Theor Chem Acc (2012) 131:1229
DOI 10.1007/s00214-012-1229-4

LETTER

Identifying the intermediates during the folding/unfolding of protein GB1 with MD simulations

Xiaomin Wu · Gang Yang · Lijun Zhou

Received: 23 December 2011 / Accepted: 25 April 2012 / Published online: 11 May 2012

Abstract Protein folding/unfolding has long represented one of the considerable challenges. In this work, a total of 4.0-μs explicit solvent MD simulations were performed on protein GB1 through the gradual stretching of its end-to-end distance, identifying a wide range of key intermediates and thus connecting the denatured and native states. The so-obtained results agree well with the available experimental and computational reports. In addition, they provide for the first time many other significant folding events and a clear description of protein GB1 folding/unfolding. The folding initiates at the three turn regions. Then, the three secondary structure units begin shaping in tandem, and their folding events are intersected, whereas the hydrophobic core forms at the very late stage. The non-native contacts play a significant role at the early folding stages. Unexpectedly, the tertiary contacts occur rather early and increase concomitantly with the comprising secondary structure units, which largely determine the formation of tertiary contacts.

Keywords Protein folding · Molecular dynamics · End-to-end distance · Secondary structure · Tertiary contacts

Electronic supplementary material The online version of this article (doi:10.1007/s00214-012-1229-4) contains supplementary material, which is available to authorized users.

X. Wu · G. Yang (✉) · L. Zhou
Key Laboratory of Forest Plant Ecology, Ministry of Education, Northeast Forestry University, Harbin 150040, People's Republic of China
e-mail: theobiochem@gmail.com

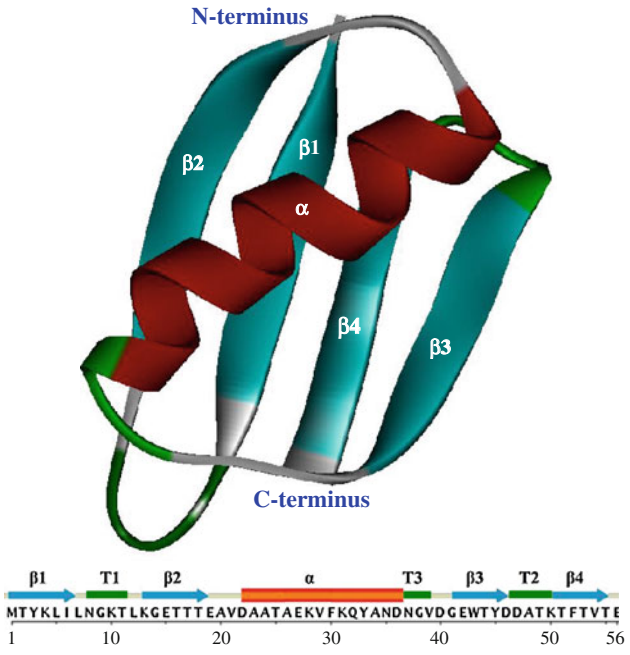
G. Yang
Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, People's Republic of China

Compared with the DNA information transfer, the mapping of proteins from the sequence to the native state is definitely more complicated, which is achieved by the so-called "folding". Protein folding has significant implications in multi-disciplines and represents one of the nature's challenges [1]. The global folding of proteins usually takes place in the order of milliseconds to seconds, whereas the underlying molecular events may occur on the nanosecond or even smaller timescales. Accordingly, the unambiguous descriptions of protein folding are not accessible to the routine experimental and molecular dynamics (MD) resolutions [2, 3].

Recently, the mechanical unfolding of proteins has often been studied using the atomic force microscopy (AFM) technique, which allows proteins to be stretched by different degrees and monitors the unfolding processes [4–6]. The B1 domain of streptococcal protein G (GB1) is regarded as a prototype to probe protein folding/unfolding mechanisms [3, 7–17]. As shown in Scheme 1, GB1 comprises three main secondary structure units, as the N-terminus β-hairpin (HP_N, Met1–Ala20), central α-helix (HE_M, Ala23–Asn35) and C-terminus β-hairpin (HP_C, Gly41–Glu56). Tertiary contacts exist between two neighboring secondary structure units. In this work, the mechanical unfolding of GB1 will be performed using the explicit solvent MD simulations, aiming to clarify its unfolding/folding mechanisms at a molecular level. The MD simulations start at the native state, then gradually stretch the end-to-end distance and finish at the completely denatured state. The adjustments of the end-to-end distances are rather gentle (about 2 Å between two neighboring MD runs), which avoids the rigorous structural transitions during the unfolding process. In this way, we present one clear and detailed pathway of GB1 unfolding/folding, which is found to be in good agreement with the

literature available. Note that the folding and unfolding of proteins are not necessarily reversible although assumed to be by most studies.

The time-evolution RMSD and R_g curves in Figs. S1 and S2 show that the MD simulations of all the end-to-end distances have been well equilibrated. The similar conformations during each MD run are clustered using the *x*RMSD method (Fig. S3). The formation degrees (F_D) of the secondary structure units along the unfolding/folding pathway are given in Fig. 1, where **Ix** ($x = 1$–10) refers to the intermediate state (Fig. 2). At the beginning, GB1 is in the completely denatured state (**U**, Fig. 2) and will remain until the first intermediate **I1**, wherein the turn T1 has been partially shaped, with the formation of one H-bond (HB1, Fig. 3a). As proteins generally fold very fast, this and many other events (vide post) have not been observed before [3, 7–17]. Nonetheless, the turn T1 disappears as the folding goes on, and in the second intermediate **I2** emerges the partial formation of the turn T2. A short helix occurs during the **I1** → **I2** transition (**Tr$_{1-2}$** in Fig. 2), which is beneficial for the coil-turn transition from **I1** to **I2**. Then, the turn T3 begins to form, and its folding degree is approximately 66.7 % in the third intermediate **I3**; meanwhile, two non-native contacts form within the Asn35–Thr49 fragment (Fig. S5) and play a significant role at the early folding stages [9, 18]. Compared with the **U** state, the R_g value of **I3** has a dramatic reduction, from ca. 54.3 to 39.8 Å, indicating that protein GB1 has been somewhat compact at this stage.



Scheme 1 Ribbon representation of GB1 as well as its sequence and secondary structure. The α-helix, *β*-strands, *β*-turn regions and random coils are shown in *red*, *cyan*, *green* and *gray*, respectively

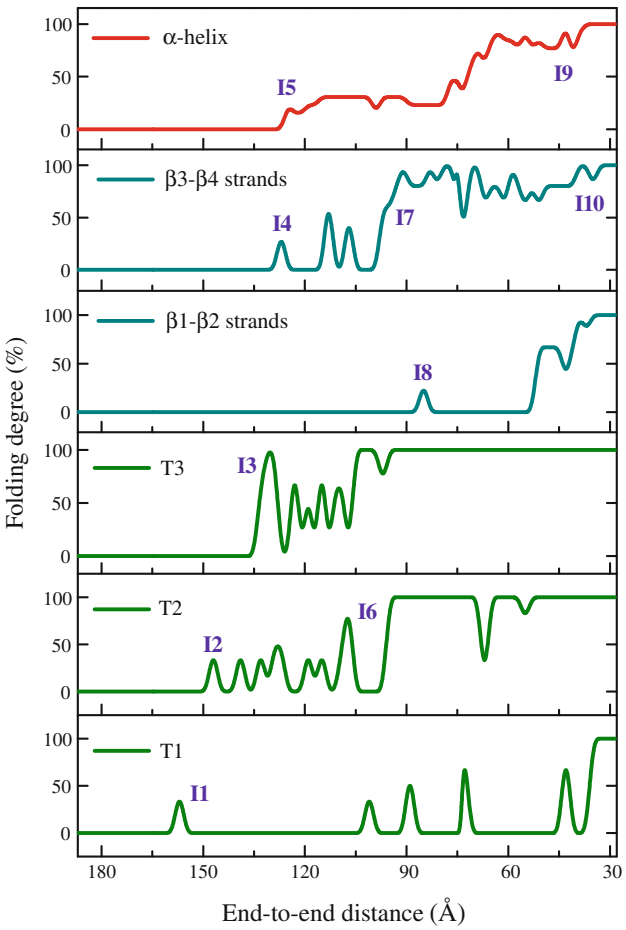


Fig. 1 The folding degrees (F_D) of the various secondary structural motifs in GB1 as function of its end-to-end distances. The definitions of *β*1–*β*2 strands, *β*3–*β*4 strands, α-helix, T1, T2 and T3 can be found in Scheme 1

After the partial folding of the three turn regions (T1, T2 and T3), the C-terminal *β*-hairpin (HP_C) begins to shape, and in the fourth intermediate **I4**, three non-native backbone H-bonds appear in the *β*3–*β*4 strands as Glu42:NH → Thr53:CO, Thr53:NH → Glu42:CO and Thr44:NH → Thr51:CO. The second turn T2 of HP_C refolds and stabilizes with two native H-bonds (HB1 and HB2, Fig. 3b). It is consistent with the experimental results that the local interactions at the turn regions and non-native interactions are essential for the formation of the C-terminal *β*-hairpin during the early folding stage [19]. The assembled *β*3–*β*4 strands are, nonetheless, short-lived and will turn into random coils in the fifth intermediate **I5**, probably due to the instability of these non-native contacts. In **I5**, three native contacts emerge at the C-terminus of the central α-helix (HE_M), together with one native helical-stabilizing H-bond (HB9); see Figs. 3c and S5. That is, HE_M has been partially shaped. The Ramachandran map of **I5** (Fig. S6) also indicates that for HE_M, the C-terminal Gln32–Asn35 residues precede to enter into the α_L region.

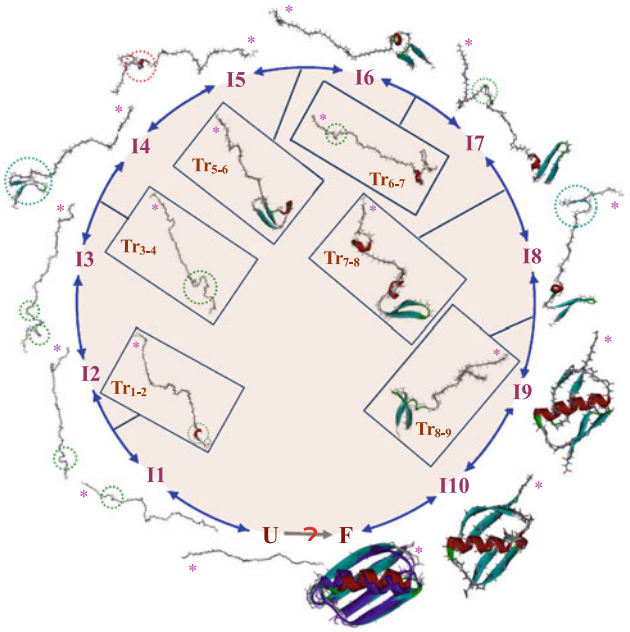


Fig. 2 Representative snapshots of GB1 along its folding process. All the structures are displayed the same manner as Scheme 1, where the N-termini are marked with *pink asterisks*. The structures that characterize the transformations between two intermediates are referred to the chosen transition structures between the (un)folded and intermediate states (**Tr**). The folded (F) state is superposed on the NMR structure (*purple*) and enlarged in Figure S8

The C-terminus is more facile to define than the rest during the folding transitions, which agrees with the mutagenesis studies of McCallister et al. [20] as well as the circular dichroism (CD) and NMR studies of Lewandowska et al. [21]. As indicated in Figs. 4 and S5, a small fraction of tertiary contacts have already appeared in **I5**, between the C-terminus of HE_M and the β3 strand of HP_C.

GB1 folds further as the end-to-end distance continues to decrease. HP_C is largely shaped in the sixth intermediate **I6**, with approximately 80 % of the β3–β4 strand residues being in order and four native H-bonds being formed (HB3, HB5, HB6 and HB7, Fig. 3b). The tertiary contacts between HE_M and HP_C increase up to 92.0 % (Fig. 4), resulting in the H-bonding between Val29 and Asp47. The concomitant production of secondary structures and tertiary contacts has also been detected by the Monte Carlo studies [15]. HP_C has been finely folded since the intermediate **I7**, where HE_M is stabilized by one native H-bond (HB9, Fig. 3) and five non-native H-bonds. HE_M and HP_C are mainly reserved during the **I7**↔**I8** transformation, but the tertiary contacts related to them are seriously damaged (Fig. 4). In the eighth intermediate **I8**, some native H-bonds of the N-terminal β-hairpin (HP_N) are reinforced, especially HB5 and HB6 (ca. 2.1 Å, see Fig. 3a). It implies that HP_N, with the folding degree of ca. 33 % in **I8**, is ready for the further folding.

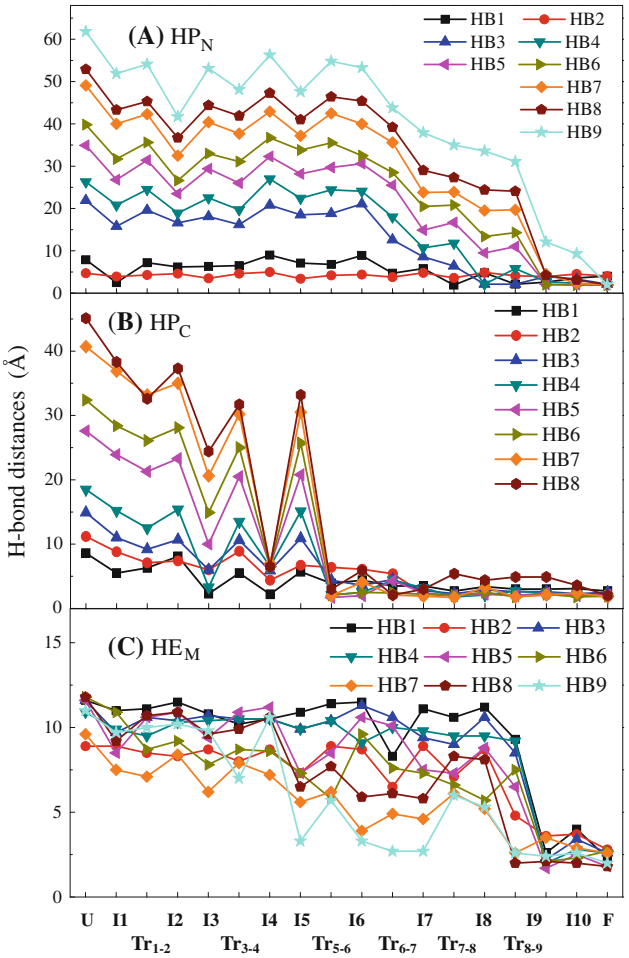


Fig. 3 The distances of the native H-bonds in GB1 changing along its folding process. The description of each H-bond can be found in Figure S4

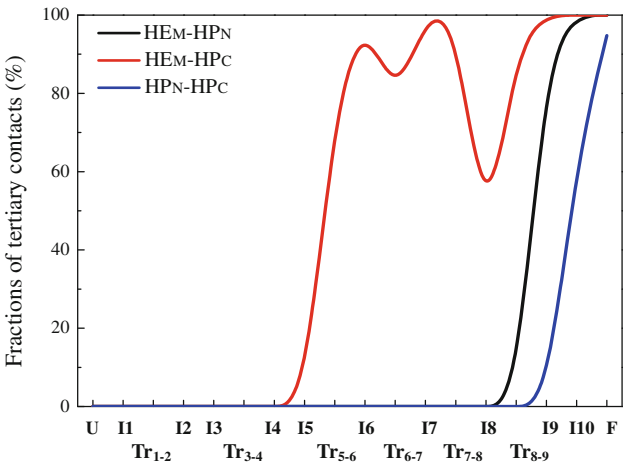


Fig. 4 The fractions of tertiary contacts in GB1, which are formed between two secondary structure subunits. The structural details of the central α-helix (HE_M), first β-hairpin (HP_N) and second β-hairpin (HP_C) can be found in Scheme 1 and Figure S4

The hydrophobic core of GB1 undergoes the largest alterations during the **I8↔I9** transition, with the radius of gyration for the hydrophobic core (R_g^{core}) decreasing from 28.0 to 8.6 Å (Fig. 5). **Tr_{8-9}** (Fig. 2) is one of the representative structures between the intermediates **I8** and **I9**, where Phe30 forms strong native tertiary contacts with the other hydrophobic residues as Phe30–Trp43 (2.1 Å), Phe30–Tyr45 (4.0 Å) and Phe30–Phe52 (2.5 Å). As the Ramachandran map of **I9** (Fig. S6) shows, all the residues in HE_M, except the two termini, have been trapped into the α_L region, and the distances of the helical native H-bonds descend significantly (Fig. 3), indicative of the fine folding of GB1 at this stage. The $\beta1$–$\beta2$ strands of HP_N are well organized, with the simultaneous production of four native H-bonds (Fig. 3). Along with the shapes of HE_M and HP_N, the tertiary contacts between them increase up to 89.5 %. The native tertiary contacts between HP_N and HP_C begin to form and stabilize with five native H-bonds within the $\beta1$–$\beta4$ strands (Figs. 4 and S7).

I10 is the next intermediate after **I9** and can be considered native-like. Its R_g value is close to that of the native state (10.4 vs. 10.1 Å). HE_M and HP_C are already in good order with their folding degrees of ca. 92.3 and 100.0 %, respectively. The folding of HP_N is significantly propagated by all its native H-bonds being formed (Fig. 3). It enhances the packing of the hydrophobic core (Fig. 5), which in turn stabilizes the protein structure [22] and completes the tertiary contacts of HE_M with HP_N and HP_C. Nonetheless, the inner $\beta1$–$\beta4$ strands are still loose, and the tertiary contacts between them amount to 63.2 % (Figs. 4 and S5). The **I10↔F** transition witnesses the formation of all the native H-bonds of the $\beta1$–$\beta4$ strands (Fig. S7); meanwhile, the native tertiary contacts of $\beta1$–$\beta4$ strands become in good order, that is, Tyr3 and Tyr45 (2.63 Å), Tyr3 and Phe52 (2.02 Å), Lys5 and Phe52 (2.77 Å), and Tyr3 and Phe30 (2.53 Å). The representative structure of the **F** state is superposed to the NMR structure (Fig. S8) and has a RMSD of 1.1 Å [7], indicative of the excellent agreement. Thus, we provide an unambiguous and detailed description of protein GB1 folding/unfolding at a molecular level.

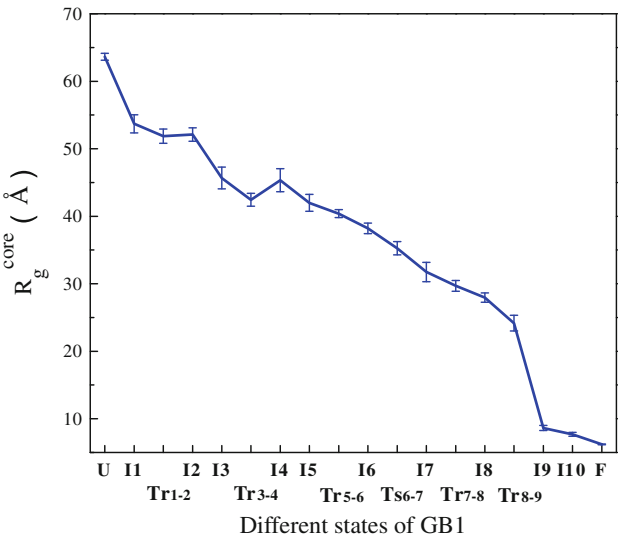


Fig. 5 The radii of gyration for the hydrophobic core (R_g^{core}) of the GB1 representative structures at the various folding states, together with their standard deviations (SD)

1 Computational section

Molecular dynamics (MD) simulations were performed using the GROMACS 4.0.5 software package [23, 24] and OPLSAA force field [25, 26]. The B1 domain of streptococcal protein G (GB1) from the NMR experiments (PDB code: 2GB1 [7]) was solvated in 9156 SPC (simple-point-charge) water molecules, that is, within a cubic box with length of 66.0 Å. Four water molecules were replaced by Na^+ ions to neutralize the GB1 system. After removing bad contacts by energy minimizations and relaxing water molecules by position-restrained MD, 50.0-ns MD simulations were run at constant temperature and pressure (300 K and 1 bar), using the Berendsen and Parrinello-Rahman barostats [27]. The particle mesh Ewald (PME) procedure was applied for long-range electrostatic interactions [28]. The cutoff radii of the coulomb and van der Waals interactions were 10.0 and 14.0 Å, respectively. The non-bonded interactions were updated every 10 time steps. All the covalent bonds with hydrogen atoms were constrained using the LINCS algorithm [29]. The time step was 2.0 fs, and the atomic coordinates were saved every 20.0 ps.

The mechanical unfolding of protein GB1 was carried out by the gradual stretching of its end-to-end distance (the two α-C atoms of residues Met1 and Glu56). Altogether, 80 end-to-end distances were covered, at 28.5 Å (the native state), at 31.0 Å and then with an interval of 2.0 Å increasing up to 187.0 Å (the completely denatured state). The end-to-end distance was restrained during each 50.0-ns MD run, and the final conformation of the previous MD run was used to start the next MD run. It resulted in a total of 4.0-μs simulation time and 2×10^5 configurations. All the simulation conditions are identical to the native state or elsewhere [30], except the shapes of the water boxes, which need altering along with the changes of the end-to-end distances. Nonetheless, in all the cases, the protein structures and box edges were at least 12.0 Å, ensuring the sufficient solvation of protein GB1.

The overall and local structural motions of protein GB1 are characterized by the backbone root-mean-square deviations (RMSD), backbone and core radii of gyration

(R_g and R_g^{core}), the folding degree of secondary structural motif (F_D) and the tertiary contacts and H-bonds. The secondary structures were analyzed according to the DSSP algorithm [31]. A native contact was defined with the atom pairs within the distance 6.5 Å. An H-bond was counted if the distance of the two heavy atoms (N and O) is below 3.5 Å and the N···H···O angle is beyond 120°. The cross-RMSD (*x*RMSD) was used to identify the similar structures of each MD run, and the conformation at the center of the largest cluster was defined as the representative structure [30, 32]. The conformational analysis was performed on all the MD runs, and those forming new structural features or/and having distinct structural changes were selected as intermediates. The transitions between two neighboring intermediates were connected by the transition states, for example, $\mathbf{Tr_{1-2}}$ discussed in the text. Unless otherwise noted, the NMR structure [7] was used as the reference.

Acknowledgments We are grateful for the financial supports from National Natural Science Foundation of China (No. 20903019), Special Fund for Basic Scientific Research of Central Colleges (No. DL09EA04-2) and Cultivated Funds of Excellent Dissertation of Doctoral Degree Northeast Forestry University (No. 140-602051).

References

1. Scheraga HA, Khalili M, Liwo A (2007) Annu Rev Phys Chem 58:57–83
2. Freddolino PL, Harrison CB, Liu YX, Schulten K (2010) Nat Phys 6:751–758
3. Sheinerman FB, Brooks CL III (1998) Proc Natl Acad Sci USA 95:1562–1567
4. Best RB, Clarke J (2002) Chem. Commun 183–192
5. Borgia A, Steward A, Clarke J (2008) Angew Chem Int Ed 47:6900–6903
6. Karsai Á, Kellermayer MSZ, Harris SP (2011) Biophys J 101:1968–1977
7. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM (1991) Science 253:657–661
8. Kuszewski J, Clore GM, Gronenborn AM (1994) Protein Sci 3:1945–1952
9. Sheinerman FB, Brooks CL III (1998) J Mol Biol 278:439–456
10. Park SH, Shastry MCR, Roder H (1999) Nat Struct Biol 6:943–947
11. Nauli S, Kuhlman B, Baker D (2001) Nat Strut Biol 8:602–605
12. Shimada J, Shakhnovich EI (2002) Proc Natl Acad Sci USA 99:11175–11180
13. Hubner IA, Shimada J, Shakhnovich EI (2004) J Mol Biol 336:745–761
14. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G (2006) Proc Natl Acad Sci USA 103:11904–11909
15. Kmiecik S, Kolinski A (2008) Biophys J 94:726–736
16. Toptygin D, Woolf TB, Brand L (2010) J Phys Chem B 114:11323–11337
17. Camilloni C, Broglia RA, Tiana G (2011) J Chem Phys 134:045105
18. Blanco FJ, Ortiz AR, Serrano L (2007) Fold Des 2:123–133
19. Lewandowska A, Oldziej S, Liwo A, Scheraga HA (2010) Biophys Chem 151:1–9
20. McCallister EL, Alm E, Baker D (2000) Nat Struct Biol 7:669–673
21. Lewandowska A, Oldziej S, Liwo A, Scheraga HA (2010) Biopolymers 93:469–480
22. Lewandowska A, Ołdziej S, Liwo A, Scheraga HA (2010) Proteins 78:723–737
23. Christen M, Hunenberger PH, Bakowies D, Baron R, Bürgi R, Geerke RP, Heinz TN, Kastenholz MA, Kräutler V, Oostenbrink C, Peter C, Trzesniak D, van Gunteren WF (2005) J Comput Chem 26:1719–1751
24. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) J Chem Theory Comput 4:435–447
25. Jorgensen WL, Tirado-Rives J (1988) J Am Chem Soc 110:1657–1666
26. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 1996(118):11225–11236
27. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684–3690
28. Darden T, York D, Pedersen L (1993) J Chem Phys 98:10089–10092
29. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) J Comput Chem 18:1463–1472
30. Wu XM, Yang G, Zu YG, Fu YJ, Zhou LJ, Yuan XH (2011) Comput Theor Chem 973:1–8
31. Kabsch W, Sander C (1983) Biopolymers 22:2577–2637
32. Luttman E, Fels G (2006) Chem Phys 323:138–147