dynamic properties of the monomeric insect erythrocruorin-III from Chironomus thummi-thummi. Relationships between structural flexibility and functional complexity. *Biophys. J.* 73:2742–2751.

Ellgaard, L., M. Molinari, and A. Helenius. 1999. Setting the standards: quality control in the secretory pathway. *Science.* 286:1882–1888.

Fischer, K. H., and J. A. Hertz. 1991. Spin Glasses. Cambridge University Press, Cambridge UK.

Gerstein, M., J. Tsai, and M. Levitt. 1995. The volume of atoms on the protein surface calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* 249:955–966.

Koradi, R., M. Billeter, and K. Wüthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14: 51–55.

Mezard, M., G. Parisi, and M. A. Virasoro. 1987. Spin Glass Theory and Beyond. Word Scientific, Singapore.

Pliska, P., and E. Marinari. 1993. On heteroplymer shape dynamics. *Europhysics Lett.* 22:167–173.

Ritter, C. 2002. Characterization of the substrate properties recognized by the ER folding sensor UDP-glucose:glycoprotein glucosyltransferase. *In* Biology. Swiss Federal Institute of Technology-ETH, Zurich. 111.

Ritter, C., and A. Helenius. 2000. Recognition of local glycoprotein misfolding by the ER folding sensor UDP-glucose: glycoprotein glucosyltrasferase. *Nat. Struct. Biol.* 7:278–280.

Rosa, J. J., and F. M. Richards. 1981. Hydrogen exchange from identified regions of the S-protein component of ribonuclease as a function of temperature, pH, and the binding of S-peptide. *J. Mol. Biol.* 145:835–851.

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

Sadoc, J.-F., and R. Mosseri. 1999. Geometrical Frustration. Cambridge University Press, Cambridge, UK. 307

Stella, L., M. Nicotra, G. Ricci, N. Rosato, and E. E. Di Iorio. 1999. Molecular dynamics simulations of human glutathione transferase P 1–1: analysis of the induced-fit mechanism by GST binding. Proteins. 37:1–9.

Tavernelli, I., S. Cotesta, and E. E. Di Iorio. 2003. Protein dynamics, thermal stability, and free energy landscapes: a molecular dynamics investigation. *Biophys. J.* 85:2641–2649.

Tavernelli, I., and E. E. Di Iorio. 2001. The interplay between protein dynamics and frustration of non-bonded interactions as revealed by molecular dynamics simulations. *Chem. Phys. Lett.* 345:287–294.

Tironi, I. G., R. Sperb, P. E. Smith, and W. F. van Gunsteren. 1995. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* 102:5451–5459.

van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, A. E. Scott, and I. G. Tironi. 1996. Biomolecular Simulation: The GROMOS96 Manual and User Guide. VDF Hochschulverlag AG an der ETH Zürich, Zürich.

Wolynes, P. G., and W. A. Eaton. 1999. The physics of protein folding. *Physics World.* 12:39–44.

Wolynes, P. G., Z. Luthey-Schulten, and J. N. Onuchic. 1996. Fast folding experiments and the topography of protein folding energy landscapes. Chemistry & Biology. 3:425–432.