Solvent environment conducive to protein aggregation

Ariel Fernández[a,b,1,*], Maria de las Mercedes Boland[c]

[a] *Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA*
[b] *Institute for Protein Research, Osaka University, Yamadaoka 3-2, Suita, Osaka 565-0871, Japan*
[c] *100 Morningside Drive, New York, NY 10027, USA*

Received 15 August 2002; accepted 28 August 2002

First published online 16 September 2002

Edited by Thomas L. James

Abstract The effect of solvent structuring induced by molecular crowding is elucidated within a competitive situation involving protein folding and aggregation. Two patterned fragments of amyloidogenic proteins are chosen as study cases and analyzed by molecular dynamics with an implicit treatment of the solvent. The extent of crowding needed to induce aggregation is determined. The results constitute a first step to assess the relevance of in vivo environments in understanding fibrillogenesis. The approach is independently validated by satisfactorily reproducing the results of an all-atom explicit solvent trajectory.

Key words: Solvent environment; Molecular crowding; Protein aggregation; Protein folding

1. Introduction

A proper assessment of complex solvent environments is necessary to understand in vivo folding, aggregation and their interconversion [1–5]. Thus, disease-related fibrillogenesis often occurs at high protein concentration [6–8], in environments poorly understood beyond the excluded volume and confinement effect [1,5,9]. Here we delineate the role of complex environments on the stabilization of intermolecular hydrogen bonds, inducing the aggregation of fragments of amyloidogenic proteins [5,8,10]. The study involves all-atom molecular dynamics simulations with an implicit-solvent treatment needed to make relevant timescales computationally accessible. The treatment accounts for conformation-dependent solvent polarizability [11–13] and exogenous solvent-structuring perturbations arising from crowding conditions. We show that certain peptides that adopt ephemeral β-hairpin conformations as they approach the solvent-structuring moieties, eventually undergo a transition from the folded state to an antiparallel aggregate. Our simulations are independently validated by contrasting our results on *villin headpiece* folding in vitro with the longest available all-atom simulation [14].

2. Results

The enhancement of dielectric-dependent intramolecular interactions depends on the partial structuring, immobilization or removal of surrounding water. Thus, nearby hydrophobes induce a structuring of the solvent needed to create a cavity around them and the net effect of this structuring is a localized reduction of the solvent polarizability and consequently, a lowering in the local dielectric coefficient ε. This effect has been quantified in recent work [11–13], where we delineated the role of hydrophobic clustering in the enhancement of dielectric-dependent intramolecular interactions. Specifically, we determined ε as a function of the positions $\{\mathbf{r}_j, j=1, 2,...,n\}$ of surrounding carbonaceous groups (CH_i, $i=1,2,3$), which might be external or part of the chain. Thus, if we place a charge at the center of (intrinsic) coordinates and a partner charge at position $\mathbf{r}$, we get (Section 3):

$$\varepsilon = [(\varepsilon_0^{-1}-\varepsilon_w^{-1})\Omega(\{\mathbf{r}_j\})(1+r/\xi)\exp(-r/\xi)+\varepsilon_w^{-1}]^{-1} \quad (1)$$

where ε_w, ε_0 are the permittivities of bulk water and vacuum, respectively, $\xi=5$ Å is a water dipole–dipole correlation length and

$$\Omega(\{\mathbf{r}_j\}) = \prod\nolimits_{j=1,...,n}[1+\exp(-\|\mathbf{r}_j\|/\Lambda)][1+\exp(-\|\boldsymbol{r}-\boldsymbol{r}_j\|/\Lambda)] \quad (2)$$

where $\|\cdot\|$ denotes vector norm and $\Lambda=1.8$ Å, the characteristic length associated with the water-structuring effect induced by the solvent organization around the hydrophobes.

Eqs. 1 and 2 reveal that the spatial distribution of the hydrophobic groups around the charge is responsible for an enhancement of the electric field when their positions lie at distances comparable to the characteristic length ξ. Thus, the equations quantify the local inhibition of solvent polarizability resulting in a net decrease in electrostatic shielding. This effect is illustrated by the stabilization of pre-formed amide-carbonyl H-bonds upon desolvation of the backbone [11–13,15].

Just as the electrostatic modulation due to hydrophobic clustering becomes thus quantifiable in our implicit-solvent treatment, the hydrophobic interactions should also be rescaled in a consistent way, taking into account their dependence on the extent of burial [13]. Thus, the energetic cost of exposing a hydrophobic group to a solvent with dielectric coefficient ε is obtained by multiplying the cost of exposing it to bulk solvent [16] by the rescaling factor $\varepsilon/\varepsilon_w$.

All-atom molecular dynamics simulations with implicit solvent at $T=316$ K were performed using a modified version of the TINKER 3.7 molecular design package [17] and adopting the AMBER95 (parm94) force field [18]. The electrostatic shielding and local dielectric coefficient computed using Eqs.

*Fax: (1)-773-702 0439.
E-mail address: ariel@uchicago.edu (A. Fernández).

[1] On leave from: Instituto de Matemática, Universidad Nacional del Sur, CONICET, Bahía Blanca 8000, Argentina.

0014-5793/02/$22.00
PII: S0014-5793(02)03392-6

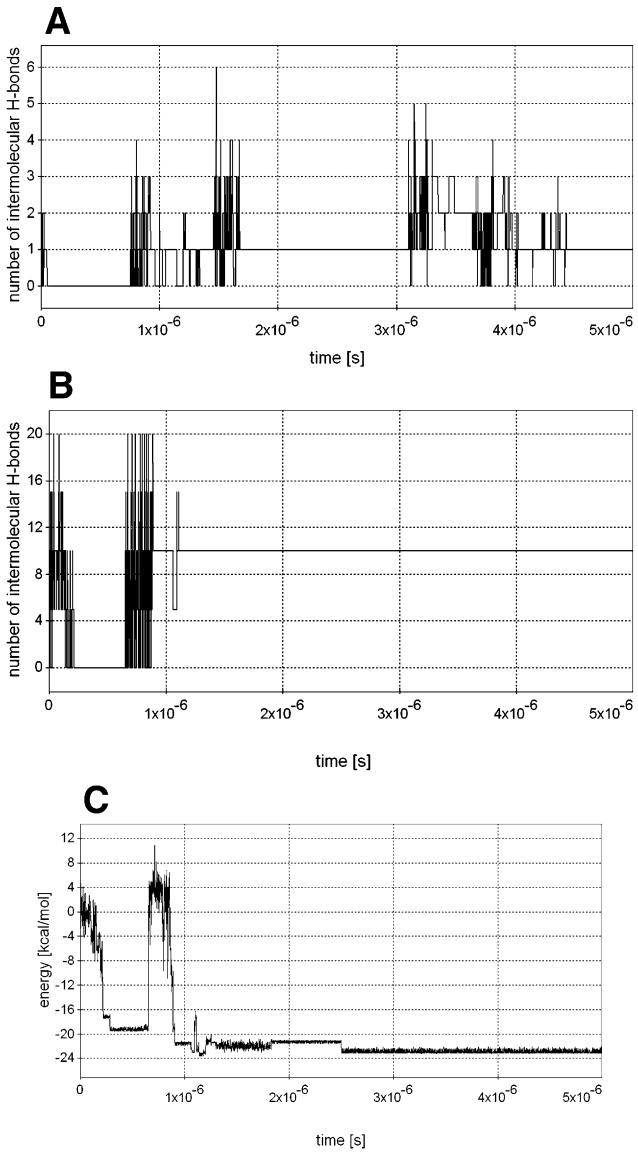

Fig. 1. A: Time-dependent number of intermolecular backbone H-bonds from a typical MD run for two identical chains of the core-modified 12–24 fragment of the Aβ-peptide under non-crowding conditions ($d = 8$). B: Idem, but subject to crowding conditions ($d = 25$ Å). C: Energy profile for the run specified in B.

1 and 2, was evaluated at each site of the chain and at each folding step (every 20 fs). The solvent-accessible surface area potential [16] was modified to incorporate the solvent-structuring effects requiring the computation of the local factor $\varepsilon/\varepsilon_w$. We used the Pastor–Karplus [19] scheme for calculating the atomic friction coefficients, with a water viscosity of $\eta = 0.84$ cp. Our implicit-solvent model was independently validated by comparing our results on *villin headpiece* folding with the all-atom Duan–Kollman trajectory [14] (Section 3).

The core fragment $_{12}$VHHQKLVFFAEDV$_{24}$ from the amyloidogenic Aβ-peptide [6] was first chosen for study. We noticed that this fragment is palindromic with regard to hydrophobic/polar complementarity if we take into account the dual polar/hydrophobic nature of both K and Q. Thus, an antiparallel β-sheet aggregate would be possible if it were not for the fact that the lysine charge would be buried [20]. To prevent this adverse effect and still warrant the solubility of the peptide, the mutation K16Q was introduced in silico.

The folding of two identical independent chains was studied under controlled conditions at neutral pH in two different solvent contexts: (A) The chains fold in bulk solvent, and the local dielectric environments are determined intra- and intermolecularly by the chains as they fold. (B) Two external solvent-structuring perturbations are introduced by placing two external hydrophobes at a distance $d = 25$ Å from each other.

Thirty runs composed of 9×10^7 steps each were performed under situation A, each with a new random choice of initial conformation. No reproducible or sustainable folding or aggregation pattern was observed. To assess the extent of aggregation, the number of intermolecular amide–carbonyl H-bonds was displayed in Fig. 1A as a function of time for the run that presented the highest persistence of dimolecular association. A single intermolecular H-bond prevailed for extended periods (~ 1 μs) but no supramolecular pleated structure occurred. An N–H···O H-bond satisfies the following operational constraints [12]: (I) N–O distance shorter than 3.45 Å; (II) N–H–O angle within 60° latitude.

By contrast, situation B yielded a sustainable aggregation, revealed in 25 out of 33 runs, each made up of 9×10^7 steps. The remaining runs yielded no sustainable structure and were not reproducible. The most reproducible run is displayed in Figs. 1B,C and 2. Fig. 1B reveals that 10 stable intermolecular H-bonds are formed after an induction period $\tau = 1.24$ μs. The induction period ends with large fluctuations reflecting a competition between aggregation and intramolecular folding. The metastable folded conformation is sustained during the period 3×10^{-7} s $< t < 6.5 \times 10^{-7}$ s and is only 4 kcal/mol less stable (Fig. 1C) than the antiparallel aggregate. Nevertheless, the fold → aggregate conversion appears to be irreversible (Fig. 1C), in accordance with known facts on amyloid nucleation [6]. This implies that the energy basin corresponding to the aggregated state is far broader, or has far larger microcanonical entropy [13], than that of the folded state.

The three most representative conformations are displayed in Fig. 2 using all-atom and ribbon backbone descriptions. The large spheres centered at the external hydrophobes represent regions of maximum solvent structuring defined by the inflection point in the exponential decay of the solvent-structuring effect (Section 3). Fig. 2a shows the metastable state ($t = 5 \times 10^{-7}$ s), where each chain is independently folded and they mutually stabilize their backbone H-bonds by reciprocal desolvation [12,15]. Fig. 2b is a snapshot of the system taken at 8×10^{-7} s, precisely when the dramatic enhancement of structural fluctuations (Fig. 1B) signals the dominant barrier crossing. This conformation encapsulates the competition between folding and aggregation. Fig. 2c displays a snapshot taken at 3 μs, revealing an intermolecular antiparallel pleated state of high stability achieved essentially by a higher entropic content relative to the intramolecularly folded state (cf. Fig. 1C).

The folded → aggregate transition is invariably irreversible and completed after an induction period τ ranging from 1.1 to 1.5 μs on all reproducible runs. The transition results from a combination of two factors: (a) the weakening of the hydrophobic effect (the free energy cost for exposing hydrophobes to an inherently more structured solvent becomes lower); and (b) the concurrent enhancement of the H-bond stability

brought about by the externally induced decrease in solvent polarity ($\varepsilon_w \rightarrow \varepsilon$). Thus, in generic terms, while the hydrophobic effect contribution decreases by the local factor $\varepsilon/\varepsilon_w$, the H-bonds are enhanced by the reciprocal factor $\varepsilon_w/\varepsilon$.

The results are robust to moderate changes in the separation d between the two external solvent-structuring moieties, which may be regarded as a crowding parameter. Thus, if the hydrophobes are placed beyond a threshold distance $d = d^* = 30$ Å, no reproducible or sustainable aggregation is observed. Furthermore, repeated runs with different values of the crowding parameter reveal that the induction period τ on successful runs yielding the aggregate fits an empirical exponential:

$$\tau \sim \exp[12\ \text{Å}/(d^* - d)],\ 16\ \text{Å} < d < d^* \quad (3)$$

Similar results were obtained for the fragment $_{21}$NFLNCYVSGFH$_{31}$ contained in the so-called K3 fragment of β2-microglobulin (residues 20–41), known to be amyloidogenic under non-extreme conditions [21]. This peptide was chosen for its computational tractability and for its near-palindromic h/p complementarity which holds except for an L–G mismatch. The results obtained under conditions identical to those of Fig. 2 are displayed in Fig. 3a–c, showing three snapshots taken respectively at 0.8, 1.4 and 4 μs. The intramolecularly folded conformation (Fig. 3a) is now less stable, with an average lifetime of 0.1 μs, due to the higher inability of the shorter peptide to internally protect its H-bonds from water attack. The large structural fluctuations needed to overcome the energy barrier that separates the folded from the aggregated state extend now over a longer time period: 1.5 μs, as befits a smaller system than that previously studied. The overall induction period for aggregation is sensibly larger than in the previous study case: 2.3 μs. However, the aggregation kinetics, the competition between aggregation and folding, the inductive role played by the external hydrophobes and the critical crowding parameter (now $d^* = 29.2$) are all very similar.

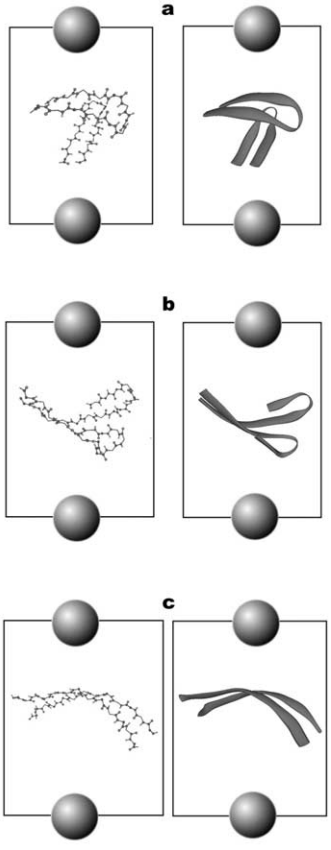

Fig. 2. a–c: Three snapshots of the representative run described in Fig. 1B,C taken respectively at 0.5, 0.8 and 3 μs.

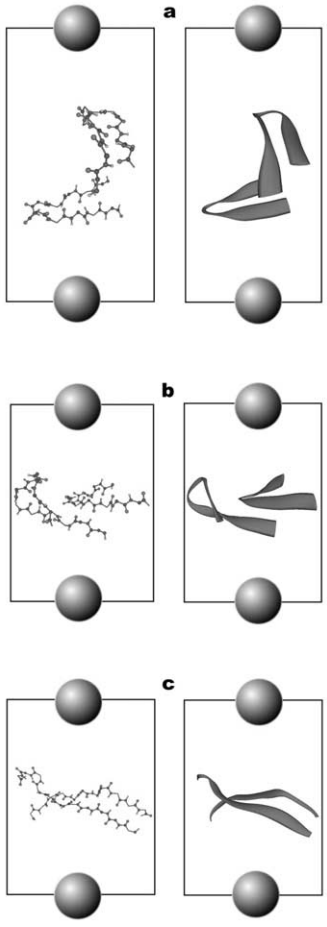

Fig. 3. a–c: Three snapshots taken respectively at 0.8, 1.4 and 4 μs for two chains of the 21–31 fragment of β2-microglobulin taken from a typical run under conditions identical to those of Figs. 1B,C and 2.

These results illustrate the influence of crowding conditions interplaying with the local contexts determined by the peptide chains as they adopt different conformations. The local environments evolve providing a route for denaturation of intramolecular folds and concurrently induce the stabilization of aggregated states. Thus, this work delineates the role of complex in vivo contexts in nucleating amyloidogenic aggregation and offers a means of dealing computationally with such environments.

3. Methods

3.1. Hydrophobic clustering effect on electrostatic contributions

The aim of this section is to outline the effect of desolvation on the modulation of electrostatic interactions which take place as proteins

fold or aggregate. To investigate a specific interaction, an intrinsic coordinate system is needed. Thus, we define such a system by placing the charge q at its center, and a test charge along the x-axis. We assume the electrostatic pairwise interaction to be surrounded by n spherical hydrophobes (carbonaceous groups) centered at positions $\mathbf{r}_j$, $j=1, 2,...,n$. Since the solvent makes cavities around them, the hydrophobes, whether part of the chain itself or external to it, are treated as solvent-structuring entities.

To evaluate the electric field $E(\mathbf{r})$ at position $\mathbf{r}$, we replace the position-dependent dielectric [22–25] by an integral kernel. This kernel is convoluted with the electric field at position $\mathbf{r}$ to represent the correlations with the field at neighboring positions. This prompts us to introduce the equation:

$$\mathrm{div}[\int K(\mathbf{r},\ \mathbf{r}',\ \{\mathbf{r}_j\})E(\mathbf{r}')\mathrm{d}\mathbf{r}'] = 4\pi q\delta(\mathbf{r}) \tag{4}$$

where the kernel $K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\})$ is parametrically dependent on the fixed hydrophobe positions. In the absence of vicinal hydrophobic units, the correlations decay as $\exp(-\|\mathbf{r}-\mathbf{r}'\|/\xi)$ (ξ = characteristic correlation length). In the limit $\xi\to 0$, we get: $K(\mathbf{r}, \mathbf{r}')\sim\delta(\mathbf{r}'-\mathbf{r})$, and thus Eq. 4 becomes the Poisson equation.

The correlation kernel may be identified taking into account the relationship between diffraction and dielectric. For bulk water, we get $K(\mathbf{r}, \mathbf{r}') = K(\mathbf{r}-\mathbf{r}')$ by inverse-transforming its frequency $\mathbf{k}$-vector representation:

$$K(\mathbf{r}-\mathbf{r}') = \int \exp[\mathrm{i}\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')]\ L(\mathbf{k})\mathrm{d}\mathbf{k} \tag{5}$$

where the distribution $L(\mathbf{k}) = [(\varepsilon_w-\varepsilon_0)/(1+\varepsilon_w\|\mathbf{k}\|^2\xi^2/\varepsilon_0)+\varepsilon_0]$ reflects the fact that, due to dipole-reorientation inertia, water becomes only significantly polarized when interacting with low-frequency radiation.

To obtain the correlation kernel with n hydrophobic units at fixed positions, we modify Eq. 5 to incorporate their solvent-structuring effect:

$$K(\mathbf{r},\ \mathbf{r}',\ \{\mathbf{r}_j\}) = [\int \exp[\mathrm{i}\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')]L(\mathbf{k})\mathrm{d}\mathbf{k}]\times[1+\sum\nolimits_{j=1,...,n}\Gamma_j(\mathbf{r},\mathbf{r}')] \tag{6}$$

where $\Gamma_j(\mathbf{r}, \mathbf{r}')\sim\exp[-(\|\mathbf{r}-\mathbf{r}_j\|+\|\mathbf{r}'-\mathbf{r}_j\|)/\Lambda]$. Here we introduce the characteristic length Λ to assess the water-structuring effect.

We now solve Eq. 4 using Eq. 6 by Fourier transformation and inversion:

$$\int E(\mathbf{r})\mathrm{d}\mathbf{r} = -(4\pi q)\int \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})\|\mathbf{k}\|^{-2}[K(\mathbf{k},\ \{\mathbf{k}_j\})]^{-1}\mathrm{d}\mathbf{k} \tag{7}$$

$$E(\mathbf{r}) = (q/r^2)[(\varepsilon_0^{-1}-\varepsilon_w^{-1})\Omega(\{\mathbf{r}_j\})(1+r/\xi)\exp(-r/\xi)+\varepsilon_w^{-1}] \tag{8}$$

where: $\Omega(\{\mathbf{r}_j\}) = \Pi_{j=1,...,n}[1+\exp(-\|\mathbf{r}_j\|/\Lambda)][1+\exp(-\|\mathbf{r}-\mathbf{r}_j\|/\Lambda)]$.

The effective permittivity given in Eq. 1 results directly from Eq. 8 and Coulomb's law.

3.2. Validating the implicit-solvent model against the longest all-atom computation

A 1 μs trajectory for *villin headpiece* ($N=35$) was generated by molecular dynamics adopting the implicit-solvent model described above at conditions identical to those reported by Duan and Kollman [14]. These authors performed the longest all-atom simulation [26–28] so far reported, covering 1 μs of the folding process. In our run, the local dielectric coefficient was evaluated for each residue along the chain at every step by identifying the hydrophobic groups within a cutoff distance of 7.2 Å = 4Λ [11,12]. Beyond this distance, the solvent-structuring effect becomes negligible. The initial chain conformation is chosen to be identical to that adopted in the all-atom trajectory, from which solvent coordinates are truncated. The RMSD between each chain conformation produced by implicit-solvent ($x(t)$) and all-atom ($X(t)$) computation is obtained every 20 fs and displayed in Fig. 4. At the start ($t=0$) there is an immediate jump in RMSD to values comparable to their time average (5.3 Å). Clear upper and lower bounds at 3.0 and 7.6 Å can be established for at least 99.5% of the trajectory. Nearly identical bounds were found for six out of the nine runs performed for *villin headpiece* (the other three runs yielded non-reproducible results). Such tight agreement provides validation for our implicit-solvent model.

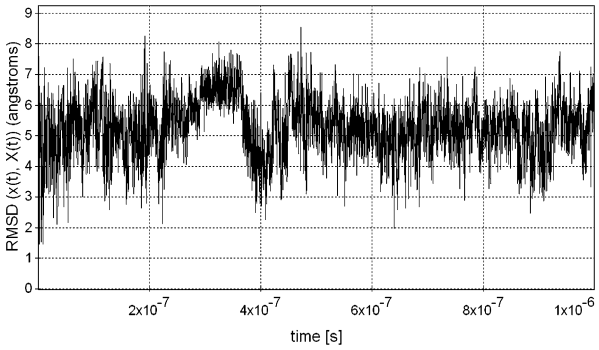

Fig. 4. RMSD between the respective chain conformations of *villin headpiece* generated by our typical implicit-solvent MD run ($x(t)$) and the all-atom Duan–Kollman trajectory ($X(t)$).

Acknowledgements: The author thanks Prof. Y. Duan for allowing him to access the Duan–Kollman trajectory [14] for the purpose of study.

References

[1] Minton, A.P. (1998) Methods Enzymol. 295, 127–149.
[2] Eaton, W.A. and Hofrichter, J. (1990) Adv. Protein Chem. 40, 63–279.
[3] Rivas, G., Fernández, J.A. and Minton, A.P. (2001) Proc. Natl. Acad. Sci. USA 98, 3150–3155.
[4] van den Berg, B., Wain, R., Dobson, C.M. and Ellis, R.J. (2000) EMBO J. 19, 3870–3875.
[5] Uversky, V.N., Cooper, E.M., Bower, K.S., Li, J. and Fink, A.L. (2001) FEBS Lett. 515, 99–103.
[6] Rochet, J.C. and Lansbury Jr., P.T. (2000) Curr. Opin. Struct. Biol. 10, 60–68.
[7] Gejyo, F., Yamada, T., Odani, S., Nakagawa, Y., Arakawa, M. and Kunitomo, T. (1985) Biochem. Biophys. Res. Commun. 129, 701–706.
[8] Hatters, D.M., Minton, A.P. and Howlett, G.J. (2002) J. Biol. Chem. 277, 7824–7830.
[9] Klimov, D.K., Newfiled, D. and Thirumalai, D. (2002) Proc. Natl. Acad. Sci. USA 99, 8019–8024.
[10] Hoshino, M., Katou, H., Hagihara, Y., Hasegawa, K., Naiki, H. and Goto, Y. (2002) Nat. Struct. Biol. 9, 332–336.
[11] Fernández, A. (2002) Phys. Lett. A 299, 217–220.
[12] Fernández, A., Sosnick, T.R. and Colubri, A. (2002) J. Mol. Biol. 321, 659–675.
[13] Fernández, A. (2001) J. Chem. Phys. 114, 2489–2502.
[14] Duan, Y. and Kollman, P.A. (1998) Science 282, 740–744.
[15] Vila, J.A., Ripoll, D.R. and Scheraga, H.A. (2000) Proc. Natl. Acad. Sci. USA 97, 13075–13079.
[16] Ooi, T., Oobatake, M., Nemethy, G. and Scheraga, H.A. (1987) Proc. Natl. Acad. Sci. USA 84, 3086–3090.
[17] Ponder, J.W., Rubenstein, S., Kundrot, C., Huston, S., Dudek, M., Kong, Y., Hart, R., Hodsdon, M., Pappu, R., Mooij, W. and Loeffler, G. (1999) TINKER: Software tools for molecular design, version 3.7, Washington University, St. Louis, MO.
[18] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz Jr., K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) J. Am. Chem. Soc. 117, 15179–15197.
[19] Pastor, R.W. and Karplus, M. (1988) J. Phys. Chem. 92, 2636–2641.
[20] Richardson, J.S. and Richardson, D.C. (2002) Proc. Natl. Acad. Sci. USA 99, 2754–2759.
[21] Kozhukh, G.V., Hagihara, Y., Kawakami, T., Hasegawa, K.,

Naiki, H. and Goto, Y. (2002) J. Biol. Chem. 277, 1310–1315.
[22] Warshel, A. and Russell, S.T. (1984) Q. Rev. Biophys. 17, 283–422.
[23] Petrey, D. and Honig, B. (2000) Protein Sci. 9, 2181–2191.
[24] Warshel, A. and Papazyan, A. (1998) Curr. Opin. Struct. Biol. 8, 211–217.
[25] Bryant, R.G. (1996) Annu. Rev. Biophys. Biomol. Struct. 25, 29–53.
[26] Brooks, B.R. et al. (1983) J. Comput. Chem. 4, 187–217.
[27] Daggett, V. and Levitt, M. (1992) Proc. Natl. Acad. Sci. USA 89, 5142–5146.
[28] Brooks III, C.L. and Case, D. (1993) Chem. Rev. 93, 2487–2502.