Use of Molecular-Dynamics Simulation for Optimizing Protein Stability: Consensus-Designed Ankyrin Repeat Proteins

by **Moritz Winger** and **Wilfred F. van Gunsteren***

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich, ETH, CH-8093 Zürich (phone: +41-44-6325502; fax: +41-44-6321039; e-mail: wfvgn@igc.phys.chem.ethz.ch, igc-sec@igc.phys.chem.ethz.ch)

In earlier work, two highly homologous (87% sequence identity) ankyrin repeat (AR) proteins, E3_5 and E3_19, were studied using molecular-dynamics (MD) simulation. Their stabilities were compared, and it was found that the C-terminal capping unit is unstable in the protein E3_19, in agreement with CD experiments. The different stabilities of these two very similar proteins could be explained by the different charge distributions among the AR units of the two proteins. Here, another AR protein, N3C, with yet another charge distribution has been simulated using MD, and its stability was analyzed. In agreement with the experimental data, the structure of N3C was found to be less stable than that of E3_5, but, in contrast to E3_19, secondary structure was only slightly lost, while structurally N3C is closer to E3_19 than to E3_5. The results suggest that a homogeneous charge distribution over the repeat units does enhance the stability of design AR proteins in aqueous solution, which, however, may be modulated by the bulkiness of amino-acid side chains involved in the mutations.

Introduction. – Ankyrin repeat (AR) proteins are involved in protein–protein interactions in most species [1–4]. The numbers of identified ARs runs into the ten thousands. They are a part of thousands of proteins. Usually, they consist of 33 amino acids, each AR forming a structural module (β_2,α_2) consisting of a β-turn, followed by two antiparallel α-helices, and a loop connecting to the turn of the next AR. The AR architecture allows for modification of the size and character of the binding surface to a target protein with the aim of high-affinity binding. X-Ray and NMR structures of AR proteins have offered structural insights into the molecular basis for their wide variety of biological functions. Molecular-dynamics (MD) simulation studies may provide a more detailed, structural, dynamic, and energetic insight into their properties.

In a previous simulation study by *Yu et al.* [5], the highly homologous consensus-designed ankyrin repeat (AR) proteins E3_5 [6] and E3_19 [7] (*Fig. 1*) were studied using MD simulations with explicit water. CD Experiments [6][8] had indicated that the protein E3_19 is significantly less stable than E3_5. According to the simulation trajectories, the difference in stability is mainly due to the difference in stability of the C-terminal capping AR, while the proteins have similar properties for the internal ARs [5] . Analysis of the charge redistribution when mutating E3_5 into E3_19 reveals that the third internal AR, which is spatially closest to the C-terminal capping AR, becomes more negatively charged. This explains the unfolding of the C-terminal capping in the MD simulation of E3_19. The study of *Yu et al.* illustrates the complementarity between experiment and simulation when designing proteins with specific properties [5].

Simulation studies offer detailed insight into energetic and structural properties of proteins in solution that are not accessible through experiment. This makes design suggestions possible.

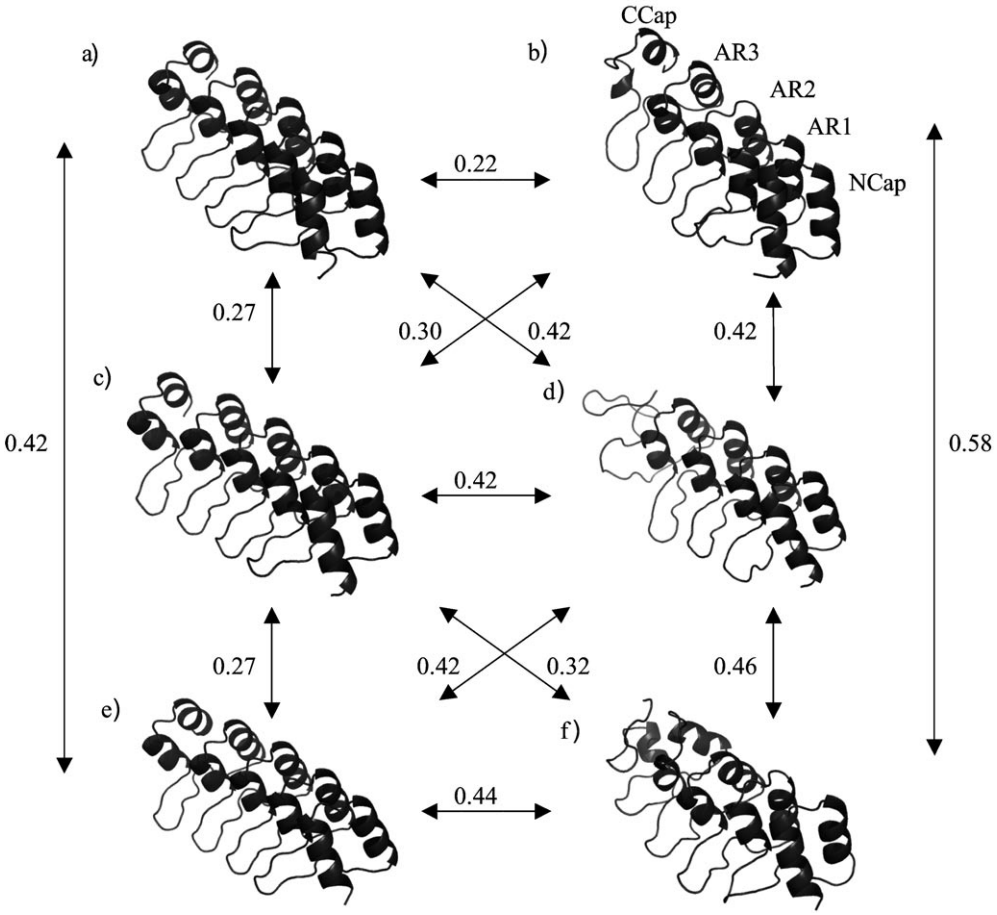

Fig. 1. *X-Ray crystal structures of the proteins E3_5* [6] (PDB: 1MJ0) (*a*), *E3_19* [7] (PDB: 2BKG) (*c*), *N3C* [9] (*e*). *Structures after 12-ns MD simulation: E3_5* (*b*), *E3_19* (*d*), *N3C* (*f*). Backbone (C_α, N, C) atom-positional RMS differences in nm between the different structures are shown at the black arrows. The least squares superposition of structures involved all backbone (C_α, N, C) atoms.

In the present study, a third AR protein, N3C, is investigated using MD simulation in explicit water. N3C is again highly homologous to the previously studied two proteins, E3_5 and E3_19 (87 and 88% sequence identity, resp., *Table 1*). The crystal structures of these two proteins have a backbone (N, C_α, C) atom-positional root-mean square difference of 0.27 nm (*Fig. 1*). Structurally, N3C is closer to E3_19 than to E3_5, but the N3C structure seems experimentally to be the most stable of the three, while E3_19 being the least stable. The latter finding might be explained by the observation that N3C has the lowest overall charge ($-12\,e$) compared to that of the other two

proteins ($-16\,e$), and its charge is moreover more evenly distributed over the five AR units. Here, we investigate the relative stability of the three proteins by MD simulation in order to obtain insight into the features determining protein stability. A MD simulation of N3C in aqueous solution was carried out, and the obtained trajectory was compared to those of the proteins E3_5 and E3_19, which were described by *Yu et al.* [5]. In experimental unfolding experiments, the unfolding of the entire protein was measured [5], here the stability of the AR units is investigated only.

Table 1. *Mutations That Distinguish the Proteins E3_5, E3_19, and N3C from Each Other*

Residue	E3_5	E3_19	N3C
33	Thr	Glu	Lys
35	Asn	Thr	Lys
36	Asp	Tyr	Asp
38	Tyr	Asp	Tyr
46	Ser	Arg	Arg
47	Asn	Val	Glu
59	Asn	Asn	Ala
66	Ser	Leu	Lys
68	Leu	Phe	Lys
69	Thr	Ser	Asp
71	Ile	Ser	Tyr
79	Ala	Lys	Arg
80	Thr	Arg	Glu
92	His	Tyr	Ala
99	Tyr	Asp	Lys
101	Asn	Thr	Lys
102	Asp	Ile	Asp
104	His	Ser	Tyr
112	Lys	Asp	Arg
113	Tyr	Thr	Glu
125	His	Tyr	Ala
156	Gln	–	Gln

Material and Methods. – *Molecular Dynamics Simulations.* MD Simulations were performed with the GROMOS software [10][11] using the force-field parameter set 45A3 [12][13]. The simulations of N3C, and of E3_5 and E3_19 by *Yu et al.* [5] are summarized in *Table 2*. Initial coordinates for N3C were taken from the X-ray structure of N3C (PDB: 2QYJ) [9]. The mutation sites that distinguish the three structures E3_5, E3_19, and N3C from each other are listed in *Table 1*. Ionization states of residues were assigned according to a pH of 8.0. The histidine side chains were protonated at the N_ε-atom. The simple-point-charge (SPC) water model [14] was used to describe the solvent molecules. In the simulations, H_2O 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 walls of the periodic box. Since, in the E3_5 and E3_19 simulations, no counterions were included [5], we did not include them in the N3C simulation for reasons of comparisons. Moreover, we did perform a 3-ns simulation of N3C including 12 Na^+ ions, which gave essentially the same results as the one without counterions. All bonds were constrained with a geometric tolerance of 10^{-4} using the SHAKE algorithm [15]. A steepest-descent energy minimization of the system was performed to relax the solute–solvent contacts, while positionally restraining the solute atoms using a harmonic interaction with a force constant of $2.5 \cdot 10^4$ 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 pairlist. 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 [16] 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 [17]. 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 $0.7513 \cdot 10^{-3}$ (kJ mol^{-1} $nm^{-3})^{-1}$. 100 ps of MD simulation with harmonic position restraining of the solute atoms with a force constant of $2.5 \cdot 10^4$ kJ mol^{-1} nm^{-2} were performed to further equilibrate the systems. The simulations E3_5, E3_19, and N3C were each carried out for 12 ns. The trajectory coordinates and energies were saved every 0.5 ps for analysis.

Table 2. *Overview of the Three 12-ns MD Simulations of the Different Systems.* E3_5: Protein E3_5 in aqueous solution, starting from the X-ray structure (PDB ID: 1MJ0); E3_19: protein E3_19 in aqueous solution, starting from the X-ray structure (PDB ID: 2BKG). N3C: protein N3C in aqueous solution, starting from the X-ray structure [9].

Simulation label	E3_5	E3_19	N3C
Protein	E3_5	E3_19	N3C
Starting structure	PDB ID: 1MJ0	PDB ID: 2BKG	PDB ID: 2QYJ
No. of H_2O molecules	9522	8790	15324
Total charge [*e*]	−16	−16	−12

Analysis. Analyses were performed with the analysis software GROMOS++ [18] and esra [19]. Atom-positional root-mean-square differences (RMSDs) between 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. Atom-positional root-mean-square fluctuations (RMSFs) over a period of simulation were calculated by performing a rotational and translational atom-positional least-squares fit of the trajectory structures on the reference structure (usually the first structure of the period) using a given set of atoms. The secondary structure assignment was achieved using the program DSSP, based on the *Kabsch–Sander* rules [20]. The percentages of intramolecular (n, $n-4$) H-bonds that are involved in the formation of α-helices have been calculated using a maximum distance criterion of 0.25 nm between the H-atom and the acceptor atom, and a minimum donor–H-acceptor angle criterion of 135°.

Results. – The atom-positional RMSDs from the starting structures for the atoms (N, C_α, C) in the simulations E3_5, E3_19, and N3C are shown in *Fig. 2* and have been calculated for all residues (solid lines) and for the three internal AR residues only (dotted lines). The protein E3_5 remains close to its crystal structure and converges to an RMSD of *ca.* 0.25 nm. The simulations E3_19 and N3C do not converge to a constant value within the 12 ns of simulation, the protein N3C showing larger structural rearrangements than the protein E3_19. The RMSDs of the internal ARs reach a value of 0.1 nm after *ca.* 2 ns and remain constant throughout the whole simulation period. The stability of the internal repeat units proves that the structural rearrangements occur either in the N- or C-terminal capping units.

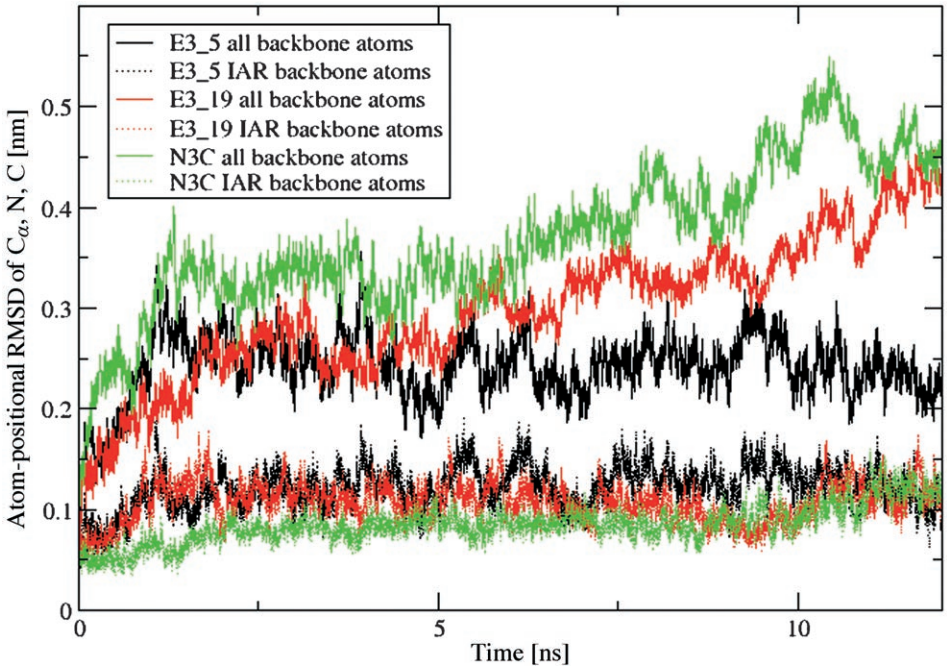

Fig. 2. *Atom-positional root-mean-square deviations* (RMSDs) *of the backbone atoms* (N, C_α, C) *of E3_5* (*black*), *E3_19* (*red*), *and N3C* (*green*) *with respect to the starting* (X-ray) *coordinates in the simulations.* The dotted lines represent the backbone RMSDs for the three internal AR (IAR) units of the corresponding protein. The least-squares superposition of structures involved all backbone (C_α, N, C) atoms (*solid lines*) and those of the three IAR units (*dotted lines*).

The atom-positional RMSFs for the C_α-atoms were calculated for the entire 12 ns of the trajectories (*Fig. 3*). All simulations show reasonably small fluctuations for the internal AR units as well as for the N-terminal capping unit. Larger fluctuations are observed only in the C-terminal capping unit. This indicates that the latter unit is relatively unstable.

The secondary structure assignment in *Fig. 4* shows that the helical structural elements are very stable for the simulation E3_5. In the simulation E3_19, secondary structure features of the C-terminal capping unit vanish completely after 5 ns. In the simulation N3C, the secondary structure of the C-terminal AR unit is still in place for the majority of the residues. According to these secondary-structure time histories, the elevated RMSDs and RMSFs originate from movement of or between the two helices of the C-terminal capping repeat and less from destruction of its helical secondary structure.

Backbone (n, $n-4$) H-bond occurrences for the C-terminal capping unit are tabulated in *Table 3*. Only the H-bonds involved in the formation of an α-helix are shown. The (n, $n-4$) H-bonds support the statement drawn from the secondary-structure analysis: secondary structure of E3_5 is stable, the C-terminal capping in

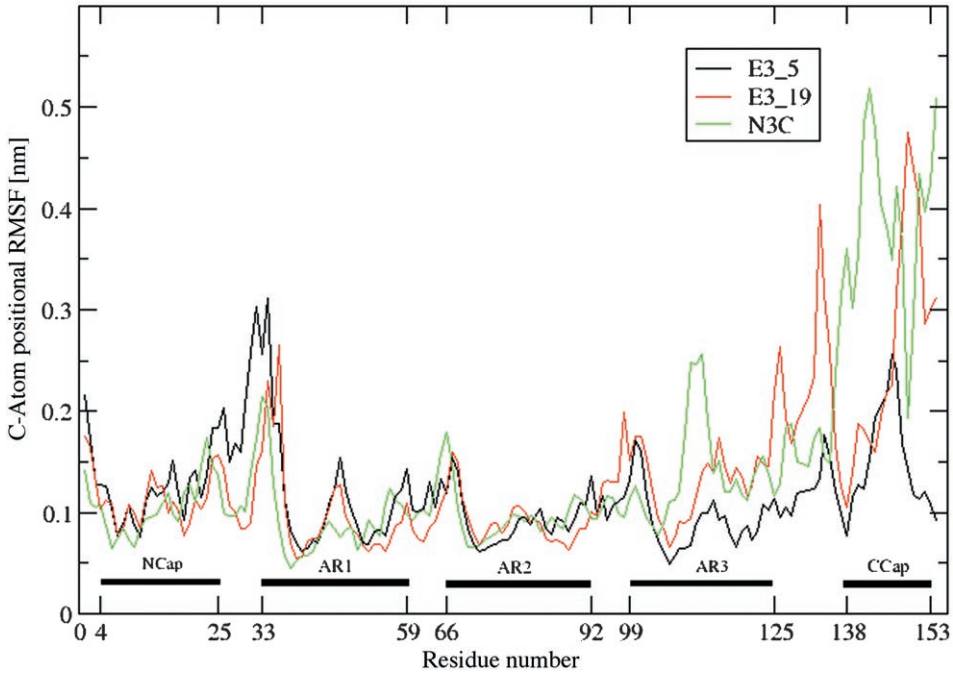

Fig. 3. *Atom-positional root-mean-square fluctuations* (RMSFs) *of the* C_α*-atoms of E3_5* (*black*), *E3_19* (*red*), *and N3C* (*green*) *over 12-ns of simulation.* AR Units are indicated by horizontal black bars. The least squares superposition of structures involved all C_α-atoms.

E3_19 unfolds, and N3C retains most of its helical features, though at somewhat lower percentages.

Final structures from the simulations are depicted in *Fig. 1*. The unfolding of the C-terminal AR unit in the simulation E3_19 is clearly seen; in the two other simulations, parts of this secondary structure are still visible. The backbone atom-positional RMSD values for the various pairs of X-ray and simulated structures show that structurally N3C is closer to E3_19 than to E3_5.

In short, we can conclude that the structures of E3_5 and N3C both are reasonably stable throughout the simulation periods. The structure of the protein E3_19 proves to be unstable in agreement with experiment.

Discussion. – This work presents the results of a MD simulation of the ankyrin repeat protein N3C using the GROMOS force-field parameter set 45A3. Root-mean-square deviations and fluctuations, as well as H-bond analysis and secondary-structure assignments have been performed to evaluate the protein's stability. These data were compared with the data of two previously simulated proteins, E3_19 and E3_5, in order to be able to investigate the stabilities of AR proteins as a function of charge distribution over the repeat units. The protein E3_5 proves to be the most stable of the simulated proteins, followed by the protein N3C, which loses a small amount of secondary structure during the 12 ns of simulation, while E3_19 loses all secondary

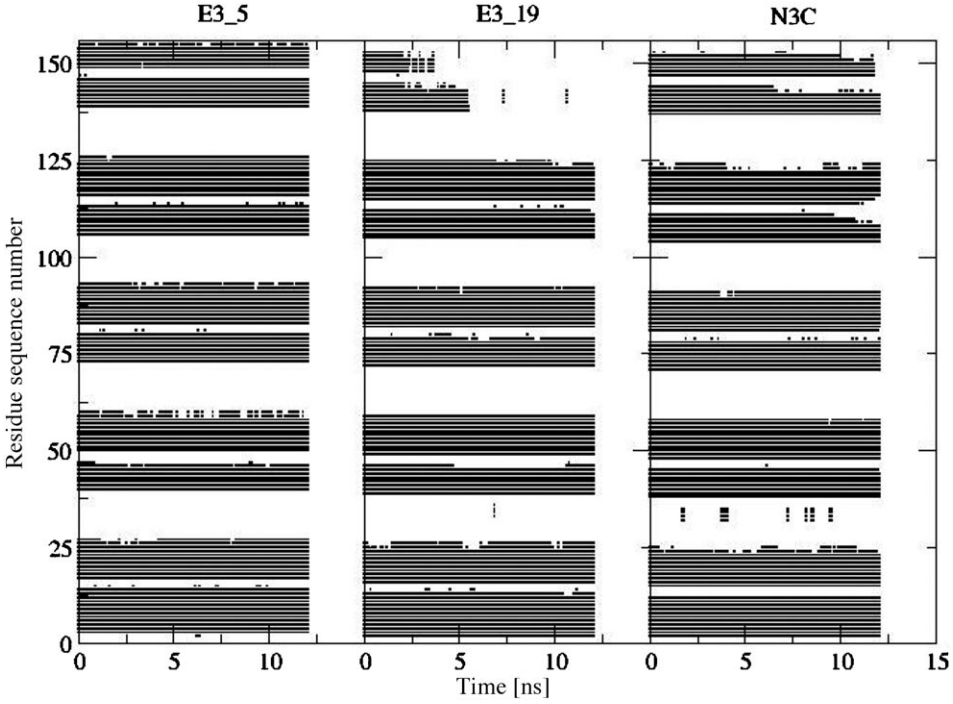

Fig. 4. *Secondary structure elements of E3_5* (*left*), *E3_19* (*middle*), *and N3C* (*right*). α-Helix (*black*). Assignments according to [20].

Table 3. *Occurrence* [%] *of all (*n, n – *4) Backbone H-Bonds* (*n* is residue sequence number) *from the 12-ns MD Simulations*

Residue number		Occurrence [%] of H-bonds		
Donor (NH)	Acceptor (O)	E3_5	E3_19	N3C
142	138	88	9	74
143	139	87	13	74
144	140	78	13	74
145	141	81	31	38
146	142	77	11	–
147	143	64	–	–
148	144	76	–	–
149	145	–	–	–
150	146	–	–	89
151	147	–	8	79
152	148	76	18	61
153	149	82	–	64
154	150	52	10	–
155	151	53	–	–

structure in its C-terminal repeat unit. Yet, structurally N3C is closer to E3_19 than to E3_5. These two observations from MD simulation confirm analoguous experimental findings [6–9].

Considering the hypothesis that a more or less uniform distribution of the charges among the AR units or a reduction of the total charge of the last internal AR unit would increase stability, one would expect that the protein N3C would be very stable in solution (*Table 4*), even more stable than the protein E3_5. This is not the case, although the protein retains most of its helical features in the C-terminal capping AR.

Table 4. *Charge Distribution over the Various ARs in the Proteins E3_5, E3_19, and N3C* (in *e*). The atomic partial charges were taken from the GROMOS force-field parameter set 45A3 [12][13]. NCap: residues 1–32; AR1: residues 33–65; AR2: residues 66–98; AR3: residues 99–131; CCap: residues 132–156.

	NCap	AR1	AR2	AR3	CCap	Total
E3_5	–1	–4	–3	–3	–5	–16
E3_19	–1	–4	–1	–5	–5	–16
N3C	–1	–2	–2	–2	–5	–12

Other factors than charge–charge interactions between AR units must be responsible for the higher stability of E3_5 compared to N3C. Comparing the amino-acid sequences of the three proteins, we find 21 differences between E3_5 and E3_19, 18 differences between E3_5 and N3C, and again 21 differences between N3C and E3_19. This suggests more similarity between N3C and E3_5 than with E3_19. Comparing the bulkiness of the side chains, N3C is more like E3_19 for AR1 and AR2, and lies in between E3_5 and E3_19 for AR3. This suggests that charge considerations are not sufficient to explain protein stability, but that other factors such as polarity and volume of side chains should also be accounted for.

Moritz Winger acknowledges *Haibo Yu* for the MD simulations of the proteins E3_5 and E3_19, and *Daniel Trzesniak* for technical support. The designed ankyrin repeat proteins come from a library generated by *Andreas Plückthun*, the crystal structure of the repeat protein N3C was provided by *Markus G. Grütter*, *Tobias Merz*, and *Per R. E. Mittl*. Financial support was obtained from the *Swiss National Science Foundation* (*M. G. G.*) and through the *National Center of Competence in Research* (*NCCR*) *Structural Biology* of the *Swiss National Science Foundation*, which is gratefully acknowledged.

REFERENCES

[1] P. Bork, *Proteins* **1993**, *17*, 363.
[2] S. G. Sedgwick, S. J. Smerdon, *Trends Biochem. Sci.* **1999**, *24*, 311.
[3] P. J. Mohler, A. O. Gramolini, V. Bennett, *J. Cell. Sci.* **2002**, *115*, 1565.
[4] L. K. Mosavi, T. J. Cammet, D. C. Desrosiers, Z. Y. Peng, *Protein Sci.* **2004**, *13*, 1435.
[5] H. Yu, A. Kohl, H. K. Binz, A. Plückthun, M. G. Grütter, W. F. van Gunsteren, *Proteins* **2006**, *65*, 285.
[6] A. Kohl, H. K. Binz, P. Forrer, M. T. Stumpp, A. Plückthun, M. G. Grütter, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1700.
[7] H. K. Binz, A. Kohl, A. Plückthun, M. G. Grütter, *Proteins* **2006**, in press.
[8] H. K. Binz, M. T. Stumpp, P. Forrer, P. Amstutz, A. Plückthun, *J. Mol. Biol.* **2003**, *332*, 489.
[9] T. Merz, S. K. Wetzel, S. Firbank, A. Plückthun, M. G. Grütter, P. R. E. Mittl, *J. Mol. Biol.* **2008**, *376*, 232.

[10] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, 'Biomolecular Simulation: The GROMOS Manual and User Guide', vdf Hochschulverlag, ETH Zürich, Switzerland, 1996.
[11] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, W. F. van Gunsteren, *J. Phys. Chem. A* **1999**, *103*, 3596.
[12] X. Daura, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **1998**, *19*, 535.
[13] L. D. Schuler, X. Daura, W. F. van Gunsteren, *J. Comput. Chem.* **2001**, *22*, 1205.
[14] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, in 'Intermolecular Forces', Ed. B. Pullman, Reidel, Dordrecht, 1981; pp. 331 – 342.
[15] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.
[16] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys.* **1995**, *102*, 5451.
[17] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684.
[18] M. Christen, P. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D. Geerke, T. Heinz, M. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, W. van Gunsteren, *J. Comput. Chem.* **2005**, *26*, 1719.
[19] V. Kräutler, M. Kastenholz, P. H. Hünenberger, available at http://sourceforge.net/project/esra, 2005.
[20] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577.

Received April 11, 2008