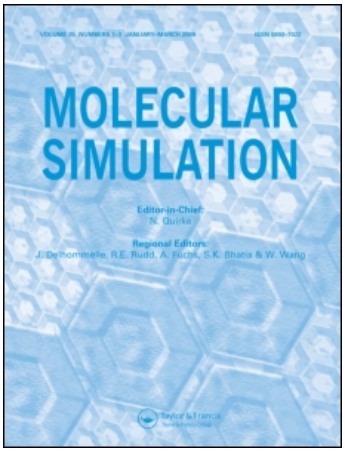

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Refolding molecular dynamics simulations of small- and middle-sized proteins in an explicit solvent

Atsushi Suenaga[ab]; Noriaki Okimoto[a]; Toshikazu Ebisuzaki[a]
[a] Computational Science Division, Advanced Computing Centre, The Institute of Physical and Chemical Research (RIKEN), Wako, Saitama, Japan [b] Computational Biology Research Centre (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan

Online publication date: 26 October 2010

To cite this Article Suenaga, Atsushi , Okimoto, Noriaki and Ebisuzaki, Toshikazu(2002) 'Refolding molecular dynamics simulations of small- and middle-sized proteins in an explicit solvent', Molecular Simulation, 28: 4, 337 — 357
To link to this Article: DOI: 10.1080/08927020290018732
URL: http://dx.doi.org/10.1080/08927020290018732

Molecular Simulation, 2002 VOL. 28 (4), pp. 337–357

REFOLDING MOLECULAR DYNAMICS SIMULATIONS OF SMALL- AND MIDDLE-SIZED PROTEINS IN AN EXPLICIT SOLVENT

ATSUSHI SUENAGA[a,b,*], NORIAKI OKIMOTO[a] and TOSHIKAZU EBISUZAKI[a]

[a]*Computational Science Division, Advanced Computing Centre, The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan;*
[b]*Computational Biology Research Centre (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6 Aomi, Koutou-ku, Tokyo 135-0064, Japan*

In order to elucidate the protein folding problem, we performed molecular dynamics simulations for small- and middle-sized two unfolding and six refolding proteins in an explicit solvent. Histidine-containing phosphocarrier protein and small designed protein were chosen for the simulations. We found that the protein folding process of these proteins was divided into three phases: an α-helix formation phase, a packing phase and a β-sheet formation phase. In the α-helix formation phase, an α-helix was developed from a β-turn structure through a 3_{10}-helix state. In the packing phase, proteins became compact, and tertiary structures (α/α or pre-β/β packing) were formed. Formation of a hydrophobic nucleus occurred concomitant with the α-helix formation and packing phase. Finally, in the β-sheet formation phase, a β-sheet was developed owing to the sequential formation of hydrogen bonds between two neighbouring strands, just like a "closing zipper".

Keywords: Protein folding; Molecular dynamics; α-helix; β-sheet; Tertiary structure

INTRODUCTION

The protein folding problem, known as how a chain of amino acids folds into a well-defined unique structure, is one of the most fundamental enigmas in bioscience. Despite the hard work by many researchers, the protein folding

*Corresponding author. Address: Computational Biology Research Centre (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6 Aomi, Koutou-ku, Tokyo 135-0064, Japan. Tel.: +81-3-3599-8058. Fax: +81-3-3599-8081. E-mail: suenaga-a@aist.go.jp

ISSN 0892-7022 © 2002 Taylor & Francis Ltd
DOI: 10.1080/08927020290018732

problem is still not well understood and how a sequence can find its native structure in a huge conformational space remains a riddle. Knowledge of the protein folding mechanism will result in a huge advance in general bioscience, especially in the field of the drug design and pharmaceutical chemistry. Prion disease and Alzheimer's disease, for example, have been found to be caused by miss-folding of proteins [1,2].

Molecular dynamics (MD) simulation is one of the most promising approaches for solving the protein folding problem. Simulation of the temporal evolution of a molecular system at a high time resolution and molecular behaviour at the atomic level as it occurs in nature is possible. In MD simulations, new positions of atoms are calculated by numerical integration of Newton's equation of motion. The time step must be sufficiently small to resolve the highest frequency mode in the system (a few femtoseconds). However, the time scale on which the fastest folding proteins form their native structures is in the microsecond range [3,4], and 10^9 integration steps are necessary to simulate the whole process of folding. Since such a huge number of time steps requires a large computer resource, particularly, for a large molecular system, folding simulations have been performed mainly for peptide or tiny protein systems [5–13]. Such simulations have shown formation of district secondary structures, such as a β-turn [5], α-helix [6,10,13], type VI reverse turn [7], and β-hairpin [8,11,12]. Duan and Kollman [13] were the first to perform a microsecond MD simulation, and they demonstrated the effectiveness of MD simulations for solving the protein folding problem. In other protein folding simulations, attempts have been made to simplify the system by representing amino acid residues as linked beads or by implicit representation of a solvent [14–17]. These studies have shown that secondary structures fold in only a few nanoseconds and that they use small peptides [5–12] or small proteins including only α-helices [13]. In the present study, we performed several refolding MD simulations using small- and middle-sized proteins including both α-helices and β-sheet.

Until now, several models have been proposed for the protein folding pathway. The "framework model" proposed by Kim and Baldwin [18] suggests that the formation of each secondary structural element, stabilised in comparatively short-range interactions, precedes the formation of the tertiary structure. The "hydrophobic collapse model" assumes the initial condensation of hydrophobic elements that gives rise to compact states without secondary structures [19]. The development of native-like tertiary interactions in the compact states prompts the subsequent formation of stable secondary structures. These two models are different in the timing of formation of the secondary structure and the major driving force of protein folding [20,21]. The "nucleation condensation model" assumes the early formation of a relatively stable substructure, directing the rest

of the protein to fold around it [22]. However, it is very difficult to clarify the general pathway of protein folding via experimental observations due to the very short lifetimes of the folding intermediates. From our multiple refolding MD simulations starting at different levels (fully or partially unfolded structures) of the process, we have tried to assemble a protein folding pathway.

In this study, we chose two real proteins: a small designed protein (psv) with only 28 residues and a middle-sized histidine-containing phosphocarrier protein (HPr) with 87 residues. The structures of these proteins are known [23,24]; Psv contains an α-helix and a two-stranded β-sheet (Fig. 1A), and HPr has three α-helices and a four-stranded β-sheet (Fig. 1B). Using our MD simulations for these proteins, we found that there are three phases in the folding process, (I) an α-helix formation phase, (II) a packing phase, and (III) a β-sheet formation phase. The unfolding process followed opposite way (III → II → I).

METHODS

Molecular Dynamics Simulation

All calculations were performed on a Fujitsu VPP700E supercomputer using an AMBER 5.0 simulation software program [25]. To neutralise the systems, counter ions (Na^+ or Cl^-) were placed near charged residues by the EDIT module of

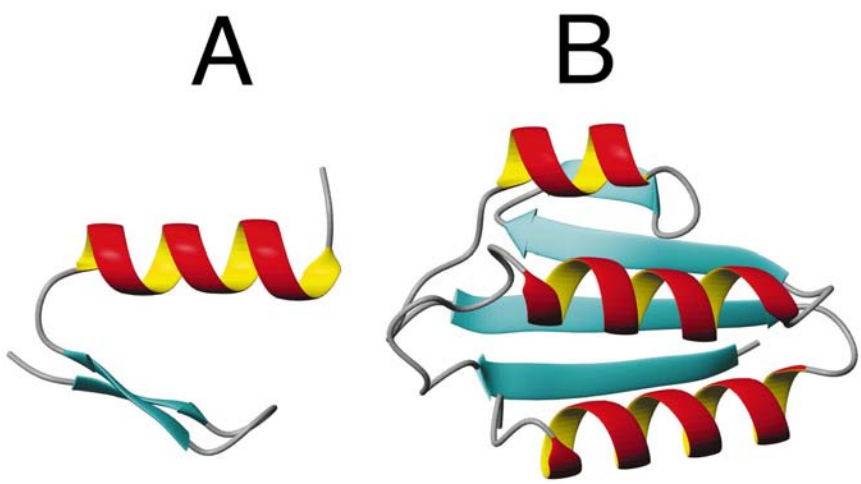

FIGURE 1 Ribbon diagrams showing the secondary structures of native (A) small designed protein (psv) and (B) histidine-containing phosphocarrier protein (HPr). The α-helices are drawn in red and yellow, and the β-sheets are drawn in blue.

AMBER. The NMR solution structure (SS) of psv (PDB code; 1psv [24]) and the crystal structure (CS) of HPr (PDB code; 1ptf [23]) were placed in boxes filled with a TIP3P water model [26]. The sizes of the boxes ($52 \times 44 \times 42$ for psv and $73 \times 58 \times 49$ Å for HPr) were chosen so that the distances of atoms in the protein from the wall were more than 10.0 Å. The total numbers of atoms for the simulations of psv and HPr were 8176 and 17,863, respectively. We adopted the force field of Cornell *et al.* [27] and imposed the periodic boundary condition. The temperature and pressure were kept constant by the method of Berendsen *et al.* [28], and the solute and solvent were separately coupled to a temperature bath with coupling constants of 0.2 ps and pressure relaxation time of 0.5 ps in all simulations.

Unfolding/refolding Simulation

We performed two unfolding simulations and six refolding simulations of psv and HPr. Table I shows details of these simulations. In all unfolding simulations (temperature is set at 373 K), non-bonded interactions were evaluated using a twin range cutoff method (9×17 for psv and 9×18 Å for HPr), and all bonds were constrained to an equilibrium length using the SHAKE method [29]. In all folding simutations, the particle-mesh Ewald was used to calculate long-range Coulomb force. Only bonds involving hydrogen atoms were constrained by the SHAKE method [29]. The temperature of the system of HPr was gradually lowered from 373 to 300 K during a 60 ps period, and the system of psv was

TABLE I Summary of simulations

Protein	*Simulation*	*Name*	*Initial**	*Simulation time* (ns)
Psv	Control	Control	SS†	3.0
	Unfolding	Usim-psv	SS†	3.0
	Refolding	Rsim1-psv	EX‡	3.0
		Rsim2-psv	PU1¶	1.5
		Rsim3-psv	PU2§	2.6
HPr	Control	Control	CS‖	2.0
	Unfolding	Usim-HPr	CS‖	1.0
	Refolding	Rsim1-HPr	PU1#	4.5
		Rsim2-HPr	PU2**	1.7
		Rsim3-HPr	PU3††	1.5

* Initial structure for the simulation.
† Solution structure.
‡ Extended structure.
¶ Partially unfolded structure from Usim-psv at 0.8 ns.
§ Partially unfolded structure from Usim-psv at 0.1 ns.
‖ Crystal structure.
Partially unfolded structure from Usim-HPr at 1.0 ns.
** Partially unfolded structure from Usim-HPr at 0.35 ns.
†† Partially unfolded structure from Usim-HPr at 0.15 ns.

gradually cooled to 280 K during a 80 ps period. The time steps of unfolding and folding simulations were set at 2.0 and 0.5 fs, respectively.

In order to investigate the structural stability of psv and HPr, we propagated two control MD simulations at 280 K for psv (3 ns) and 300 K for HPr (2 ns) under the same conditions as those for the refolding MD simulations. The EX of psv was longer than the native structure, and the box size was very large ($123 \times 68 \times 59$ Å). The total number of atoms in the system of Rsim1-psv was 45,655.

Analysis

The secondary structures were recognised by using the SECSTR module of the program PROCHECK [30] throughout this study. Figures 1, 7–10 were prepared using the program MOLMOL [31].

RESULTS AND DISCUSSION

Control MD Simulations At a low Temperature

The control MD simulations of psv (at 280 K) and HPr (at 300 K) were studied to assess the stability of the proteins. The secondary structure contents of psv and HPr as functions of time are shown in Figs. 2A and 3A, respectively. Psv and HPr remained native secondary structures throughout the 3.0 ns (or 2.0 ns for HPr) MD simulation. The three-dimensional (tertiary) structures were very close to the respective native structures during the 3.0 ns (or 2.0 ns for HPr) MD simulations with the main-chain root mean square deviations (RMSDs) being 1.290 Å for psv and 3.753 Å for HPr (Fig. 4A for psv and Fig. 4C for HPr, Tables II and III). Although the RMSD of HPr was large after the 2.0 ns MD simulation, in particular at the loop and the C-terminal α-helix region, the whole fold of HPr was very similar to that of the native structure. The tertiary contact maps are shown in Fig. 5. The tertiary structures of psv and HPr remained the native during the 3.0 ns (or 2.0 ns for HPr) MD simulations (Fig. 5A,E). Other structural parameters, radius of gyration (R_g, Fig. 4B,D) and solvent accessible surface area (ASA, Fig. 6), suggest that these two proteins are very stable at 280 or 300 K (See also Tables II and III for details of structural parameters of two control MD simulations).

Unfolding Simulation of Psv

The unfolding process of Usim-psv over a 3.0 ns period is shown in Fig. 7. The β-sheet disappeared rapidly within 0.02 ns (Fig. 2B). The α-helix began to

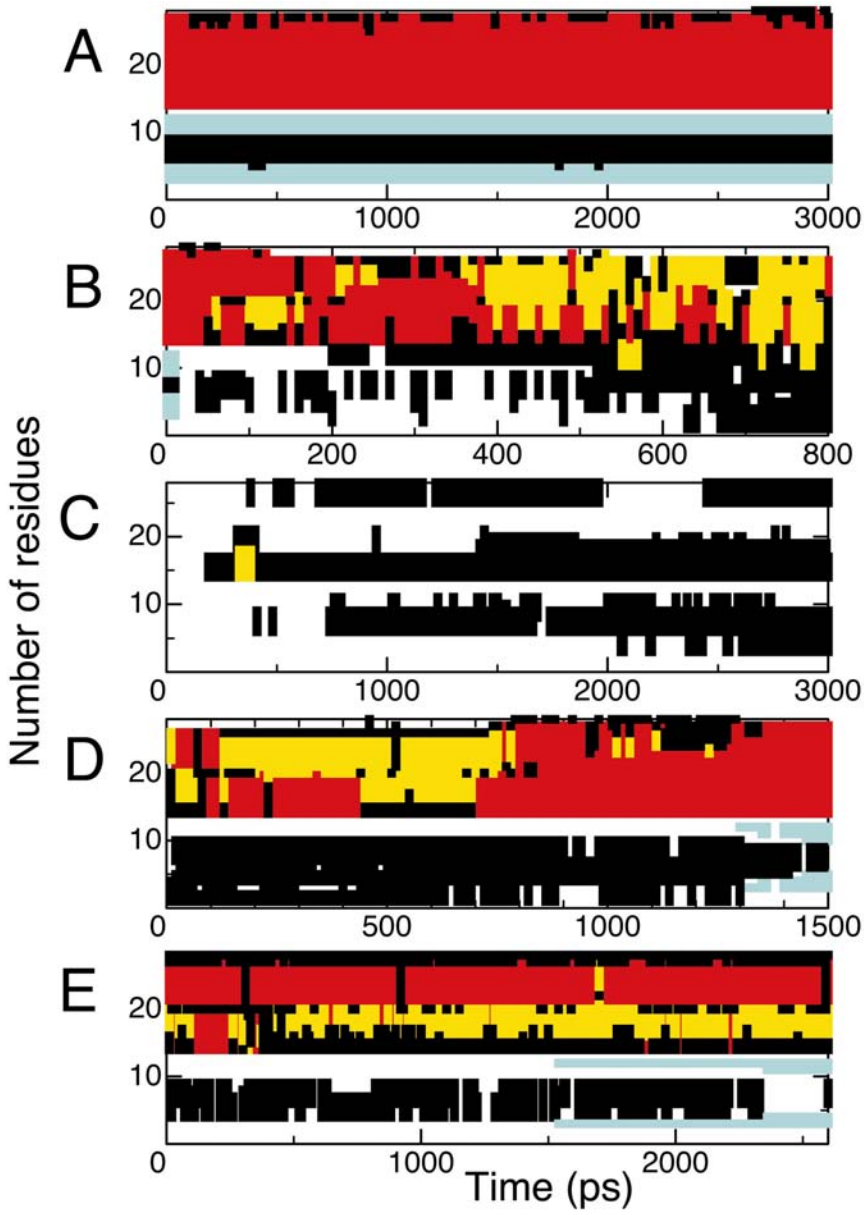

FIGURE 2 Content of the secondary structure as a function of time from (A) control MD simulation of psv, (B) Usim-psv, (C) Rsim1-psv, (D) Rsim2-psv and (E) Rsim3-psv. The α-helix is shown in red boxes, the 3_{10}-helix in yellow boxes, the β-sheet in blue boxes, and β-turn in black boxes.

TABLE II Structural parameters for psv

*State**	*SS*	*Control*†	*EX*	*PR1*†	*PU1*	*FR*†	*PU2*	*PR2*†
RMSD (Å)‡	–	1.070	22.939	10.629	4.055	2.688	4.554	1.917
R_g (Å)¶	8.589	8.697	28.971	15.938	9.693	8.351	8.634	8.618
Helix (%)§	42.9	43.1	0.0	0.0	10.7	42.9	35.7	41.8
Sheet (%)‖	21.4	19.6	0.0	0.0	0.0	21.4[#]	0.0	14.3[#]
ASA (Polar)	2051.4	2429.1	4210.8	2692.4	2321.6	2310.6	2407.3	2416.6
(Non-polar)	348.9	431.4	1646.1	1241.3	835.3	538.1	519.0	424.3

* SS, solution structure; Control, control MD simulation at 280 K; EX, extended structure; PR1, partially refolded structure from EX; PU1, partially unfolded structure from Usim-psv at 0.8 ns; FR, fully refolded structure from PU1; PU2, partially unfolded structure from Usim-psv at 0.1 ns; PR2, partially refolded structure from PU2.
† Average values over the last 0.5 ns period of simulation.
‡ Main chain RMSD from solution structure.
¶ Radius of gyration.
§ Percentage of helical content.
‖ Percentage of β-sheet content.
[#] The value at 1.5 ns of Rsim2-psv for FR and at 2.1 ns of Rsim3-psv for PR2.

TABLE III Structural parameters for HPr

*State**	*CS*	*Control†*	*PU3*	*PR1†*	*PU2*	*PR2†*	*PU1*	*PR3†*
RMSD (Å)‡	–	3.753	13.328	12.559	9.208	8.129	4.756	4.418
R_g (Å)¶	11.638	11.331	16.194	14.625	13.777	12.332	12.396	11.605
Helix (%)§	37.9	37.8	36.8	36.6	35.6	36.4	39.1	37.4
Sheet (%)‖	26.4	26.4	0.0	0.0	0.0	1.4	11.5	25.9
ASA (Polar)	3939.7	3844.3	5265.2	4321.1	4595.4	4044.7	3978.5	4340.1
(Non-polar)	1222.4	1245.2	3560.8	2661.1	2562.2	1918.1	2006.4	1428.6

* CS, crystal structure; Control, MD simulation at 300 K; PU3, partially unfolded structure from Usim-HPr at 1.0 ns; PR1, partially refolded structure from PU3; PU2, partially unfolded structure from Usim-HPr at 0.35 ns; PR2, partially refolded structure from PU2; PU1, partially unfolded structure from Usim-HPr at 0.15 ns; PR3, partially refolded structure from PU1.
† Average values over the last 0.5 ns period of simulation.
‡ Main chain RMSD from the crystal structure.
¶ Radius of gyration.
§ Percentage of helical content.
‖ Percentage of β-sheet content.

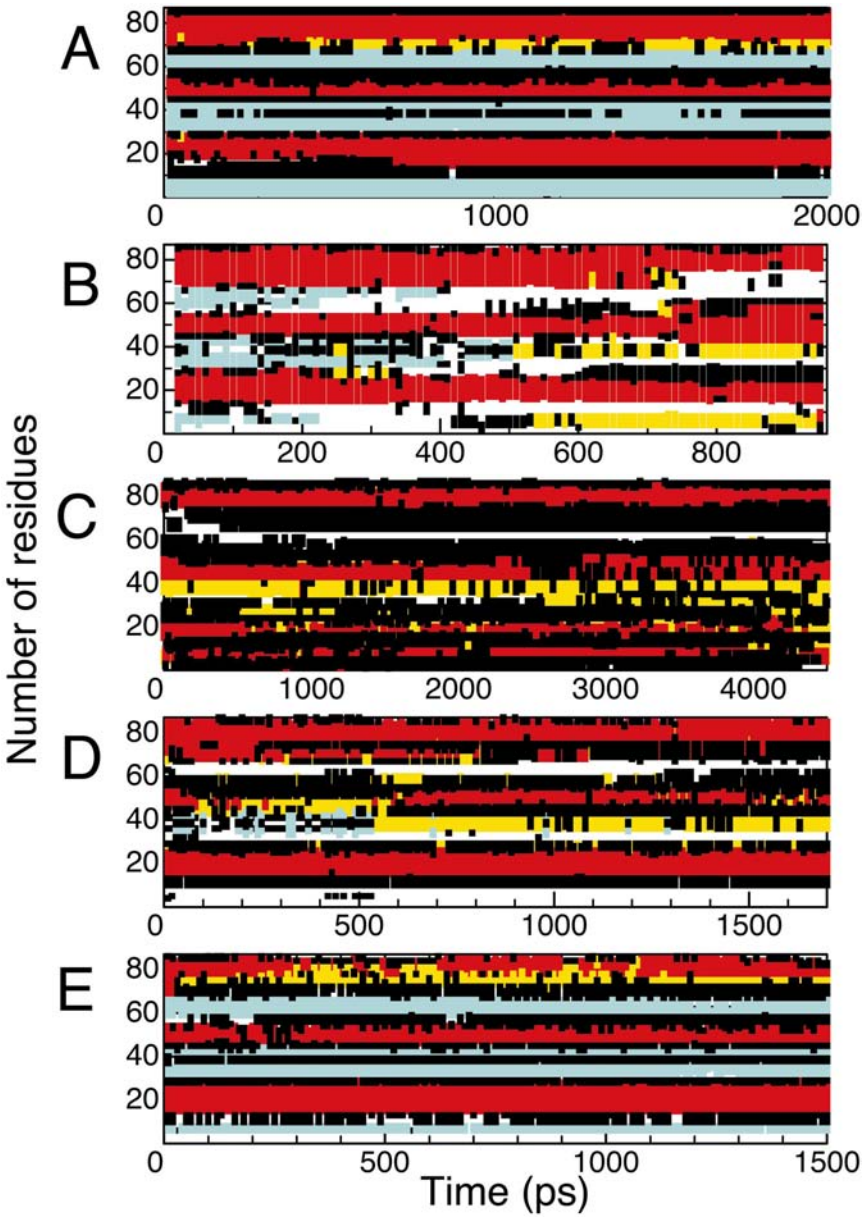

FIGURE 3 Content of the secondary structure as a function of time from (A) control MD simulation of HPr, (B) Usim-HPr, (C) Rsim1-HPr, (D) Rsim2-HPr and (E) Rsim3-HPr. See the legend to Fig. 2 for the colour scheme.

collapse at the central part and fully deformed at 0.8 ns (Figs. 2B and 7). Subsequently, the α-helix region was transformed into a 3_{10}-helix and then into a β-turn structure (Fig. 2B). After 0.8 ns (main-chain RMSD from the native structure was 4.055 Å at 0.8 ns), the psv had greatly unfolded (Fig. 4A) and the structure of psv had become a random coil after 1.0 ns (main-chain RMSD was around 7.0 Å). ASA of the nonpolar part also increased through two steps (Fig. 6A,B). These steps in the increase of ASA took place during periods of 0–0.6 ns (first step) and 0.8–3.0 ns (second step). These results indicate that the structure formed at about 0.8 ns is an intermediate state of the unfolding of psv. The tertiary structure of the intermediate state, especially the contacts between strand 2 of the β-sheet (residue numbers 9–12) and the α-helix (upper-left diagonals of Fig. 5A,C), was lost.

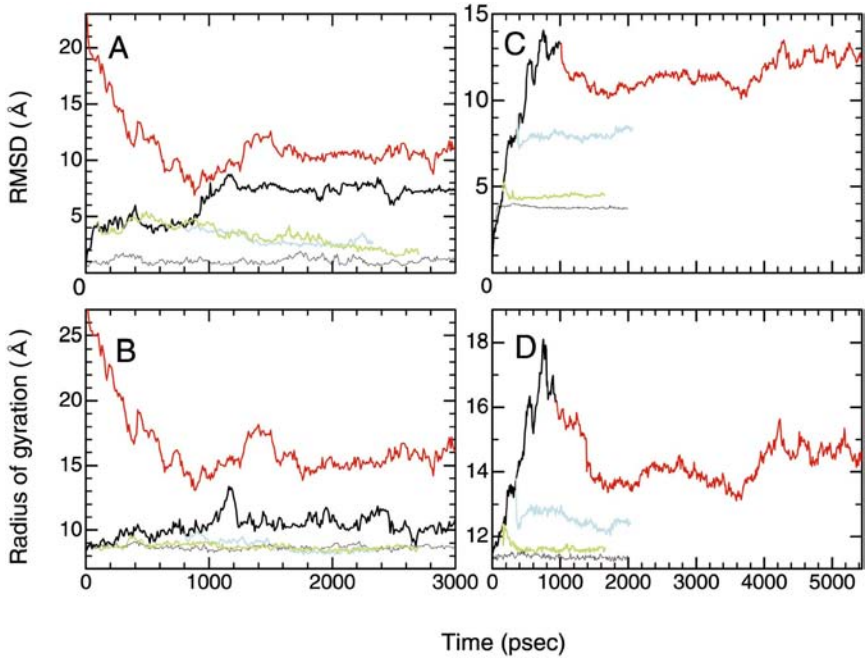

FIGURE 4 (A) Backbone root mean square deviations (RMSD) from the solution structure and (B) radius of gyration (R_g) of simulation structures of psv as a function of time. Black thick lines, Usim-psv; black thin lines, control simulation of psv; red lines, Rsim1-psv; blue lines, Rsim2-psv; and green lines, Rsim3-psv. (C) Backbone RMSD from the crystal structure and (D) R_g of simulation structures of HPr as a function of time. Black thick lines, Usim-HPr; black thin lines, control simulation of HPr; red lines, Rsim1-HPr; blue lines, Rsim2-HPr; and green lines, Rsim3-HPr.

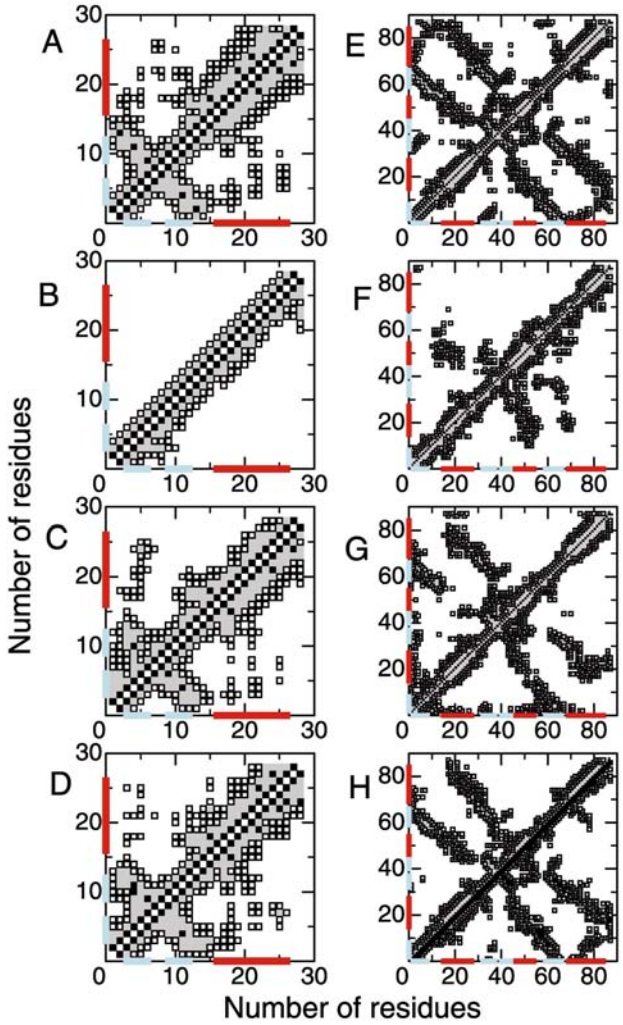

FIGURE 5 Tertiary contact maps of psv [(A) to (D)] and HPr [(E) to (H)]. (A) Solution structure of psv (upper-left diagonal)/3.0 ns control MD simulation of psv (lower-right diagonal), (B) initial structure of Rsim1-psv (extended structure)/PR obtained by Rsim1-psv at 3.0 ns, (C) PU obtained by Usim-psv at 0.8 ns/PR obtained by Rsim2-psv at 1.5 ns, (D) PU obtained by Usim-psv at 0.1 ns/PR obtained by Rsim3-psv at 2.6 ns, (E) crystal structure of HPr/2.0 ns control MD simulation of HPr, (F) PU obtained by Usim-HPr at 1.0 ns/PR obtained by Rsim1-HPr at 4.5 ns, (G) PU obtained by Usim-HPr at 0.35 ns/PR obtained by Rsim2-HPr at 1.7 ns, (H) PU obtained by Usim-HPr at 0.15 ns/PR obtained by Rsim3-HPr at 1.5 ns. The darkest blocks represent Cα–Cα distances under 5 Å; shaded blocks, under 8 over 5 Å; and open blocks, under 12 over 8 Å. Blue and red lines on the *X*- and *Y*-axes indicate the β-strand and α-helical regions in the native protein, respectively.

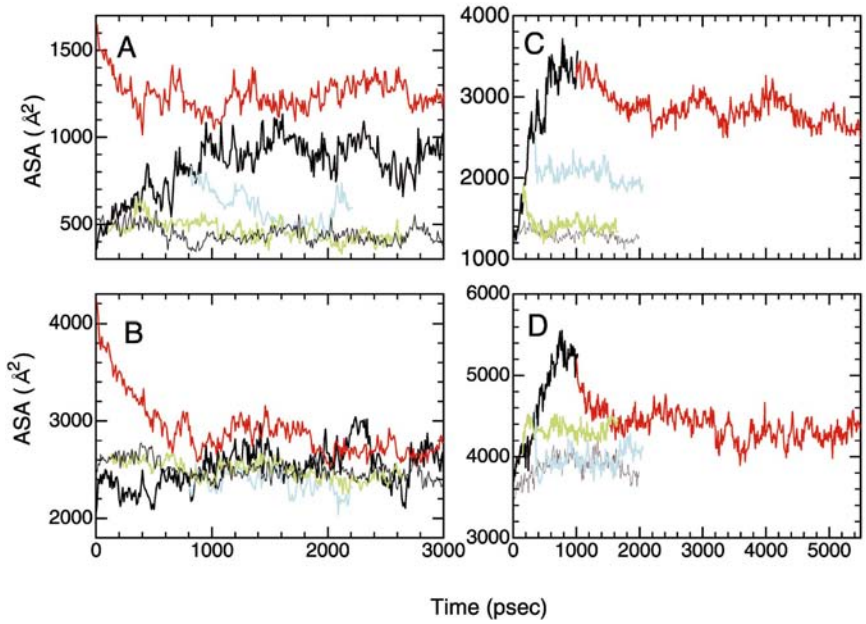

FIGURE 6 Nonpolar and polar solvent-accessible surface area (ASA) as a function of time. (A) Nonpolar and (B) polar ASA from simulations of psv. Black thick lines, Usim-psv; black thin lines, control simulation of psv; red lines, Rsim1-psv; blue lines, Rsim2-psv; and green lines, Rsim3-psv. (C) Nonpolar and (D) polar ASA from simulations of HPr. Black thick lines, Usim-HPr; black thin lines, control simulation of HPr; red lines, Rsim1-HPr; blue lines, Rsim2-HPr; and green lines, Rsim3-HPr.

Unfolding Simulation of HPr

Figure 8 shows the unfolding process of Usim-HPr over a period of 1.0 ns. At first, a C-terminal α-helix started to separate from the hydrophobic core at about 0.15 ns (Fig. 8) (main-chain RMSD from the native structure was 4.756 Å at 0.15 ns). After that, two intermediate states appeared at about 0.35 and 1.0 ns, respectively (Fig. 8). The first intermediate state (at about 0.35 ns) retained all of

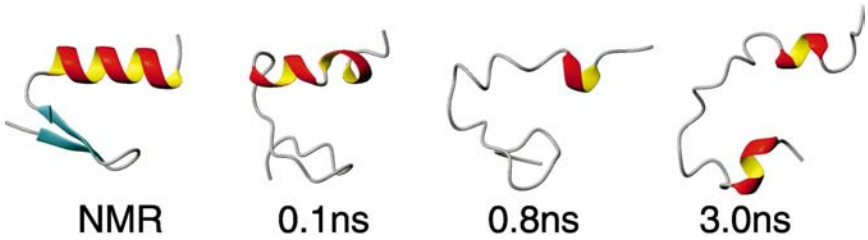

FIGURE 7 Ribbon illustrations of snapshots from Usim-psv trajectories. See legend to Fig. 1 for definitions of colours of the secondary structures.

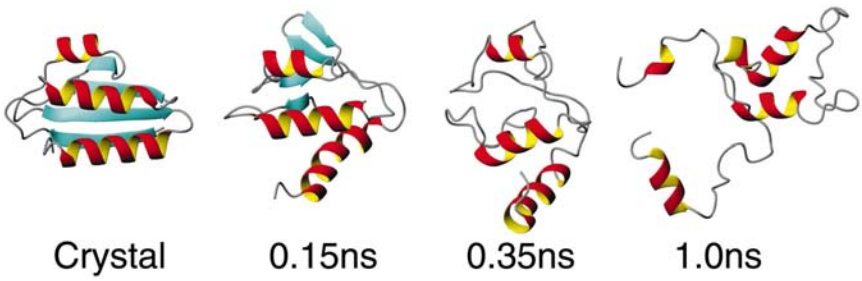

FIGURE 8 Ribbon illustrations of snapshots from Usim-HPr trajectories. See legend to Fig. 1 for definitions of colours of the secondary structures.

the native α-helices (Fig. 3B) and the tertiary structure (upper-left diagonals of Fig. 5E,G) except the β-sheet structure (Fig. 3B) (main-chain RMSD from the native structure was 9.208 Å). In the second intermediate state (at about 1.0 ns), no tertiary structure was not able to found (upper-left diagonals of Fig. 5E,F), though all of the native α-helices remained (Fig. 3B) (main-chain RMSD from the native structure was 13.328 Å). See Suenaga *et al.* [32] for details.

Unfolding Pathways of Psv and HPr

The unfolding process of psv started from the collapse of the hydrophobic nucleus (∼0.8 ns, Fig. 6). The β-sheet disappeared rapidly within 0.02 ns. After the β-sheet disruption, a part of the tertiary structure was lost (∼0.7 ns, upper-left diagonal of Fig. 5C). Finally, the α-helix destabilised and reformed through the pathway of 3_{10} − helix ⇔ β − turn (∼0.8 ns). These observations suggest that the unfolding pathway of psv follows: native ⟶ disruption of β-sheet structure ⟶ disruption of tertiary structure ⟶ disruption of α-helix structure ⟶ random coil. Since the ASA of the nonpolar part was continuously increasing during the path from "native" to "disruption of the α-helix structure" (0–0.8 ns), it was very difficult to determine where exactly the "collapse of the hydrophobic nucleus" is in the pathway.

The unfolding pathway of HPr was same as that of psv, i.e. native ⟶ disruption of β-sheet structure ⟶ disruption of tertiary structure ⟶ disruption of α-helix structure ⟶ random coil. In the MD simulation of unfolding of HPr, we did not investigate after the path from "disruption of α-helix" to "random coil", because of the short simulation time. The "collapse of the hydrophobic nucleus" continued from "native" to "disruption of α-helix" of the path as in the unfolding pathway of psv. See Suenaga *et al.* [32] for details of the pathway.

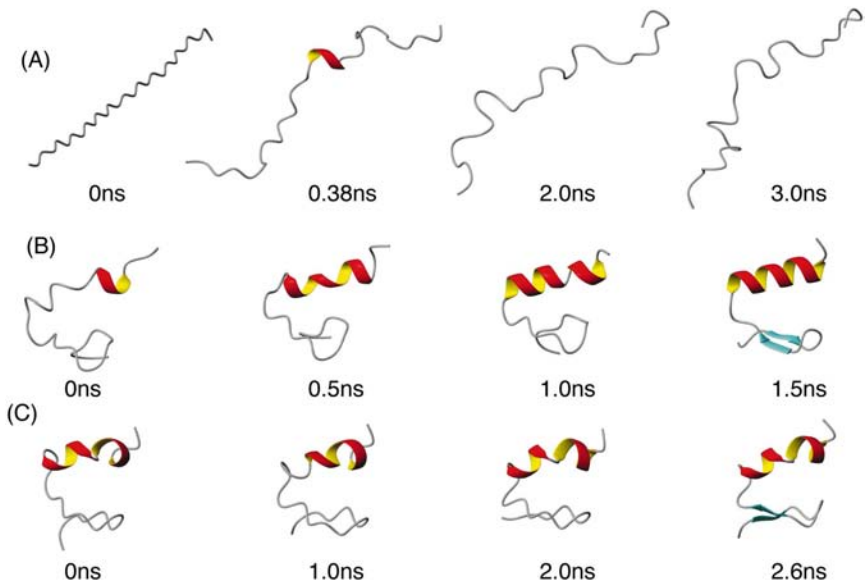

FIGURE 9 Ribbon illustrations of snapshots from Rsim-psv trajectories. (A) (top line) shows the conformational change of Rsim1-psv, (B) (middle line) shows the conformational change of Rsim2-psv, and (C) (bottom line) shows the conformational change of Rsim3-psv. See legend to Fig. 1 for definitions of colours of the secondary structures.

Refolding Simulation of Psv from Extended Structure: Initial Phase of Folding

Figure 9A shows the results of refolding simulation of Rsim1-psv representing the initial phase of the folding process over a period of 3.0 ns. At first, four β-turns appeared (Fig. 2C). These four turns were located at residues 14–17 (type I β-turn at the central part of the peptide, TC), residues 16–19 (type III β-turn at the helical region of the native structure, TH), residues 25–28 (type III β-turn at the C-terminus, TT) and residues 8–11 (type II β-turn in the middle of the two-stranded β-sheet, TB). Since the hydrogen bond in TH [Lys16(CO)–Arg19(HN)] easily developed into an α-helical hydrogen bond [Lys16(CO)–Asp20(HN)], one round of the α-helix (nascent α-helix) was tentatively formed through the 3_{10}-helix structure. These results indicate that the initial step of protein folding is the formation of a β-turn and an α-helix. Structural parameters, RMSD, R_g, and ASA, gradually decreased, approaching those of the control simulation, but the Rsim1-psv could not demonstrate the final refolding (Figs. 4A,B and 6A,B and Table II). R_g and ASA of the polar part showed particularly large decreases, indicating that the protein had become compact.

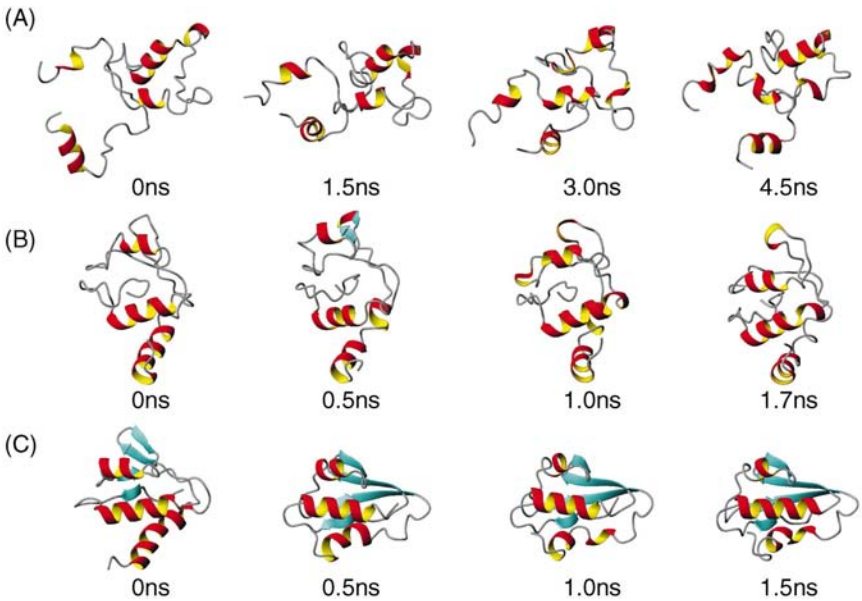

FIGURE 10 Ribbon illustrations of snapshots from Rsim-HPr trajectories. A (top line) shows the conformational change of Rsim1-HPr, B (middle line) shows the conformational change of Rsim2-HPr, and C (bottom line) shows the conformational change of Rsim3-HPr. See legend to Fig. 1 for definitions of colours of the secondary structures.

Refolding Simulation of Psv from a Partially Unfolded Structure Obtained by Unfolding Simulation At 0.8 ns: Middle Phase of Folding

Figure 9B shows snapshots of simulation of Rsim2-psv over a period of 1.5 ns. At first, the β-turn in the helical region of the native structure (residues 16–19, TH) changed to a 3_{10}-helix and then developed into a nascent α-helix (Fig. 2D). This is exactly the reverse process of the unfolding shown in Usim-psv and also observed in Rsim1-psv. A full-length α-helix was rapidly formed from the nascent α-helix at 0.65 ns (Figs. 2D and 9B). These results are consistent with previous observations [10,11,33–36] of the α-helix formation/deformation taking a unique pathway: $\alpha - \text{helix} \Leftrightarrow 3_{10} - \text{helix} \Leftrightarrow \beta - \text{turn}$.

After formation of the α-helix, the peptide was gradually packed until 1.2 ns. The radius of gyration (R_g), which indicates the compactness of a protein, decreased from +13 to −3% (Fig. 2B and Table II). RMSD (Fig. 2A) and ASA (Fig. 6A,B) also decreased. The decrease in ASA of the nonpolar parts indicates nucleation of the hydrophobic core.

Finally, β-sheet formation and α/β packing (upper-left and lower-right diagonals of Fig. 5C) took place. Since the PU (at 0.8 ns) of psv (initial structure

for Rsim2-psv) had a hydrogen bond between two neighbouring strands [Tyr3(CO)–Phe12(HN)], located in a β-sheet of the native structure, the native (-like) β-sheet propagated by forming hydrogen bonds sequentially just like a "closing zipper" from 1.2 to 1.5 ns (Fig. 2D). A similar "closing zipper"-like propagation was observed in a previous folding simulation of the β-hairpin [11]. The first hydrogen bond between the two neighbouring strands played the role of a "junction of the zipper". The final structure of Rsim2-psv seemed to be a transition state of folding, since the final structure was very similar to the native structure. The structural parameters of the final structure however were slightly different from those of the native structure. When we performed Rsim2-psv at 300 K, a β-sheet was not formed during a period of 8.0 ns (data not shown), and this finding is consistent with the results of an NMR experiment showing that psv took a defined three-dimensional structure at 280 K but not at 300 K [37]. Although we performed several Rsims starting from other PUs of psv (more unfolded structures than PU at 0.8 ns of Usim-psv) that had no "junction of the zipper", we did not observe the "closing zipper" of the β-sheet. Thus, the seed of a hydrogen bond for the β-sheet, which was not formed in our simulation trials, is absolutely necessary.

Refolding Simulation of Psv from a Partially Unfolded Structure Obtained by Unfolding Simulation At 0.1 ns: Final Phase of Folding

Figure 9C shows the results of simulation of Rsim3-psv over a period of 2.6 ns. The starting structure, PU at 0.1 ns of Usim-psv, had an α-helix and the structural parameters were close to those of the native structure (Fig. 4A,B, upper-left diagonal of Figs. 5D and 6A,B, and Table II). In this simulation, we observed the final stage of the peptide folding. The α-helix was relatively stable during a period of 2.6 ns (Fig. 2E). At first, a tertiary structure was formed during a period of 1.2 ns (data not shown). Next, since the initial structure of Rsim3-psv had the "junction" of β-sheet [Tyr3(CO)– Phe12(HN)] like PU at 0.8 ns of the Usim-psv, the β-sheet was formed rapidly as in the case of Rsim2-psv (Fig. 2E). The partially refolded structure (PR) obtained by Rsim3-psv at 2.6 ns seemed to be a transition state of folding, because it was similar to the native structure. Structural parameters of the PR were slightly different from those of the native structure (Table II). When we performed Rsim3-psv at 300 K, a β-sheet was not formed in 6.0 ns simulation.

Overall, the results indicate that the β-sheet is formed at the final phase of folding and that the "junction" of the β-sheet plays an important role in the β-sheet formation.

Refolding Simulation of HPr from a Partially Unfolded Structure Obtained by Unfolding Simulation At 1.0 ns

Figure 10A shows the results of simulation of Rsim1-HPr over a period of 4.5 ns. At first, a hydrophobic nucleus was formed by packing of residual α-helices; R_g, ASA (especially of the nonpolar part), and RMSD decreased (Figs. 4C,D, 5F, 6C,D and Table III). Furthermore, two strands (residues 32–37 and 60–66) of the β-sheet in the native structure came close to each other (pre-β/β packing) at the hydrophobic nucleus in the partially refolded structure (PR) [Figs. 5F and 10A (3.0 ns)], although the β-sheet was not formed. The partially unfolded structure (PU) obtained by Usim-HPr at 0.35 ns was quite similar to the PR by Rsim1-HPr at 4.5 ns, except for the N-terminus ($\Delta R_g\% = 7.3\%$, ΔASA% of the nonpolar part = 8.1%, and ΔASA% of the polar part = 17.0%; Table III). The difference in secondary structure contents between PU obtained by Usim-HPr at 0.35 ns and the PR by Rsim1-HPr at 4.5 ns was also small ($\Delta\%_{\alpha-\mathrm{helix}} = 1.0\%$, $\Delta\%_{\beta-\mathrm{sheet}} = 0.0\%$; Table III). The tertiary contact maps of the PU (upper-left diagonal of Fig 5G) and the PR (lower-right diagonal of Fig. 5F) were very similar. However, Cα atom positional RMSD of the PR obtained by Rsim1-HPr at 4.5 ns from the PU by Usim-HPr at 0.35 ns was 11.841 Å, i.e. PU obtained by Usim-HPr at 0.35 ns ↔ (PR by Rsim1-HPr at 4.5 ns) ↔ PU by Usim-HPr at 1.0 ns. These facts indicate that the pathway is unique for the local unfolding/refolding process.

Refolding Simulation of HPr from a Partially Unfolded Structure Obtained by Unfolding Simulation At 0.35 ns

Figure 10B shows the snapshots of simulation of Rsim2-HPr over a period of 1.7 ns. The α-helices in PU were similar to the native ones (Fig. 3D). In this simulation, the protein packing proceeded further [~ 1.7 ns, Fig. 10A (1.7 ns)]; R_g and ASA decreased (Figs. 4D and 6C). The C-terminal α-helix, which was located far from the hydrophobic core in Usim-HPr (~0.15 ns, Fig. 8), was packed to the hydrophobic core (Figs. 5G and 10B). Furthermore, subsequent to the α/α packing, pre-β/β packing was also observed, as in the simulation of Rsim1-HPr. As a result, almost all of the tertiary structure was formed in the simulation (Fig. 5G). During a period of 1.7 ns, a short two-stranded β-sheet was formed in the central β-sheet region of the native structure (Figs. 5G and 10B). The short two-stranded β-sheet was not stable (Fig. 3D).

The PR obtained by Rsim2-HPr at 1.7 ns was very close to the PU by Usim-HPr at 0.15 ns (Fig. 10B,C), with Cα atom positional RMSD from the PU by Usim-HPr at 0.15 ns of 3.782 Å. At first, Cα atom positional RMSD of the PU

obtained by Usim-HPr at 0.35 ns, as the initial structure for Rsim2-HPr, from the PU by Usim-HPr at 0.15 ns was 4.737 Å. The structural parameter differences between the PU obtained by Usim-HPr at 0.15 ns and the PR by Rsim2-HPr at 1.7 ns, ΔR_g% and ΔASA% of the polar part and ΔASA% of the nonpolar part, were 0.5, 7.2 and 1.7%, respectively. These results also indicate that the pathway is unique for the local unfolding/refolding process, similar to the results of Rsim1-HPr, PU obtained by Usim-HPr at 0.15ns ↔ (PR by Rsim2-HPr at 1.7ns) ↔ PU by Usim-HPr at 0.35 ns.

Refolding Simulation of HPr from a Partially Unfolded Structure Obtained by Unfolding Simulation At 0.15 ns

Figure 10C shows the results of simulation of Rsim3-HPr over a period of 1.5 ns. During the first 0.2 ns, protein packing (packing of the C-terminal α-helix) proceeded further; and R_g and ASA decreased (Figs. 4D and 6C). As a result of the packing, a tertiary structure, similar to the native structure, was formed (Fig. 5H). Next, a native-like β-sheet developed just like a "closing zipper" from the preexisting "junction" of the β-sheet (between residues 37 and 60) at 0.3 ns (Figs. 3E and 10C). Judging from the structural parameters, RMSD, R_g, ASA, the secondary structure and tertiary structure (Figs. 3E, 4C, 4D, 5H, 6C,D and Table III), we found that the structure of the PR was quite similar to the native structure.

In our calculations, the total simulation period is of the order of a few nanoseconds. We demonstrated several essential steps of folding by our simulations: α − helix → tertiary structure → β − sheet → transition state.The rate constant for folding of HPr in water has been shown to be 14.9 s^{-1} [38], so that it is difficult to describe full refolding simulation of HPr.

Folding Pathways of Psv and HPr

The refolding process of psv started from the formation of an α-helix (from Rsim1 and Rsim2-psv). After the path of α-helix formation, a tertiary structure was formed (from Rsim2-psv). Concomitantly with the formation of the α-helix and tertiary structure, a hydrophobic nucleus was formed. Finally, a β-sheet was formed (from Rsim2 and Rsim3-psv). These results indicate that the folding pathway of psv is random coil → α − helix formation phase → packing (tertiary structure formation) phase → β − sheet formation phase → transitionstate → native.

The refolding process of HPr started from the formation of a hydrophobic nucleus (all α-helices being similar to the native ones) and it continued until the path of β-sheet formation. Subsequently, formation of a tertiary structure (packing) took place (from Rsim1, Rsim2 and Rsim3-HPr). Finally, a β-sheet was formed (Rsim3-HPr). These results suggest that the refolding pathway of HPr is random coil → α − helix formation phase → packing (tertiary structure formation) phase → β − sheet formation phase → transitionstate → native, the same as the refolding pathway of psv.

The refolding and unfolding pathways of psv and HPr were exactly reverse. Thus, the pathway is considered to be a general pathway of folding of small globular proteins. The first formation of an α-helix and the slow formation of a β-sheet observed in our simulations were consistent with experimental observations [3,39]. In this pathway, an α-helix and a tertiary structure were formed concurrently with (in parallel to) the "formation of hydrophobic nucleus" process. This pathway was also found similarly in chymotrypsin inhibitor 2 by experiment [40] and by MD simulation [41]. Accordingly, the nucleation condensation model better describes the folding mechanisms of psv and HPr.

CONCLUSIONS

In this study, we found that the protein folding process of both the psv and HPr is divided into three phases: (I) an α-helix formation phase, (II) a packing (tertiary structure formation) phase, and (III) a β-sheet formation phase. The formation of a hydrophobic nucleus occurred concomitantly with the α-helix formation and packing phase. In the α-helix formation phase, a β-turn appears first, a 3_{10}-helix develops from the β-turn (Fig. 2D), and then the 3_{10}-helix is transformed into an α-helix. During the packing phase, the protein becomes compact, first by α/α packing (formation of a tertiary structure) and then by pre-β/β packing (Fig. 5F). After a long time, probably a few microseconds, a hydrogen bond is formed between the two neighbouring strands; in this study, we did not simulate this step. Once the hydrogen bond is formed, β-sheet formation takes place like a "closing zipper" (Figs. 9B and 10C). Finally, the configuration of the structure may be adjusted (from a transition state) to the native structure.

We successfully followed the process of partial folding by getting together several pieces of refolding simulations. Although complete simulation of the whole folding process is not possible at present, the rapid increase in computational power should enable us to perform a microsecond MD simulation of a protein in the near future; a molecular dynamics machine (MDM) was developed by Narumi *et al.* [42].

Acknowledgements

We are grateful to Dr T. Huber for critical reading of the manuscript and his support in the use of parallelised amber for the VPP supercomputer, and to Prof I. Yamato, Dr K. Yasuoka, Prof P. A. Kollman and Prof M. Karplus for their helpful comments. This work was supported by a Special Postdoctoral Researchers Program from RIKEN (to A.S.).

References

[1] Yankner, B.A., Dawes, L.R., Fisher, S., Villa-Komaroff, L., Oster-Granite, M.L. and Neve, R.L. (1989) "Neurotoxicity of a fragment of the amyloid precursor associated with Alzheimer's disease", *Science* **245**, 417.
[2] Prusiner, S.B. (1997) "Prion diseases and the BSE crisis", *Science* **278**, 245.
[3] Hagen, S.J., Hofrichter, J., Szabo, A. and Eaton, W.A. (1996) "Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding", *Proc. Natl Acad. Sci. USA* **93**, 11615.
[4] Burton, R.E., Huang, G.S., Daugherty, M.A., Calderone, T.L. and Oas, T.G. (1997) "The energy landscape of a fast-folding protein mapped by Ala → Gly substitutions", *Nat. Struct. Biol.* **4**, 305.
[5] Tobias, D.J., Mertz, J.E. and Brooks III, L. (1991) "Nanosecond time scale folding dynamics of a pentapeptide in water", *Biochemistry* **30**, 6054.
[6] Sung, S.-S. and Wu, X.-W. (1996) "Molecular dynamics simulations of synthetic peptide folding", *Proteins* **25**, 202.
[7] Demchuk, E., Bashford, D. and Case, D A. (1997) "Dynamics of a type VI reverse turn in a linear peptide in aqueous solution", *Fold Des.* **2**, 35.
[8] Prevost, M. and Ortmans, I. (1997) "Refolding simulations of an isolated fragment of barnase into a native-like β hairpin: evidence for compactness and hydrogen bonding as concurrent stabilizing factors", *Proteins* **29**, 212.
[9] Daura, X., Jaun, B., Seebach, D. and van Gunsteren, W.F. (1998) "Reversible peptide folding in solution by molecular dynamics simulation", *J. Mol. Biol.* **280**, 925.
[10] Chipot, C., Maigret, B. and A., Pohorille (1999) "Early events in the folding of an amphipathic peptide: a multinanosecond molecular dynamics study", *Proteins* **36**, 383.
[11] Wang, H., Varady, J., Ng, L. and Sung, S.-S. (1999) "Molecular dynamics simulations of β-hairpin folding", *Proteins* **37**, 325.
[12] Bonvin, A.M.J.J. and van Gunsteren, W.F. (2000) "β-hairpin stability and folding: molecular dynamics studies of the first β-hairpin of Tendamistat", *J. Mol. Biol.* **296**, 255.
[13] Duan, Y. and Kollman, P.A. (1998) "Pathway to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution", *Science* **282**, 740.
[14] Skolnick, J. and Kolinski, A. (1990) "Simulations of the folding of a globular protein", *Science* **250**, 1121.
[15] Šali, A., Shakhnovich, E. and Karplus, M. (1994) "How does a protein fold?", *Nature* **369**, 248.
[16] Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D. and Chan, H.S. (1995) "Principles of protein folding-a perspective from simple exact models", *Protein Sci.* **4**, 561.
[17] Lazaridis, T. and Karplus, M. (1997) "New view of protein folding reconciled with the old through multiple unfolding simulations", *Science* **278**, 1928.
[18] Kim, P.S. and Baldwin, R.L. (1982) "Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding", *Annu. Rev. Biochem.* **51**, 459.
[19] Dill, K.A. (1985) "Theory for the folding and stability of globular proteins", *Biochemistry* **24**, 1501.
[20] Baldwin, R.L. and Rose, G.D. (1999) "Is protein folding hierarchic? I. Local structure and peptide folding", *Trends Biochem. Sci.* **24**, 26.
[21] Baldwin, R.L. and Rose, G.D. (1999) "Is protein folding hierarchic?II. Folding intermediates and transition states", *Trends Biochem. Sci.* **24**, 77.

[22] Itzhaki, L.S., Otzen, D.E. and Fersht, A.R. (1995) "The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods", *J. Mol. Biol.* **254**, 260.
[23] Jia, Z., Vandonselaar, M., Hengstenberg, W., Quail, J.W. and Delbaere, L.T.J. (1994) "The 1.6 Å structure of histidine-containing phosphotransfer protein HPr from *Streptococcus faecalis*", *J. Mol. Biol.* **236**, 1341.
[24] Dahiyat, B.I., Sarisky, C.A. and Mayo, S.L. (1997) "*De novo* protein design: towards fully automated sequence selection", *J. Mol. Biol.* **273**, 789.
[25] Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, III, T.E., Rose, III, W.S., Simmerling, III, C.L., Darden, III, T.A., Merz, III, K.M. and Stanton, III, R.V. (1997) AMBER 5.0 (University of California, San Francisco).
[26] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) "Comparison of simple potential functions for simulating liquid water", *J. Chem. Phys* **79**, 926.
[27] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) "A second generation force field for the simulation of proteins and nucleic acids", *J. Am. Chem. Soc.* **117**, 5179.
[28] Berendsen, H.J.C., Postma, J.M.P., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) "Molecular dynamics with coupling to an external bath", *J. Comp. Phys* **81**, 3684.
[29] Ryckaert, J.-P., Ciccotti, G. and Berendsen, J.C. (1977) "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes", *J. Comp. Phys* **23**, 327.
[30] Laskowski, R.A., MacArthur, M.W. and Thornton, J.M. (1993) "PROCHECK: a program to check the stereochemical quality of protein structures", *J. Appl. Crystallog.* **26**, 283.
[31] Koradi, R., Billeter, M. and Wüthrich, K. (1996) "MOLMOL: a program for display and analysis of macromolecular structures", *J. Mol. Graph.* **14**, 51.
[32] Suenaga, A., Komeiji, Y., Uebayasi, M., Meguro, T. and Yamato, I. (1998) "Molecular dynamics simulation of unfolding of histidine-containing phosphocarrier protein in water", *J. Chem. Software* **4**, 127.
[33] Sundaralingam, M. and Selharadu, Y.C. (1989) "Water-inserted α-helical segments implicate reverse turns as folding intermediates", *Science* **244**, 1333.
[34] Miick, S.M., Martinez, G.V., Fiori, W.R., Todd, A.P. and Millhauser, G.L. (1992) "Short alanine-based peptides may form 3(10)-helices and not alpha-helices in aqueous solution", *Nature* **359**, 653.
[35] Williams, S., Causgrove, T.P., Gilmanshin, R., Fang, K.S., Callender, R.H., Woodruff, W.H. and Dyer, R.B. (1996) "Fast events in protein folding: helix melting and formation in a small peptide", *Biochemistry* **35**, 691.
[36] Thompson, P.A., Eaton, W.A. and Holfrichter, J. (1997) "Laser temperature jump study of the helix ⇔ coil kinetics of an alanine peptide interpreted with a "kinetic zipper" model", *Biochemistry* **36**, 9200.
[37] Ando, T. (2000) Personal communication, Science University of Tokyo.
[38] Van Nuland, N.A.J., Meijberg, W., Warner, J., Forge, V., Scheek, R.M., Robillard, G.T. and Dobson, C.M. (1998) "Slow cooperative folding of a small globular protein HPr", *Biochemistry* **37**, 622.
[39] Munoz, V., Thompson, P.A., Hofrichter, J. and Eaton, W.A. (1997) "Folding dynamics and mechanism of β-hairpin formation", *Nature* **390**, 196.
[40] Otzen, D.E., Itzhaki, L.S., Elmasry, N.F., Jackson, S.E. and Fersht, A.R. (1994) "Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding", *Proc. Natl Acad. Sci. USA* **91**, 10422.
[41] Daggett, V., Li, A., Itzhaki, L.S., Otzen, D.E. and Fersht, A.R. (1996) "Structure of the transition state for folding of a protein derived from experiment and simulation", *J. Mol. Biol.* **257**, 430.
[42] Narumi, T., Susukita, R., Ebisuzaki, T., McNiven, G. and Elmegreen, B. (1999) "Molecular dynamics machine: special-purpose computer for molecular dynamics simulations", *Mol. Simul.* **21**, 401.