PROTEINS: Structure, Function, and Bioinformatics 65:285–295 (2006)

Molecular Dynamics Study of the Stabilities of Consensus Designed Ankyrin Repeat Proteins

Haibo Yu,[1] Andreas Kohl,[2] H. Kaspar Binz,[2] Andreas Plückthun,[2] Markus G. Grütter,[2] and Wilfred F. van Gunsteren[1]*

[1]*Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich, ETH Hönggerberg, Zürich, Switzerland*
[2]*Biochemisches Institut, Universität Zürich, Zürich, Switzerland*

ABSTRACT **Two designed ankyrin repeat (AR) proteins (E3_5 and E3_19) are high homologous (with about 87% sequence identity) and their crystal structures have a Cα atom-positional root-mean-square difference of about 0.14 nm. However, it was found that E3_5 is considerably more stable than E3_19 in guanidinium hydrochloride and thermal denaturation experiments. With the goal of providing insights into the various factors contributing to the stabilities of the designed AR proteins and suggesting possible mutations to enhance their stabilities, homology modeling and molecular dynamics (MD) simulations with explicit solvent have been performed. Because the crystal structure of E3_19 was solved later than that of E3_5, a homology model of E3_19 based on the crystal structure of E3_5 was also used in the simulations. E3_5 shows a very stable trajectory in both crystal and solution simulations. In contrast, the C-terminal repeat of E3_19 unfolds in the simulations starting from either the modeled structure or the crystal structure, although it has a sequence identical to that of E3_5. A continuum electrostatic model was used to estimate the effect of single mutations on protein stability and to study the interaction between the internal ARs and the C-terminal capping AR. Mutations involving charged residues were found to have large effects on stability. Due to the difference in charge distribution in the internal ARs of E3_19 and E3_5, their interaction with the C-terminal capping AR is less favorable in E3_19. The simulation trajectories suggest that the stability of the designed AR proteins can be increased by optimizing the electrostatic interactions within and between the different repeats. Proteins 2006;65:285–295.**

Key words: ankyrin repeat; stability; protein design; homology modeling; molecular dynamics; continuum electrostatics; Poisson–Boltzmann

INTRODUCTION

Ankyrin repeat (AR) proteins mediate a host of important protein–protein interactions in virtually all species, ranging from viruses to humans and are found intracellularly, extracellularly and in membrane-bound form.[1–4] More than 25,000 ARs have been identified in more than 4000 different proteins in the nonredundant protein database (http://smart.embl-heidelberg.de)[5–7] since this motif was first discovered in the yeast cell cycle.[8,9] They are built of tightly joined repeats of usually 33 amino acids. Each repeat forms a structural unit ($\beta_2\alpha_2$) which consists of a β-turn, followed by two antiparallel α-helices and a loop reaching the turn of the next repeat.[1–4] The extended β-sheet projects away from the helical pairs almost at 90° angles, resulting in a characteristic L-shaped cross-section. Some AR proteins consist solely of AR; others are multidomain molecules, in which ARs are combined with other structural modules. The number of repeats in different protein is highly variable. Usually four to six repeats assemble to form a domain, but proteins containing up to over 30 consecutive repeats have been found.[4] The AR architecture permits adapting the size and varying and modulating the binding surface to a target protein, leading to high-affinity interactions. AR proteins are involved in a wide variety of biological processes, such as transcription initiation or inhibition, cell cycle regulation, maintenance of cytoskeletal integrity, ion transport, and cell–cell signalling.[1–4] Recent X-ray and NMR structures of AR proteins, alone and in complex with their target proteins, have provided invaluable insights into the molecular basis for their wide variety of biological functions.[2,4] Up to now, the atomic structures of 18 (including 13 naturally occurring and 5 designed) AR proteins have

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Haibo Yu's present address is Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 University Ave., Madison, WI 53706.

Andreas Kohl's present address is Institutionen för medicinsk biokemi och biofysik (MBB), Scheeles väg 2, Solna, Karolinska Institutet, 171 77 Stockholm, Sweden.

Grant sponsor: Swiss National Science Foundation; Grant sponsor: National Center of Competence in Research. Structural Biology, Switzerland.

*Correspondence to: Wilfred F. van Gunsteren, Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH Honggerberg, 8093 Zurich, Switzerland. E-mail: wfvgn@igc.phys.chem.ethz.ch or igc-sec@igc.phys.chem.ethz.ch

Received 11 April 2005; Revision 7 February 2006; Accepted 8 February 2006

Published online 31 August 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20991

been deposited in the Protein Data Bank.[10] Even though different parts of the surface of an AR protein could be involved in protein–protein interactions, most AR proteins interact with their target proteins via the protruding β-turns and the following α-helices.[2,4]

For practical biotechnological applications of proteins, in addition to specific functions of binding or catalysis, it is important to use molecules that are well behaved with regard to biophysical properties, such as stability, high yield of folding, and low propensity for aggregation. Various design and engineering approaches have been proposed in protein design and protein engineering.[11] A particularly promising route exploiting the sequence information and creating molecules with the desired biophysical properties is consensus design. It makes use of statistical analyses of sequence alignments of families of homologous proteins.[12–14] Its underlying idea is that structurally important residues of proteins are more conserved than other residues during protein evolution. In contrast, residues important for the specific binding function of an individual molecule are not conserved across the family. Thus, replacing a residue with a corresponding consensus amino acid may improve the stability or the folding efficiency of a protein. Generally, natural AR proteins are considered to be thermodynamically rather unstable and prone to aggregation. Recently, consensus design has been applied to AR proteins. Consensus protein with two to four repeats have been reported[15] and a library approach has been presented[14,16–19] in which two to four internal repeats were placed between N- and C-terminal capping repeats, where the internal repeats contain randomized interaction residues. It has been shown that designed AR proteins are well expressed and thermodynamically very stable. The structures of two N3C AR proteins (E3_5 and E3_19), with three internal ARs, were determined by X-ray crystallography.[16,19] E3_19 differs from E3_5 by 20 amino acids at the randomized positions (all within the internal ARs, about 13% of the sequence). In Figure 1, the superposition of the two crystal structures, based on a least-squares fitting of the Cα atoms, is shown. The Cα atom-positional root-mean-square difference (RMSD) is about 0.14 nm. The major structural difference is the orientation of two helices in the C-terminal capping AR. GdmCl unfolding and thermal denaturation experiments followed by CD showed that E3_19 is considerably less stable than E3_5.[16,17]

Complementary to experimental studies,[16,17] we here use molecular dynamics (MD) simulations in combination with homology modeling and continuum electrostatics to study the stability of the designed AR proteins E3_5 and E3_19. Because the crystal structure of E3_19[19] was solved about one year after that of E3_5[16] and the sequences of E3_5 and E3_19 are highly homologous, a homology model was used as starting structure for E3_19 in one of the simulations. With this level of homology, structural modeling approaches have been proven to be highly successful.[20] Results from MD simulation of E3_19 in aqueous solution starting from the modeled structure (E3_19MS) are compared with those obtained starting from the X-ray crystal structure of E3_19 (E3_19S). In addition, the simulated properties of E3_5 in aqueous solution (E3_5S) and in crystal (E3_5C) are compared. From the resulting trajectories, we examined the difference in the fluctuations between the two proteins E3_5 and E3_19. Continuum–electrostatic model calculations based on static structures were used to study the effect of 20 single mutations of the protein E3_5 on its stability in the native fold and on the interactions between the internal ARs and the C-terminal capping AR. Based on our simulation studies, an avenue to increase the stability of the designed AR proteins is proposed.

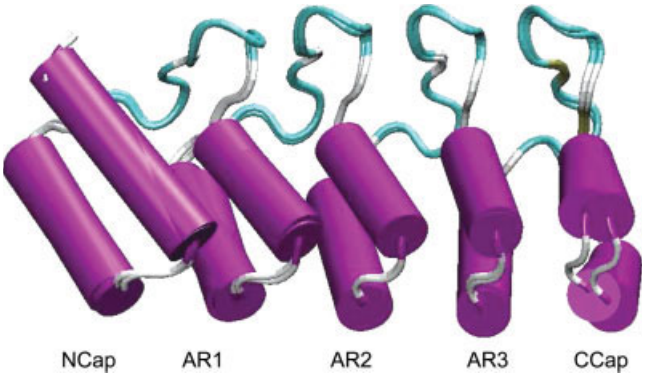

Fig. 1. Superposition of the crystal structures of proteins E3_5[16] (PDB: 1MJ0) and E3_19[19] (PDB: 2BKG) based on least-squares fitting of the positions of the Cα atoms. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Recently, the stability of AR proteins as function of the number of their repeat units has been studied using an Ising-type of model.[21,22] Each repeat unit was modeled as one bead or particle with one parameter describing its tendency to fold. The second parameter of the model was the strength of the first neighbour interaction between beads or repeats that are adjacent to each other. For such a one-dimensional (1D) Ising model the partition function can be calculated analytically as function of the two parameters. The relative stabilities of AR proteins of different lengths were calculated. However, due to its simplicity, the effects of mutations can not be calculated. This excludes a comparison of the results with those obtained here.

MATERIALS AND METHODS

Molecular Dynamics Simulations

MD simulations were performed with the GROMOS software package[23,24] using the force-field parameter set 45A3.[25,26] The simulations are summarized in Table I. Initial coordinates were either taken from the X-ray structures of E3_5 (PDB ID: 1MJ0)[16] and E3_19 (PDB ID: 2BKG)[19] or from the structure of E3_19 obtained by homology modeling from the X-ray structure of E3_5. The modeled structure of E3_19 was generated by substituting the side-chain configurations of the mutated sites of E3_5 by those of a set of standard configurations of amino acids.[23] The 20 mutation sites are: Thr33Glu, Asn35Thr, Asp36Tyr, Tyr38Asp, Ser46-Arg, Asn47Val, Ser66Leu, Leu68Phe, Thr69Ser, Ile71Ser, Ala79Lys, Thr80Arg, His92Tyr, Tyr99Asp, Asn101Thr,

TABLE I. Overview of the MD Simulations of the Different Systems

Simulations Label	E3_5C	E3_5S	E3_19MS	E3_19S
Protein	4 × E3_5	E3_5	E3_19	E3_19
Boundary conditions	Rectangular	Truncated octahedron	Truncated octahedron	Truncated octahedron
Starting structure	PDB ID: 1MJ0	PDB ID: 1MJ0	Modeled	PDB ID: 2BKG
Number of water molecules	712	9522	9528	8790
Number of ligand/ions				
$SO4^{2-}$	8			
TRS	1			
Na^+	80			
Total charge (e)	0	−16	−16	−16
Simulation time (ns)	6	12	12	12

E3_5C: four E3_5 proteins in a crystal unit cell, starting from the X-ray structure (PDB ID: 1MJ0); E3_5S: protein E3_5 in aqueous solution, starting from the X-ray structure (PDB ID: 1MJ0); E3_19MS: protein E3_19 in aqueous solution, starting from the model structure derived from the X-ray structure of E3_5; E3_19S: protein E3_19 in aqueous solution, starting from the X-ray structure (PDB ID: 2BKG).

Asp102Ile, His104Ser, Lys112Asp, Tyr113Thr, and His125-Tyr. Ionization states of residues were assigned according to a pH value of 8.0. The histidine side-chains were protonated at the N_ε atom. For the crystal simulation denoted as E3_5C, the crystallographically determined 178 water molecules, two sulfate ions, and one tris(hydroxymethyl)aminomethane molecule were included. The simple-point-charge (SPC) water model[27] was used to describe the solvent molecules. In the solution simulations, water molecules were added around the protein within a truncated octahedron with a minimum distance of 1.4 nm between the protein atoms and the square wall. In the crystal simulation E3_5C the starting coordinates of the four symmetry-related E3_5 molecules in the unit cell were obtained by performing the $P2_12_12$ symmetry transformations using the experimental unit cell edge lengths $a = 7.3864$ nm, $b = 4.7360$ nm, and $c = 4.7003$ nm. A total of 80 counterions (Na^+) were included in the crystal simulation. All the bonds were constrained with a geometric tolerance of 10^{-4} using the SHAKE algorithm.[28] A steepest-descent energy minimization of the systems was performed to relax the solute–solvent contacts, while positionally restraining the solute atoms using a harmonic interaction with a force constant of 2.5×10^4 kJ mol^{-1} nm^{-2}. Next, steepest-descent energy minimization of the system without any restraints was performed to eliminate any residual strain. The energy minimizations were terminated when the energy change per step became smaller than 0.1 kJ mol^{-1}. For the nonbonded interactions, a triple-range method with cutoff radii of 0.8/1.4 nm was used. Short-range van der Waals and electrostatic interactions were evaluated every time step based on a charge-group pair-list. Medium-range van der Waals and electrostatic interactions, between (charge group) pairs at a distance longer than 0.8 nm and shorter than 1.4 nm, were evaluated every fifth time step, at which point the pair list was updated. Outside the longer cutoff radius a reaction-field approximation[29] was used with a relative dielectric permittivity of 78.5. The center of mass motion of the whole system was removed every 1000 time steps. Solvent and solute were independently, weakly coupled to a temperature bath of 295 K with a relaxation time of 0.1 ps.[30] In the solution simulations, the systems were also weakly coupled to a pressure bath of 1 atm with a relaxation time of 0.5 ps and an isothermal compressibility of 0.7513×10^{-3} (kJ mol^{-1} nm^{-3})$^{-1}$. One hundred picoseconds of MD simulation with harmonic position restraining of the solute atoms with a force constant of 2.5×10^4 kJ mol^{-1} nm^{-2} were performed to further equilibrate the systems. The simulations E3_5S, E3_19MS, and E3_19S were carried out for 12 ns and the simulation E3_5C for 6 ns. The trajectory coordinates and energies were saved every 0.5 ps for analysis.

Analysis

Atom-positional RMSDs between the structures were calculated by performing a rotational and translational atom-positional least-squares fit of one structure on the second (reference) structure using a given set of atoms. A hydrogen bond is assumed to exist based on a geometric criterion: the hydrogen-acceptor distance is smaller than 0.25 nm and the donor-hydrogen-acceptor angle is larger than 135°. The secondary structure assignment was done using the program PROCHECK.[31] The dipole moment of a set of (partial) charges with non-zero total charge is dependent on the choice of origin of the spatial coordinate system. Because the latter choice is arbitrary, we have calculated the dipole moment of the protein or a part of it in two different ways: either by distributing the non-zero total charge equally over all atoms or by shifting the origin to the Cα atom of residue 79, which atom lies more or less in the middle of the atoms for which the dipole moment is calculated.

Poisson–Boltzmann Type of Calculations

The free energy difference between a protein in its native (N) state and its denatured (D) state, ΔG_{N-D}, is a measure of its stability. For the wild-type protein this quantity is denoted by ΔG^w_{N-D} and for mutants of the protein by $\Delta G^m_{N-\mathrm{D}}$. The relative stability of the mutant with respect to the wild-type protein is then[32]

$$\Delta\Delta G_{N-D} = \Delta G^m_{N-D} - \Delta G^w_{N-D}. \quad (1)$$

Because the free energy is a thermodynamic state function, we can express $\Delta\Delta G_{N-D}$ also as a difference between the difference $\Delta G_N^{m-w} = \Delta G_N^m - \Delta G_N^w$ between the mutant

and wildtype in the native state and the difference $\Delta G_D^{m-w} = \Delta G_D^m - \Delta G_D^w$ between the mutant and wild-type in the denatured state,

$$\Delta\Delta G_{N-D} = \Delta G_N^{m-w} - \Delta G_D^{m-w}. \quad (2)$$

In the absence of structural information about the mutant, ΔG_N^{m-w} can be approximated by using in the free-energy calculation the native wild-type structure and a native mutant structure obtained by homology modeling. The quantity ΔG_D^{m-w} cannot be calculated in an analogous manner because no information on the ensemble of denatured conformations is available. However, it has been proposed by Börjesson and Hünenberger[32] to assume that the quantity ΔG_D^{m-w} is linearly related to ΔG_N^{m-w}, which yields

$$\Delta\Delta G_{N-D} = \Delta G_N^{m-w} - (\alpha\Delta G_N^{m-w} + \beta) = (1-\alpha)\Delta G_N^{m-w} - \beta. \quad (3)$$

An even simpler approximation is to use $\alpha = \beta = 0$. Then the free difference between 20 single site mutants of E3_5 and E3_5 in the native fold can be calculated as[32]

$$\begin{aligned}\Delta G_N^{m-w} &= G_N^m - G_N^w \\ &= \frac{1}{2}\sum_{i=1}^{N} q_i^m \left[\phi_{\epsilon_w}^m(\vec{r}_i) - \phi_{\epsilon_p}^m(\vec{r}_i)\right] \\ &\quad - \frac{1}{2}\sum_{i=1}^{N} q_i^w \left[\phi_{\epsilon_w}^w(\vec{r}_i) - \phi_{\epsilon_p}^w(\vec{r}_i)\right] \\ &\quad + \frac{1}{4\pi\epsilon_0\epsilon_p}\left[\sum_{i=1}^{N-1}\sum_{j>i}^{N}\frac{q_i^m q_j^m}{r_{ij}^m} - \sum_{i=1}^{N-1}\sum_{j>i}^{N}\frac{q_i^w q_j^w}{r_{ij}^w}\right] + \Delta G_{N,np}^{w-m}. \quad (4)\end{aligned}$$

The continuum electrostatics calculations were performed on static protein structures using a modified version of the GROMOS program incorporating the routines of the UHBD program[33,34] for solving the linearized Poisson–Boltzmann equation using a finite-difference algorithm to obtain the electrostatic potential $\phi(\vec{r})$ for a dielectric continuum with permittivity ε and for computing the surface-area dependent nonpolar term (np).[35,36] The partial charges were taken from the GROMOS force-field parameter set 45A3[25,26] used in the explicit solvent MD simulations. The atomic radii of the solute atoms were calculated from Lennard–Jones C_6 and C_{12}- parameters defining the interaction between the specific atom and an SPC water oxygen atom[27] as $R = (2C_{12}/C_6)^{1/6} - 0.14$ nm (the approximate radius of a water molecule subtracted from the atom–water oxygen distance at the minimum of the Lennard–Jones curve). Hydrogen atoms were treated differently and assigned a common radius of 0.01 nm. The protein was centered on a cubic grid of 7.0 nm edge length with a uniform grid spacing of 0.05 nm, and rotated to maximize the solute-to-wall distance (>0.5 nm). The value of the relative dielectric permittivity of the protein interior, ε_p, was set to 2, the ionic strength was set to 0 M. A value of $\varepsilon_w = 78.5$ was used for the relative dielectric permittivity of water. The effective microscopic interfacial tension was set to 10.46 kJ mol^{-1} nm^{-2}.[37]

Poisson–Boltzmann calculations were also used to investigate the interactions and binding energy between the internal ARs and the C-terminal capping AR of the proteins.[35,36] The internal ARs and the C-terminal capping AR were treated as two separate molecules or components. Figure 2 illustrates the thermodynamic cycle to calculate the binding energy which can be expressed as

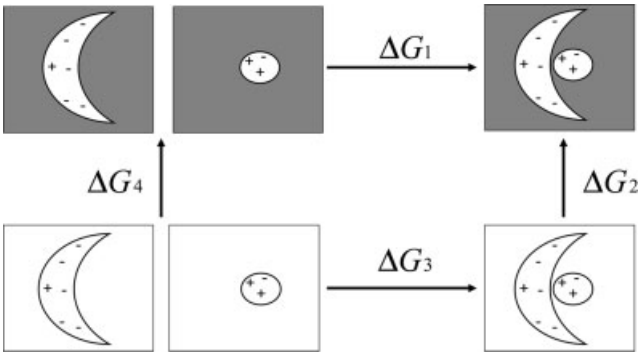

Fig. 2. Thermodynamic cycle for calculating the binding energy of a complex of two molecules or components. The white background indicates an environment identical to the biomolecular interior with a low relative dielectric permittivity (in our case $\varepsilon_p = 2$). The gray background indicates an aqueous environment with the relative dielectric permittivity set to that of the solvent, that is, in our case $\varepsilon_w = 78.5$ of bulk water. ΔG_1 is the binding energy of the complex in aqueous environment, ΔG_2 is the solvation free energy of the complex, ΔG_3 is the binding energy of the components in a constant environment identical to the biomolecular interior, and ΔG_4 is the sum of the solvation free energies of the isolated components.

$$\Delta G_{\text{bind}} = \Delta G_1 = \Delta G_2 + \Delta G_3 - \Delta G_4, \quad (5)$$

where ΔG_2 is the solvation free energy of the complex, ΔG_4 is the sum of the solvation free energies of the isolated components, and ΔG_3 is the binding energy of the components in a constant environment identical to the biomolecular interior. In practice, the solvation free energies ΔG_2 and ΔG_3 are obtained by solving the Poisson–Boltzmann equation and calculating the nonpolar contribution, while ΔG_3 is typically calculated using Coulomb's law. The parameters were the same as those used before.

RESULTS AND DISCUSSION

Comparison between the Crystal Structures and Modeled Structure of E3_5 and E3_19

Molecular dynamics with positional restraining the protein atoms in water was used to refine the modeled structure of E3_19 derived from the crystal structure of E3_5. With the large force constant used for the restraining potential energy term, little structural rearrangements were observed during the refinement, resulting in a Cα atom-positional RMSD of 0.02 nm between the modeled structure before and after MD refinement. The crystal structure of E3_19 is rather close (with Cα atom-positional RMSD of 0.14 nm) to that of E3_5 (Fig. 1). The major structural difference between E3_5 and E3_19 comes from the orientation of the C-terminal capping ARs.

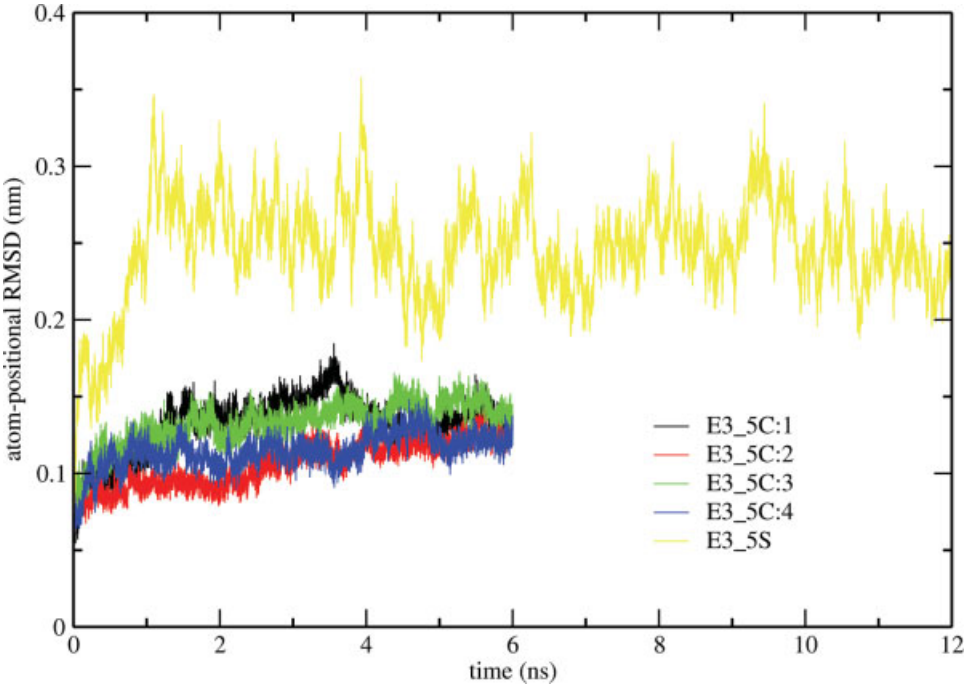

Fig. 3. Backbone atom-positional RMSD in the simulations E3_5C and E3_5S of protein E3_5 with respect to the X-ray structure as a function of simulation time. The values for the four molecules in the unit cell of the crystal simulation E3_5C are depicted in black, red, green, and blue, respectively, while those in the solution simulation E3_5S are depicted in yellow.

Convergence and Stability of Simulations

Atom-positional RMSDs were determined for the back-bone atoms in the MD trajectories with respect to the X-ray derived or modeled starting structures. In Figure 3, the back-bone atom-positional RMSDs of protein E3_5 are shown for both the E3_5C and E3_5S simulations. In the crystal simulation, the RMSDs converge more slowly than in the solution simulation. Not unexpectedly, the crystal simulation stays—in terms of RMSD—closer to the X-ray structure than the solution simulation. The four different molecules in the crystal simulation show very similar behavior. In the solution simulation, larger fluctuations are observed. In Figure 4, backbone atom-positional RMSDs for the proteins E3_5 and E3_19 in solution are shown for the simulations E3_5S, E3_19MS and E3_19S. The E3_5 protein remains close to its X-ray structure over the entire 12 ns simulation (E3_5S, black solid line) with a RMSD of 0.25 nm at the end of the simulation. Starting from the modeled structure, the E3_19 protein was stable for about 5 ns in the simulation E3_19MS (red solid line). Then an increase of RMSD indicating a large structure rearrangement is observed (details will be discussed later). The RMSD in the simulation E3_19S (green solid line) steadily increases after 5 ns. Taking together the two simulations E3_19MS and E3_19S, we can conclude that in solution the protein E3_19 is less stable than E3_5, which has also been observed experimentally.[16,17] Regarding the internal AR parts of these two proteins, there is not much difference in their stabilities. The backbone atom-positional RMSDs of the atoms in the internal AR (IAR) helices hoover between 0.07 and 0.13 nm for simulations E3_5S, E3_19MS, and E3_19S (blue, yellow, and brown solid lines, respectively). This indicates that the internal ARs of the two proteins are comparably stable and the difference between their overall stabilities mainly resides in the N-terminal and C-terminal caps. This is remarkable, as the capping repeats are of identical sequence.

Atom-Positional RMS Fluctuations and Structural Rearrangements

Cα atom-positional RMSFs were calculated for the final 4 ns of the simulation to ensure full convergence.[38] Not unexpectedly, compared to the RMSFs in the crystal simulation (E3_5C) of protein E3_5 (Fig. 5), the solution simulation (E3_5S) shows larger atomic fluctuations especially in the turn and loop regions. This is consistent with early experimental and simulation studies on the structural comparison between proteins in crystal and in solution.[39–41] The RMSFs derived from the X-ray crystallographic isotropic atomic B-factors are shown in Figure 5 for comparison. They are of comparable magnitude. In the crystal simulation, the atoms in helices are less mobile in comparison to the crystallographically derived B-factors, while for the loops the opposite is observed. As pointed out elsewhere,[38,41] these two sets of RMSF are not entirely comparable due to their different definitions.

Figure 6 shows RMSFs of Cα atoms in the solution simulations E3_5S (black), E3_19MS (red), and E3_19S (green). Generally, those of E3_5S are smaller than those

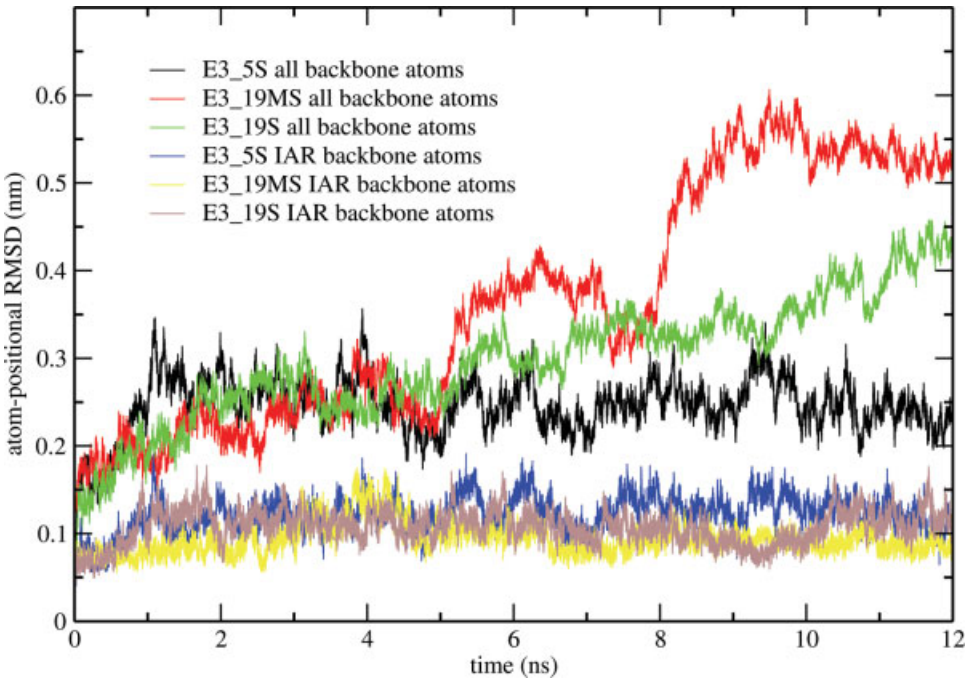

Fig. 4. Backbone atom-positional RMSD in the simulations E3_5S, E3_19MS, and E3_19S of proteins E3_5 and E3_19 in solution with respect to their respective initial structures as a function of simulation time. The values in the simulation E3_5S, E3_19MS, and E3_19S are depicted in black, red, and green, respectively. Those of the backbone atoms in the internal AR (IAR) helices are depicted in blue, yellow, and brown, respectively.

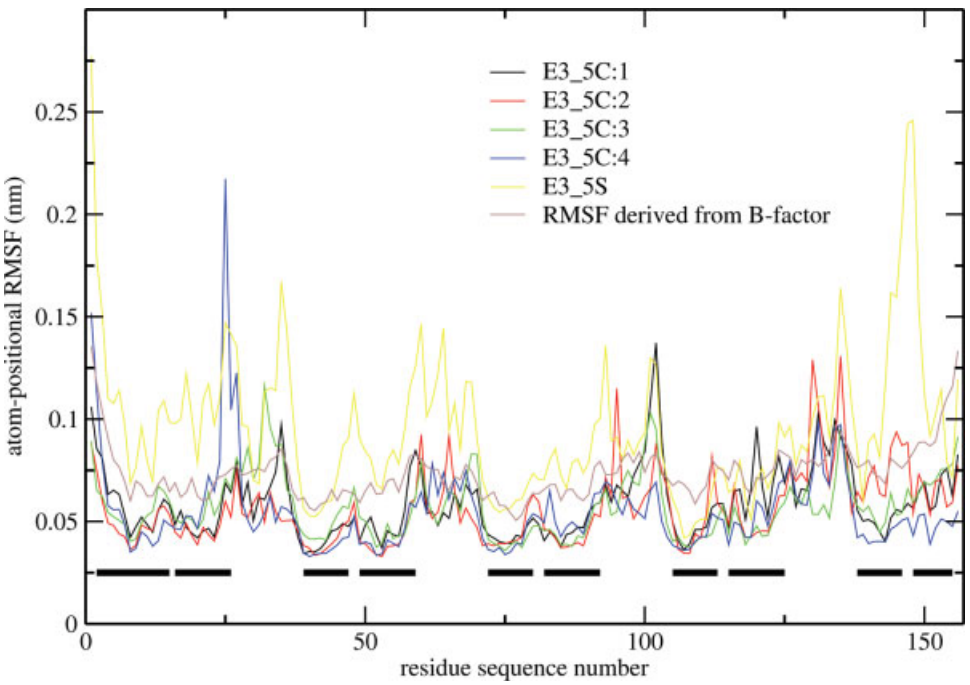

Fig. 5. Backbone atom-positional RMSF in the simulations E3_5C and E3_5S of protein E3_5. The values for the four molecules in the unit cell of simulation E3_5C are depicted in black, red, green, and blue, respectively, those in the solution simulation E3_5S are depicted in yellow, and those calculated from the crystallographically derived experimental B-factors are depicted in brown. Residues found in α-helical conformations (according to the X-ray structure) are indicated by black bars.

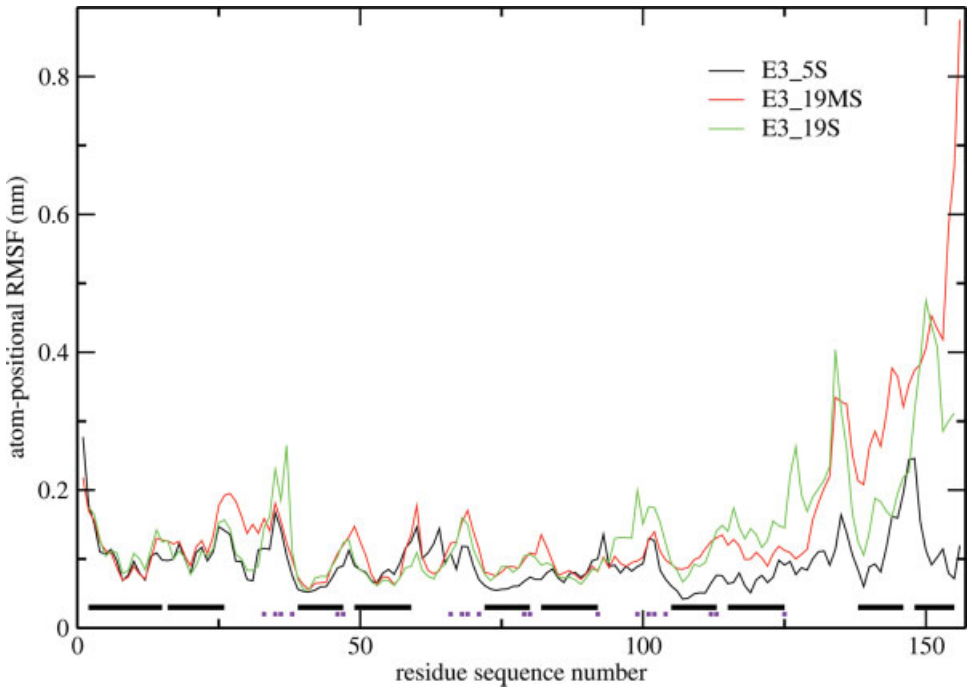

Fig. 6. Backbone atom-positional RMSF in the simulations E3_5S (black), E3_19MS (red), and E3_19S (green) of proteins E3_5 and E3_19 in solution. Residues found in α-helical conformations (according to the X-ray structure) are indicated by black bars, while the 20 residues, for which E3_5 and E3_19 differ, are indicated by purple squares.

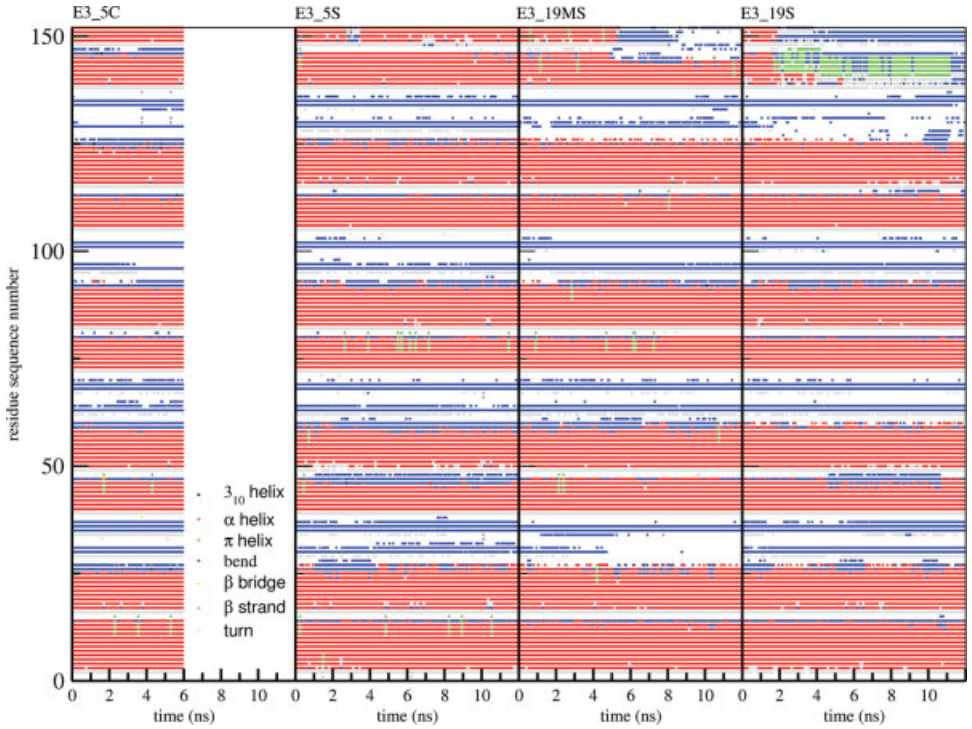

Fig. 7. Secondary structure of the proteins E3_5 and E3_19 in the simulations E3_5C, E3_5S, E3_19MS, and E3_19S as a function of time. For the crystal simulation E3_5C only the secondary structure for molecule 1 is shown. Secondary structure definition is according to Laskowski and co-workers.[31] 3_{10}-helix (black), α-helix (red), π-helix (green), bend (blue), β-bridge (yellow), β-strand (brown), and turn (gray). Each dot in the MD trajectories represents a period of 50 ps.

of E3_19MS and E3_19S, especially at the C-terminus. Although the difference in sequence between proteins E3_5 and E3_19 resides within the internal ARs (purple squares in Fig. 6), the major difference in mobility is observed at the C-terminal cap of the protein.

Figure 7 shows the secondary structure pattern in the simulations E3_5C, E3_5S, E3_19MS, and E3_19S as obtained from the program PROCHECK.[31] In the crystal simulation E3_5C, the helical secondary structure elements are well preserved in all four molecules (only the first molecule is shown), while in the solution simulation E3_5S, slightly more fraying at the ends of the helices is observed. In the E3_19MS simulation, the C-terminal helix totally disappeared after 5 ns. This occurs still earlier in the simulation E3_19S, in which the second to last helix disappeared, too, thereby partially adopting π-helical conformations.

In Figure 8, the final structures of simulations E3_5S (Panel A), E3_19MS (Panel B), and E3_19S (Panel C) are shown. The 20 mutation sites are shown in licorice format with the negatively charged residues in red, the positively charged residues in blue, and the neutral residues in cyan. The final structure of E3_5 in the simulation E3_5S is similar to the X-ray crystal structure and those in the crystal simulation (picture not shown). The final structure of E3_19 in both simulations E3_19MS and E3_19S illustrates the unfolding of its C-terminal AR.

Hydrogen Bonds

In the crystal simulation E3_5C, most of the hydrogen bonds determined in the X-ray structure are preserved and the four molecules show very similar patterns, except at the end of the helices (see Supplementary Material). In the solution simulation E3_5S, the hydrogen bond pattern is also similar to that of the X-ray structure. In the solution simulations E3_19MS and E3_19S, however, the backbone hydrogen bonds that are responsible for the eight helices of the first four ARs (residues 4–25, 39–59, 72–91, and 105–124) are stable, while those in the two C-terminal helical structure elements (residues 138–142, 148–153) are much reduced or lost in the simulation, which confirms the observations from the secondary structure analysis.

Relative Stability from Continuum Electrostatics

The free energy differences between 20 mutants of the protein E3_5 and E3_5 itself were calculated by solving the linearized Poisson–Boltzmann equation for the structure after energy minimization of the X-ray structure in explicit solvent. The values of ΔG_N^{m-w} arise from the partial cancellation of two large contributions; a Coulomb contribution favoring mutations leading to positive charges and the solvation contribution favoring the mutations leading to negative charges. The nonpolar contribution is relative small, except for the mutation from Ala to Lys. Qualitatively, we can see that mutations changing the charge of a residue have a large effect on the stability of the proteins (Table II). The calculations performed

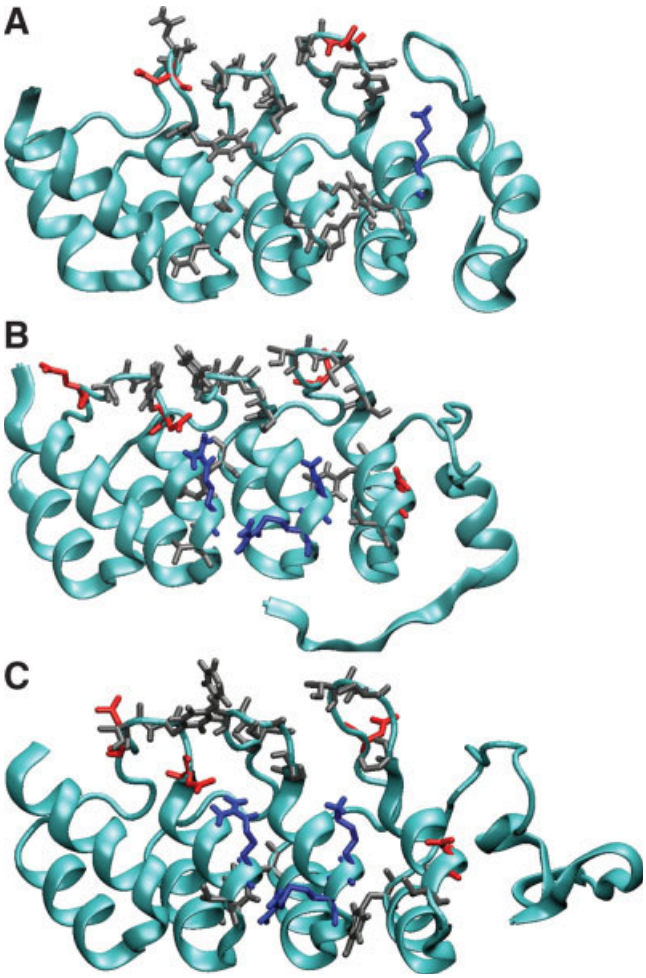

Fig. 8. Snapshots of the proteins E3_5 and E3_19 in the solution simulations E3_5S (**A**), E3_19MS (**B**), and E3_19S (**C**) at 12 ns. The 20 residues that are different between the two proteins are drawn in licorice format with the negatively charged residues in red, the positively charged residues in blue, and the neutral residues in grey in panels **A, B,** and **C.**

here are very crude for the following reasons: (1) the structures of mutants were built by replacing the corresponding side-chains; (2) the protonation states of charged residues were the same as those used in explicit solvent simulations without taking into account the protein environment[42–45]; and (3) the dielectric permittivity of the protein was set to 2 without any adjustment.[46,47] The sum of the values for the 20 individual mutations amounts to 3 kJ mol^{-1}. This would mean that E3_19 is barely less stable than E3_5. However, by applying the same approach to the X-ray structures or the modeled structure of E3_19 and E3_5, we obtain $\Delta G_N^{E3\_19M-E3\_5} = 50$ kJ mol^{-1} and $\Delta G_N^{E3\_19-E3\_5} = 40$ kJ mol^{-1}, which values are comparable in magnitude to the experimental value of 22 ± 10 kJ mol^{-1}.[16,17] The discrepancy between the results obtained by the two different approaches might stem from two possible sources: first the effects of single mutations on the free energy difference need not be additive (for noncharged mutations, see, e.g., Ref. 48), and second the total structural rearrangement induced

TABLE II. Free Energy $\Delta G_N^{m\text{-}w}$ (Eq. 4) for 20 Mutations of the Protein E3_5 Calculated from a Solvated X-Ray Structure Using a Poisson–Boltzmann Based Continnum Electrostatics Method[32]

Residue	E3_5	Charge (e)	E3_19	Charge (e)	$\Delta G_N^{m\text{-}w}$ (kJ mol^{-1})	Residue	E3_5	Charge (e)	E3_19	Charge (e)	$\Delta G_N^{m\text{-}w}$ (kJ mol^{-1})
33	Thr	0	Glu	−1	−431	79	Ala	0	Lys	1	416
35	Asn	0	Thr	0	80	80	Thr	0	Arg	1	447
36	Asp	−1	Tyr	0	522	92	His	0	Tyr	0	−30
38	Tyr	0	Asp	−1	−434	99	Tyr	0	Asp	−1	−595
46	Ser	0	Arg	1	410	101	Asn	0	Thr	0	90
47	Asn	0	Val	0	88	102	Asp	−1	Ile	0	548
66	Ser	0	Leu	0	−2	104	His	0	Ser	0	11
68	Leu	0	Phe	0	4	112	Lys	1	Asp	−1	−1110
69	Thr	0	Ser	0	−6	113	Tyr	0	Thr	0	16
71	Ile	0	Ser	0	−7	125	His	0	Tyr	0	−14

TABLE III. Charge Distribution over the Various ARs in the Proteins E3_5 and E3_19 (in e)

	NCap	AR1	AR2	AR3	CCap	Total
E3_5	−1	−4	−3	−3	−5	−16
E3_19	−1	−4	−1	−5	−5	−16

The atomic partial charges were taken from the GROMOS force-field parameter set 45A3.[25,26] NCap, residues 1–32; AR1, residues 33–65; AR2, residues 66–98; AR3, residues 99–131; CCap, residues 132–156.

by 20 single mutations (which is ignored in the present treatment) can be fairly large.

Charge Distribution and Dipole Moments

The total charges of the two proteins E3_5 and E3_19 are both −16 e (see Table I), however, the 20 mutations do change the charge distribution between the individual internal ARs (see Table III). The major difference between E3_5 and E3_19 is that negative charge has been shifted from the second to the third internal AR in E3_19 compared to E3_5. Because the total charge did not change, the change in dipole moment resulting from the redistribution of charge will dominate the change in electrostatic energy due to mutations. The dipole moments of the internal ARs (AR1, AR2, AR3) and of the 20 mutated residues were calculated in two different ways, that is, after neutralizing the overall charge or after shifting the origin of the coordinate system to the position of the Cα atom of Ala79 (E3_5) or Lys79 (E3_19). The dipole moment of the internal ARs of E3_19 is about 72 Debye larger than that of E3_5, while the dipole moment of the 20 mutation residues is about 48 Debye larger. Thus, the 20 mutations constitute a redistribution of charge within the internal ARs that enlarges its dipole moment. The shift of negative charge of AR2 to AR3, that is, in the direction of the negatively charged C-terminal capping AR, is likely to destabilize the latter by electrostatic repulsion.

Binding Energy between Internal ARs and the C-terminal Capping AR

The binding energy between the internal ARs and the C-terminal capping AR was calculated using Eq. 5. The difference in binding energy is $\Delta\Delta G_{bind} = \Delta G_{bind}^{E3\_19} - \Delta G_{bind}^{E3\_5} = -1256 - (-1281) = 25$ kJ mol^{-1}, which indicates that the binding energy between the internal ARs and the C-terminal capping AR is less favorable in E3_19 than in E3_5. This is likely to be due to the unfavorable electrostatic interaction between both the negatively charged third internal AR (−5 e) and the C-terminal capping AR (−5 e) (Table III). In the design, the sequence of the C-terminal capping AR was taken from the naturally occurring protein GABP with slight modification.[49] In crystallographic studies of other AR proteins this very C-terminal capping AR could not be located, which might also be caused by the unfavorable electrostatic interactions with internal ARs due to the highly negative charges.

CONCLUSIONS

Molecular dynamics simulations offer insights into the energetic, structural, and dynamic behavior of macromolecules in a solvent environment. They can complement a too static picture, for example, derived from a crystal X-ray structure, of molecules by highlighting the motions of the protein atoms in solution. This may hint at modes of function based on this fluctuating nature of the molecules.[50,51] On the other hand, continuum electrostatics models provide a static alternative for investigating the energetic properties of biomolecular systems,[35,36] in particular the electrostatic interactions in protein–ligand (or protein) complexes or between particular parts of a protein.

In the present work, MD simulations with explicit solvent models together with homology modeling and continuum electrostatics were used to study the stabilities of two designed AR proteins, E3_5 and E3_19. By comparison between the properties of E3_5 in the crystal simulation and in the solution simulations, it was found that the structural properties of E3_5 are similar in solution and in the crystalline state with minor differences regarding loop or side-chain conformations, while the variability in atom position is somewhat larger in solution. Complementary to the experimental data, MD simulations of E3_19 starting from a structure obtained by homology modeling and from the X-ray structure of E3_19 both demonstrate that the difference in stability

between E3_5 and E3_19 seems mainly to be due to the difference in stability of the C-terminal capping AR, while the proteins have similar properties for the internal ARs. A continuum electrostatics calculation shows that the mutations involving charged residues have a large effect on the relative stabilities of the proteins. Through analysis of the charge redistribution when mutating E3_5 into E3_19, we found that the third internal AR, which is spatially closest to the C-terminal capping AR, becomes more negatively charged, which explains the unfavorable binding energy as obtained from continuum electrostatics between the negatively charged internal ARs and negative charged C-terminal capping AR and the unfolding of the latter in the MD simulation of E3_19.

Previous experimental studies[16,17] have shown that the conserved residues contribute to the stabilities of AR proteins. Additionally, we found, through simulation, the importance to their stabilities of the electrostatic interactions between the internal ARs and the C-terminal capping AR. If one would aim at improving the stabilities of the designed AR proteins, our results suggest to optimize the distribution of charge over the ARs. One possible way might be to decrease the net charge of the C-terminal capping AR to favor electrostatic interactions with the internal ARs. However, a decrease of the net charge might affect the solubility of the AR protein. Another possibility is to redistribute the charges within individual ARs with an eye to optimize their charge–charge and dipolar interactions. In summary, our study illustrates the complementarity between experimental and simulation studies when designing proteins with specific properties. Simulation studies offer detailed energetic and structural properties of proteins in solution which are inaccessible to experimental probes, and which may suggest design changes subsequently to be investigated experimentally.

ACKNOWLEDGMENTS

Haibo Yu thanks Dr. Ulf Börjesson, Dr. Mika Kastenholz, and Dr. Chris Oostenbrink for helpful discussions on continuum electrostatics.

REFERENCES

1. Bork P. Hundreds of ankyrin-like repeats in functionally diverse proteins—mobile modules that cross phyla horizontally. Proteins 1993;17:363–374.
2. Sedgwick SG, Smerdon SJ. The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem Sci 1999;24:311–316.
3. Mohler PJ, Gramolini AO, Bennett V. Ankyrins. J Cell Sci 2002; 115:1565–1566.
4. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY. The ankyrin repeat as molecular architecture for protein recognition. Protein Sci 2004;13:1435–1448.
5. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 1998;95:5857–5864.
6. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P. Recent improvements to the smart domain-based sequence annotation resource. Nucleic Acids Res 2002;30:242–244.
7. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. SMART 4.0: towards genomic data integration. Nucleic Acids Res 2004;32:D142–D144.
8. Aves SJ, Durkacz BW, Carr A, Nurse P. Cloning, sequencing and transcriptional control of the *Schizosaccharomyces pombe* cdc10 'start' gene. EMBO J 1985;4:457–463.
9. Breeden L, Nasmyth K. Similarity between cell-cycle genes of budding yeast and fission yeast and the *Notch* gene of *Drosophila*. Nature 1987;329:651–654.
10. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
11. Regan L, Jackson SE. Engineering and design — protein design: theory and practice — editorial overview. Curr Opin Struct Biol 2003;13:479–481.
12. Lehmann M, Wyss M. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Curr Opin Biotechnol 2001;12:371–375.
13. van den Burg B, Eijsink VGH. Selection of mutations for increased protein stability. Curr Opin Biotechnol 2002;13:333–337.
14. Forrer P, Binz HK, Stumpp MT, Plückthun A. Consensus design of repeat proteins. ChemBioChem 2004;5:183–189.
15. Mosavi LK, Minor DL, Peng ZY. Consensus-derived structural determinants of the ankyrin repeat motif. Proc Natl Acad Sci U S A 2002;99:16029–16034.
16. Kohl A, Binz HK, Forrer P, Stumpp MT, Plückthun A, Grütter MG. Designed to be stable: crystal structure of a consensus ankyrin repeat protein. Proc Natl Acad Sci U S A 2003;100:1700–1705.
17. Binz HK, Stumpp MT, Forrer P, Amstutz P, Plückthun A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol 2003;332:489–503.
18. Forrer P, Stumpp MT, Binz HK, Plückthun A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. FEBS Lett 2003;539:2–6.
19. Binz HK, Kohl A, Plückthun A, Grütter MG. Crystal structure of a consensus-designed ankyrin repeat protein: implications on stability. Proteins 2006; this issue.
20. Säli A. Modeling mutations and homologous proteins. Curr Opin Biotechnol 1995;6:437–451.
21. Mello CC, Barrick D. An experimentally determined protein folding energy landscape. Proc Natl Acad Sci U S A 2004;101:14102–14107.
22. Kajander T, Cortajarena AL, Main ERG, Mochrie SGJ, Regan L. A new folding paradigm for repeat proteins. J Am Chem Soc 2005;127:10188–10190.
23. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: the GROMOS manual and user guide. ETH Zürich, Switzerland: vdf Hochschulverlag; 1996.
24. Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, van Gunsteren WF. The GROMOS biomolecular simulation program package. J Phys Chem A 1999;103:3596–3607.
25. Daura X, Mark AE, van Gunsteren WF. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. J Comput Chem 1998;19:535–547.
26. Schuler LD, Daura X, van Gunsteren WF. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. J Comput Chem 2001;22:1205–1218.
27. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Intermolecular forces. Pullman B, editor, Dordrecht: Reidel; 1981. p 331–342.
28. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of cartesian equations of motion of a system with constraints — molecular dynamics of nalkanes. J Comput Phys 1977;23:327–341.
29. Tironi IG, Sperb R, Smith PE, van Gunsteren WF. A generalized reaction field method for molecular dynamics simulations. J Chem Phys 1995;102:5451–5459.
30. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
31. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK — a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.

32. Börjesson U, Hünenberger PH. Effect of mutations involving charged residues on the stability of staphylococcal nuclease: a continuum electrostatics study. Protein Eng 2003;16:831–840.
33. Davis ME, Madura JD, Luty BA, McCammon JA. Electrostatics and diffusion of molecules in solution—simulations with the University-of-Houston-Brownian Dynamics program. Comput Phys Commun 1991;62:187–197.
34. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A, Antosiewicz J, Gilson MK, Bagheri B, Scott LR, McCammon JA. Electrostatics and diffusion of molecules in solution—simulations with the University-of-Houston-Brownian Dynamics program. Comput Phys Commun 1995;91:57–95.
35. Sharp KA, Honig B. Electrostatic interactions in macromolecules—theory and applications. Annu Rev Biophys Biophys Chem 1990;19:301–332.
36. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science 1995;268:1144–1149.
37. Hünenberger PH, Helms V, Narayana N, Taylor SS, McCammon JA. Determinants of ligand binding to cAMP-dependent protein kinase. Biochemistry 1999;38:2358–2366.
38. Hünenberger PH, Mark AE, van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. J Mol Biol 1995;252:492–503.
39. Billeter M, Kline AD, Braun W, Huber R, Wüthrich K. Comparison of the high-resolution structures of the α-amylase inhibitor tendamistat determined by nuclear magnetic resonance in solution and by X-ray-diffraction in single crystals. J Mol Biol 1989;206:677–687.
40. van Gunsteren WF, Berendsen HJC. Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. J Mol Biol 1984;176:559–564.
41. Stocker U, Spiegel K, van Gunsteren WF. On the similarity of properties in solution or in the crystalline state: a molecular dynamics study of hen lysozyme. J Biomol NMR 2000;18:1–12.
42. Bashford D, Karplus M. pK_a's of ionizable groups in proteins — atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
43. Antosiewicz J, McCammon JA, Gilson MK. The determinants of pK_a's in proteins. Biochemistry 1996;35:7819–7833.
44. Sham YY, Chu ZT, and Warshel A. Consistent calculations of pK_a's of ionizable residues in proteins: semi-microscopic and microscopic approaches. J Phys Chem B101:4458–4472, 1997.
45. Simonson T, Carlsson J, Case DA. Proton binding to proteins: pK_a calculations with explicit and implicit solvent models. J Am Chem Soc 2004;126:4167–4180.
46. Pitera JW, Falta M, van Gunsteren WF. Dielectric properties of proteins from simulation: the effects of solvent, ligands, pH, and temperature. Biophys J 2001;80:2546–2555.
47. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins 2001;44: 400–417.
48. Yu HB, Amann M, Hansson T, Köhler J, Wich G, van Gunsteren WF. Effect of methylation on the stability and solvation free enegy of amylose and cellulose fragments: A molecular dynamics study. Carbohyr Res 2004;339:1697–1709.
49. Batchelor AH, Piper DE, de la Brousse FC, McKnight SL, Wolberger C. The structure of GABP α/β: an ETS domain ankyrin repeat heterodimer bound to DNA. Science 1998;279: 1037–1041.
50. Hansson T, Oostenbrink C, van Gunsteren WF. Molecular dynamics simulations. Curr Opin Struct Biol 2002;12:190–196.
51. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002;9:646–652.