Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 23 (2004) 59–68
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Modelling of third cytoplasmic loop of bovine rhodopsin by multicanonical molecular dynamics

Yukihisa S. Watanabe [a,*], Yoshifumi Fukunishi [b], Haruki Nakamura [b,c]

[a] *Japan Biological Information Research Center (JBIRC), Japan Biological Informatics Consortium (JBiC), 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, Japan*
[b] *Biological Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology, 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, Japan*
[c] *Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan*

Received 22 August 2003; received in revised form 26 February 2004; accepted 8 April 2004
Available online 20 May 2004

Abstract

The third cytoplasmic loop (C3) of bovine rhodopsin (Rh) is an important site for its interaction with G-protein transducin. The tertiary structure of Rh was determined by X-ray crystallography, although the local conformation around the C3 loop (residues: 236–240) was not visible in electron density maps. We constructed a canonical conformation ensemble at 310 K for the C3 loop (residues: 227–244) using a multicanonical molecular dynamics simulation, and predicted several putative conformations. The conformation ensemble was classified by principal component analysis into several distinct structural clusters, some of which could provide the putative structural models of Rh and the activated state of Rh.

Keywords: Loop modelling; Rhodopsin; Third cytoplasmic loop; Multicanonical molecular dynamics; Principal component analysis; Canonical conformation ensemble

1. Introduction

G-protein-coupled receptors (GPCRs), integral membrane proteins found in the largest family of cell-surface receptors, are responsible for the transduction of several varieties of signals to the cell interior, such as autocrine, paracrine, endocrine, olfactory, gustatory, and ocular signalling factors. Thus, GPCRs are primary drug targets [1], and precise three-dimensional (3D) structures with atomic resolution have been expected for trials of structure-based drug design.

Bovine rhodopsin (Rh) is a typical molecule of GPCRs. It is activated by light stimulation and converted into its active form, metarhodopsin II (Rh^*), through several intermediates: photorhodopsin, bathorhodopsin, lumirhodopsin, and metarhodopsin I [2]. The major intracellular interaction of Rh^* occurs with a signal-transducing protein, heterotrimeric G-protein transducin (G_t). Knowledge of the mechanisms underlying these interactions is essential to understanding the intracellular signal transduction of GPCRs [2–4]. The study of the structural biology of GPCRs started recently, and the first 3D structure of Rh with an atomic resolution appeared in 2000 [5].

Three ground-state structures of Rh are registered in the Protein Data Bank (PDB), with the accession codes 1F88 [5], 1HZX [6], and 1L9H [7]. Two of these crystal structures (1F88 and 1HZX), which were analysed from the same X-ray diffraction data with a resolution of 2.8 Å, showed differences at the second cytoplasmic loop (C2), the third cytoplasmic loop (C3), and the carboxyl terminus region (C4). Disordered regions were found in C3 and C4 in all three crystal structures. In contrast, seven transmembrane helices (TMHs) were determined very clearly. The C3 region of 1L9H is almost identical to that of 1HZX.

The previous mutational experiments and a peptide competition assay suggested strongly that C2 and C3 of Rh^* could both be potential G_t-binding sites [8,9]. Other

* Corresponding author. Tel.: +81-3-3599-8098; fax: +81-3-3599-8099.
E-mail address: yswatanabe@jbirc.aist.go.jp (Y.S. Watanabe).

1093-3263/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2004.04.001

observations suggested that C4 [10] and the eighth helix (H8) [11] are also responsible for G_t-binding. In these three crystal structures, H8 is located between the seventh TMH (TMH7) and C4, lying parallel to the cytoplasmic membrane. The C3 structure, in particular, has been investigated in various experiments: nuclear magnetic resonance (NMR) spectroscopy [12], spin-labelling studies [13], and atomic force microscopy [14]. The NMR study showed that the C3 fragment formed a turn-helix-turn conformation. Although the residues from 231 to 242 formed an imperfect helix, the residues from 247 to 256, which construct the sixth TMH (TMH6) in the native Rh, did not form an α-helix in the C3 fragment [12]. The spin-labelling analysis also indicated that C3 was mainly α-helical, extending the fifth TMH (TMH5) and TMH6 by about 1.5 and 3 turns, respectively [13].

To clarify the 3D structures of the loops, several modelling studies have been conducted, taking the approaches of database analyses and ab initio modelling techniques [15,16]. Both the methods are, in general, limited to 8–12 residues in length [17]. The C3 loop is composed of 21 amino-acid residues, from 226 to 246, and those C_α atoms have mean B-factor values of about 110 $Å^2$ in a crystal structure 1F88, suggesting the very dynamic characteristic of the C3 loop. Thus, the conventional modelling technique might not be able to give any reliable models.

To predict the conformations of such long loops, enhanced conformational sampling based on a statistical mechanism algorithm is powerful enough to give a putative conformational ensemble. Shirai et al. [18] and Kim et al. [19] performed enhanced conformational samplings of the third complementarity-determining region of the antibody heavy chain CDR-H3 by multicanonical molecular dynamics (McMD) simulations, with and without the other surrounding CDR segments. With the surrounding CDR segments, the 7–12 residues in CDR-H3 were found to have predominant conformations at 300 K, corresponding to the X-ray crystal structures. The multicanonical algorithm was first proposed by Berg and Neuhaus [20] to improve sampling efficiency by providing an artificially flat energy spectrum with a 'multicanonical ensemble'. This algorithm has been applied to a variety of systems, including small peptides with implicit and explicit solvent molecules [21,22], as well as to flexible ligand docking studies [23].

We used the McMD approach to study the C3 loop of Rh. Our aim was to obtain the detailed C3 structures and their dynamics, based on a canonical ensemble having almost all possible conformations at 310 K. In contrast, conformations of conventional molecular dynamics (MD) are concentrated at a specific structure, because the MD trajectory depends on the initial structure and tends to get trapped in one of a large number of local minimum-energy states [18–23]. In order to study the cooperative motion of C3 and the other loops, we prepared two models: one composed of only a single C3 loop, and the other composed of the C2, C3, and C4 loops; in both models the N- and C-terminal TMH regions were position-restrained to those of the X-ray crystal structures. Successive principal component analysis (PCA) for the McMD results enabled us to further classify the structural ensembles.

2. Materials and methods

2.1. Initial model structures

The C3 cytoplasmic loop of Rh consists of 21 residues, from 226 to 246, of the native Rh structure. The 3D structure of the first 10 and the last six residues (226–235 and 241–246) are well determined by X-ray crystallography, and so we call these regions "rigid-C3". In contrast, the middle part of the C3 loop, with residues 236–240, is very flexible even in the crystal structures, and we call it "flexible-C3". Fig. 1(A) shows the 3D structures of whole Rh in 1F88 (blue: C2 loop, red: C3 loop, green: C4 loop). Fig. 1(B) shows the 3D structures of the cytoplasmic sides of Rh in 1F88 (blue: C2 loop, red: C3 loop, green: C4 loop, grey: others) and in 1HZX (cyan: C2 loop, purple: C3 loop, yellow: C4 loop, black: others). Table 1 indicates the root-mean-square deviations (RMSDs) of the backbone heavy atoms of Rh in 1F88, 1HZX, and 1L9H, when all the identified backbone heavy atoms of the protein pairs are superimposed as shown in Fig. 1(B).

We constructed two models, one for only a single C3 loop (Model I) and the other for the C2, C3, and C4 loops (Model II). As indicated in Fig. 1(C), Model I consists of 36 successive amino-acid residues, 221–256: the C-terminus of TMH5 (residues: 221–225), C3 (226–246), and the N-terminus of TMH6 (247–256). Model II consists of 59 amino-acid residues with four chains: the C2 loop (138–149), the Model I chain (221–256), the C-terminus of TMH7 with several successive residues (305–312), and the middle of the C4 loop (334–336).

For the initial conformations of Models I and II, the atom coordinates in 1F88 were used, except for the atoms in flexible-C3 from residues 236 to 240. For the flexible-C3 loop, an initial model was constructed using a loop search program, FRGMNT [24], and a molecular modelling program, INSIGHT 2000 (Accelrys Inc., San Diego, CA, USA). The N- and C-termini of the individual peptide chains were capped with the acetyl and *N*-methyl groups, respectively.

Table 1
Structural difference among three crystal structures

Structural pair (PDB IDs)	RMSD (Å)		
	TMH5	C3	TMH6
1F88 vs. 1HZX	0.32	3.09	0.42
1F88 vs. 1L9H	0.36	3.03	0.40
1HZX vs. 1L9H	0.26	0.55	0.31

The average RMSD values of the backbone heavy atoms between pairs of the Rh crystal structures.

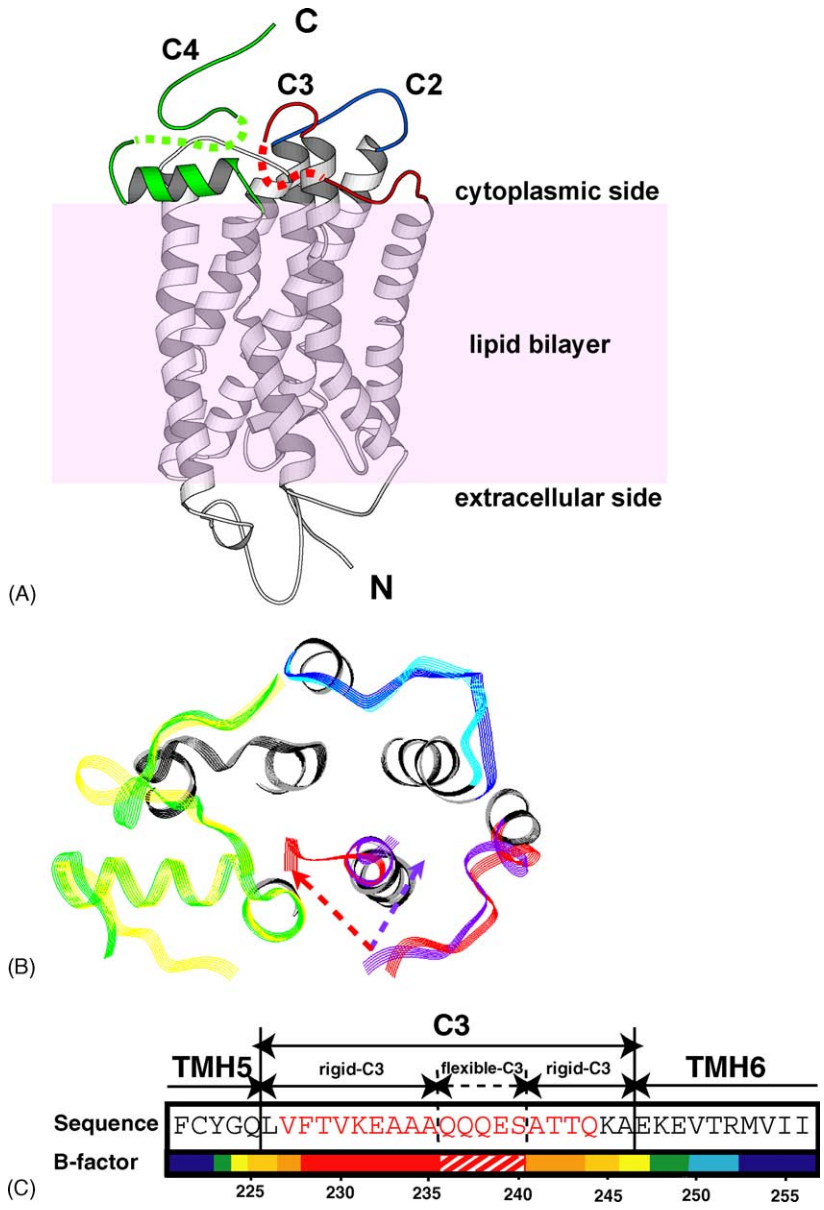

Fig. 1. (A) Ribbon drawing of the Rh crystal structure (PDB code, 1F88) is drawn by MOLSCRIPT [34]. The blue, red, and green represent C2, C3, and C4 loops, respectively. The red and green dotted lines indicate the parts of the C3 and C4 loops, respectively, whose structures were not determined. Seven vertical helices are TMHs. The pink represents lipid bilayer. (B) The cytoplasmic sides of the two Rh crystal structures are indicated by the grey ribbon (1F88) and the black ribbon (PDB code, 1HZX). The blue ribbon is for the C2 (1F88), red is C3 (1F88), green is C4 (1F88), cyan is C2 (1HZX), purple is C3 (1HZX), and yellow is C4 (1HZX). The red (1F88) and purple (1HZX) dotted lines indicate the connections by the flexible-C3 loops. (C) Residues constructing the C3 loop of Rh. The colour of the bar (blue to red) indicates the average B-factors of each Cα atom in 1F88 and 1HZX: blue for $B < 50$ Å^2, sky for 50 Å$^2 \leq B < 60$ Å^2, green for 60 Å$^2 \leq$ B ≤ 70 Å^2, yellow for 70 Å$^2 \leq$ B < 80 Å^2, bright orange for 80 Å$^2 \leq$ B < 90 Å^2, dark orange for 90 Å$^2 \leq$ B < 100 Å^2, and red for 100 Å$^2 \leq$ B. The red hatched bar indicates that those residues have no electron densities in either or both crystal structures 1F88 and 1HZX. The amino acid residues are indicated by one-letter codes. The amino acids with the red and black characters are treated as free and position-restrained, respectively, during the current calculation. The residue numbers appear at the bottom.

During all simulations described below, the atomic coordinates of a part of the C2 loop (140–145) and most of the C3 loop (227–244) were unrestrained, and other amino-acid residues $\{r_i\}$ were restrained to their initial conformations $\{r_i^{\text{init}}\}$ by the following pseudo-potential E_{rst}:

$$E_{\text{rst}} = \tfrac{1}{2}\sum_i C_i k_{\text{B}} T (r_i - r_i^{\text{init}})^2,$$

where $C_i = 0.1 \times$ (mass of the ith atom), k_{B} is Boltzmann constant, and $T = 300$ K.

2.2. Multicanonical molecular dynamics simulations

The multicanonical MD simulation [21] was performed with the AMBER force field [25] by using the program prestoX [26]. The time step, Δt, was 0.5 fs, and the cut-off

distance for non-bonded interactions was 12 Å for every McMD simulation. No explicit solvent molecules were included. Electrostatic interactions between any two atoms were calculated with a dielectric constant equal to twice the distance (Å) between the atom pairs. This simplified electrostatic model was applied so as to realise wide sampling for the long loop of a membrane protein in a limited computation time. Although the accuracy of the ensembles would be increased when we could use more realistic electrostatic models, all atom simulations including solvent molecules is too much time consuming for McMD for the wide conformational sampling. Shirai et al. [18] and Kim et al. [19] performed enhanced conformational samplings of short loops of globular proteins by McMD with the same simplified model, and they succeeded in reconstructing the crystal structures.

The details of the McMD algorithm and the computation were described previously [21]. Here, they are described briefly. The multicanonical ensemble is characterised by the following artificial flat energy distribution, P_{mc}:

$$P_{mc}(E) = Z_{mc}^{-1} n(E) e^{-W(E)} = \text{constant}$$

where E is the potential energy in the system, $W(E)$ the weight function of E, $n(E)$ the density of states, and $Z_{mc} = \sum_E n(E) e^{-W(E)}$. The function $W(E)$ is not given a priori, but instead is determined from a preliminary canonical simulation at a sufficiently high temperature T_o.

$$W(E) = \ln n(E) = \left(\frac{E}{k_B T_o}\right) + \ln P_c(E, T_o)$$

where $P_c(E, T_o)$ is the canonical energy distribution at high temperature, and k_B is the Boltzmann constant. When $P_c(E, T_o)$ does not cover a sufficiently large E range, $W(E)$ is obtained through an iteration of the McMD runs by the following relation:

$$W^{i+1}(E) = W^i(E) + \ln P_{mc}^i(E)$$

Here, T_o should be much higher than the required temperature T, to overcome the local energy barrier. Thus, an iterative procedure is necessary until $P_{mc}(E)$ covers the low energy conformations corresponding to T.

For Models I and II, canonical MD simulations with Hoover–Evans Gaussian constraint thermostat were first carried out for 5×10^6 (2.5 ns) steps at temperature (T) of 700 K. After 7 and 10 iterations, flat energy distributions were obtained from 700 to 310 K (Fig. 2(A) and (B)), respectively. Fig. 2 shows that those energy samplings overcome a large number of local minimum-energy states at 310–700 K. The final McMD simulations were performed with 2.5×10^8 (125 ns) and 1×10^8 (50 ns) steps, respectively. Every 1000 steps, a snapshot of each structure with its energy value were recorded.

2.3. Structural classification

In order to classify the conformational ensemble, the canonical ensemble of N conformations at 310 K was first made by the reweighting formula [20]. The backbone RMSD values were calculated for the all conformation pairs in the canonical ensemble, and a distance matrix of $\{N \times (N + 1)/2\}$ elements was obtained. The conformations were clustered according to the above-mentioned distance matrix by using Ward's method [27] with IMSL (Visual Numerics Inc., San Ramon, CA, USA). In addition, the Principal component analysis (PCA) method was applied in the same way as proposed by Shirai et al. [18] for the sum of the canonical ensembles of Models I and II. The distance matrix was diagonalised, and the resulting eigen vectors represented a conformational space, in which the positions of the N conformations were located.

All MD calculations were carried out on a Compaq ES40 cluster system, and the structural analyses were carried out on a Sun Fire 3800 system.

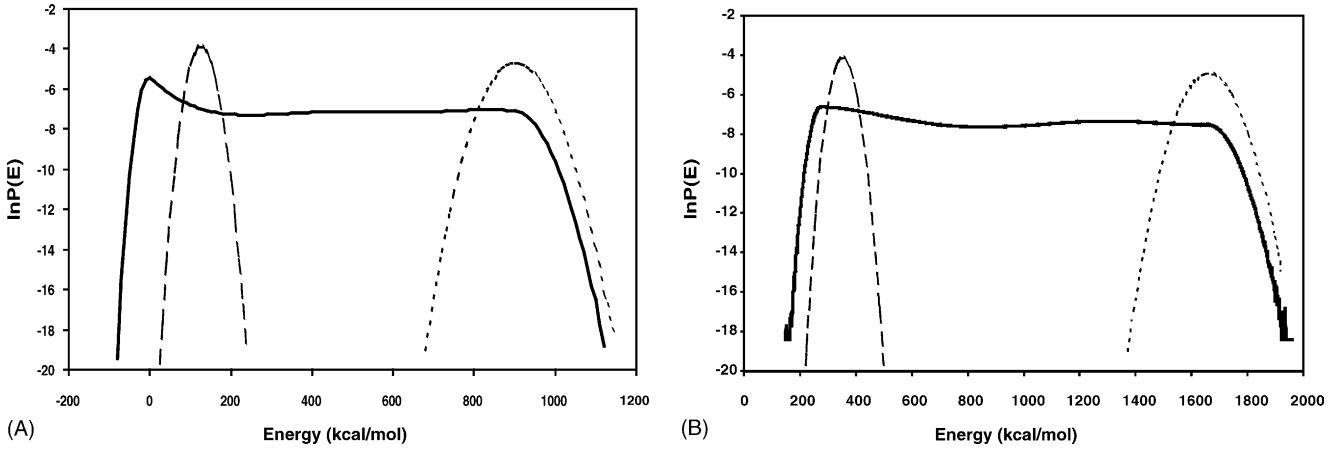

Fig. 2. Energy distributions of multicanonical ensembles (solid lines) for Models: (A) I and (B) II. The dotted curves and dashed curves represent canonical ensembles at 700 and 310 K, respectively, calculated by reweighting the multicanonical ensembles.

3. Results and discussion

3.1. Classification of conformations for Models I and II

We randomly selected 900 structures from the canonical ensembles of Models I and II, respectively, at 310 K, and classified all 1800 structures according to PCA, which was applied to only the backbone heavy atoms of the common residues in C3 (residues: 226–246). Fig. 3 shows the PCA result projected onto a two-dimensional (2D) space composed of first and second principal axes. The contributions from these axes correspond to 38.0 and 21.5% of the total components, respectively.

The clustering analysis was applied to the conformations of Models I and II individually, and Fig. 3 also shows the clustering result with a clustering level of 70 $Å^2$. Each model has three clusters, corresponding to the above PCA results. The clusters of Model I are called 'a1' (61.6% of all clusters), 'b1' (34.7%), and 'c1' (3.8%). Those of Model II are called 'a2' (41.9%), 'b2' (35.1%), and 'c2' (23.0%). Cluster b1 in Model I partly overlaps cluster b2 in Model II, as shown in Fig. 3, but no other clusters overlap with each other.

The numbers of α-helical hydrogen bonds in the flexible C3 residues were counted, and in Fig. 3, the structures having two–five hydrogen bonds are indicated by the open circles. Here, a hydrogen bond was assigned to the pair of the donor and the acceptor when the distance between them was <2.5 Å. The clusters were clearly divided into three types: those with abundant α-helical hydrogen bonds in the flexible C3 as in cluster c1 and b2; those with few as in cluster b1; and those with no such hydrogen bonds as in the other clusters.

The α-helical contents in the conformation ensembles for Models I and II at the individual residues (222–255) were

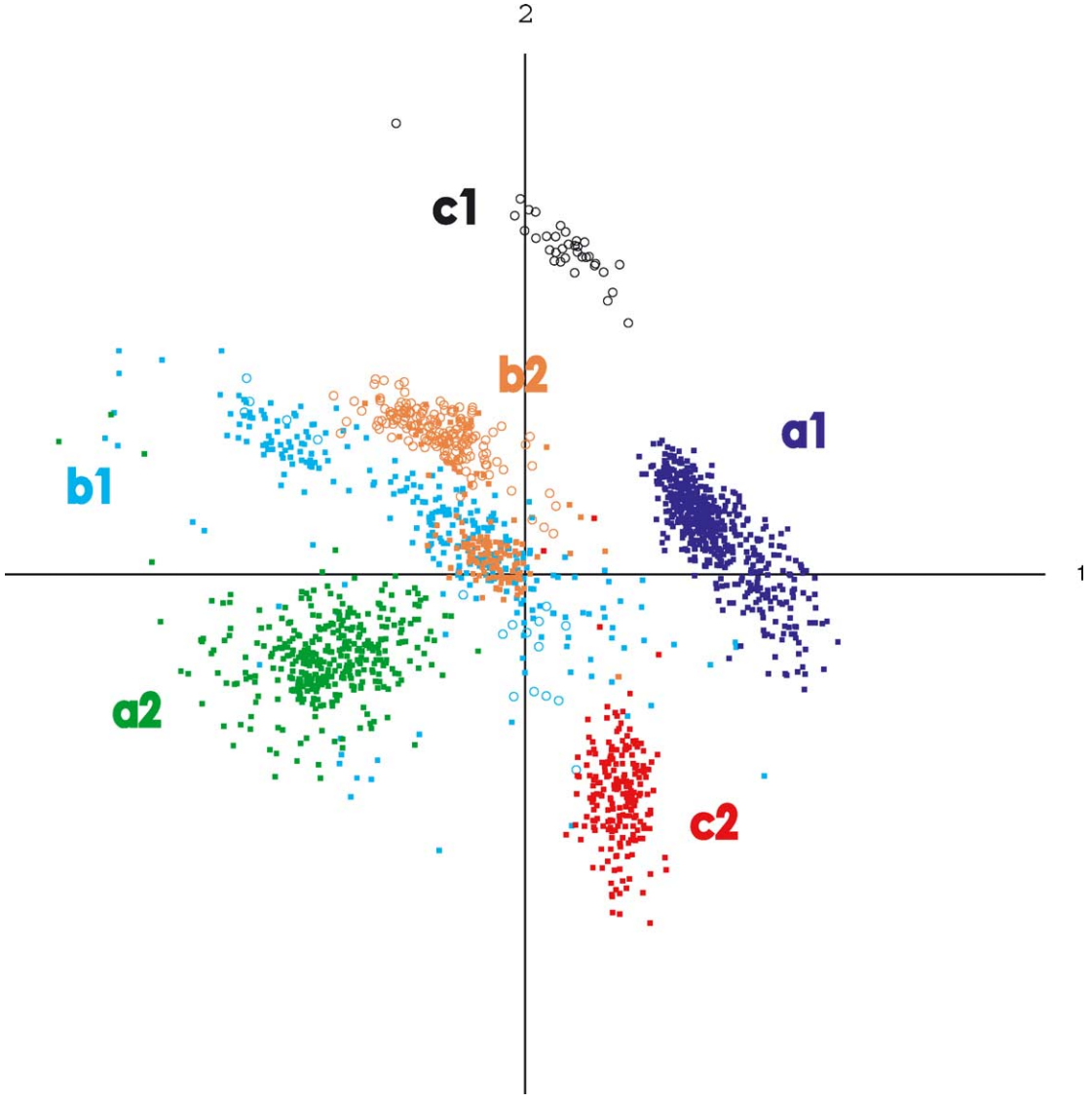

Fig. 3. Distributions of structures of the C3 region on first and second principle axes for the sum of the canonical ensembles at 310 K of Models I and II. The blue, light blue, black, green, orange, and red colours represent clusters a1, b1, c1, a2, b2, and c2, respectively. The closed squares and the open circles correspond to zero–one, and two–five hydrogen bonds in flexible-C3, respectively.

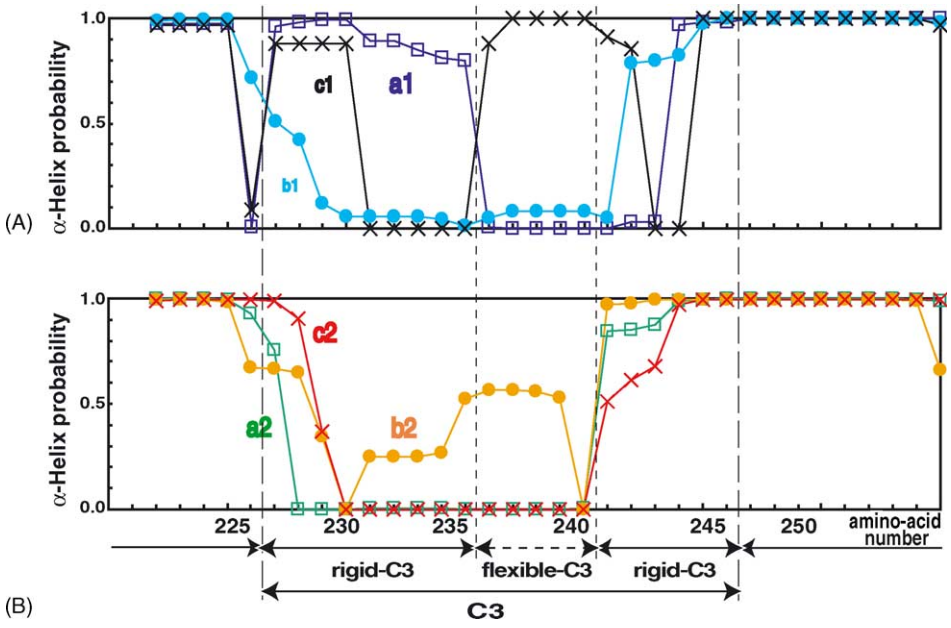

Fig. 4. Relative α-helical conformations in the canonical ensemble at 310 K for Models: (A) I and (B) II by DSSP [28]. (A) The crosses, open squares, and closed circles correspond to clusters a1, b1, and c1, respectively, for Model I. (B) The crosses, open squares, and closed circles correspond to clusters a2, b2, and c2, respectively, for Model II.

also analysed by the DSSP program [28], and the results are shown in Fig. 4(A) and (B), respectively. Clusters b1, a2, and c2 show similar patterns in their residual α-helical contents, although their 3D conformations are very different as shown in Fig. 3.

The average C3 structure of each cluster was calculated by superimposing the backbone heavy atoms on the C3 residues. The RMSD values between all structure pairs are indicated in Table 2. In addition, for each individual cluster we selected the centre structure that had the lowest RMSD value with the corresponding average C3 structure. The centre structures are shown in Fig. 5(A) and (B).

Each centre structure clearly differs from the others. The RMSD values tend to be higher among Model I centre structures than those among Model II centre structures, because Model I includes only a single C3 loop, which shows higher flexibility than actual Rh, without other cytoplasmic loops. In fact, the centre structures for both clusters b1 and c1 have very different rigid-C3 conformations from those found in the crystal structure. These centre structures even bump up against the backbone of the C2 loop. The centre structure for cluster a1 in Model I has low structural diversity, and it has large α-helical contents in the former region of the rigid-C3 (226–235), which is a very different feature from those observed in the crystal structures. Therefore, although cluster a1 is the most abundant in Model I, it may not reflect the C3 loop conformation of actual Rh. Thus, Model I without other surrounding loops did not provide any putative C3 models.

In Model II, Fig. 4(B) shows that the clusters a2 and c2 have no α-helices in the former region of the rigid-C3,

Table 2
Structural differences of C3 region among the centre structures of the current ensemble, the crystal structures, and the previous fragment models

C3 structures	Model I (centre structure) (Å)[a]			Model II (centre structure) (Å)[a]		
	a1	b1	c1	a2	b2	c2
1F88[b]	4.6	6.3	8.2	7.0	5.5	4.3
1HZX[b]	3.8	5.5	6.3	6.5	4.8	3.6
NMR[c]	4.8	4.9	4.0	4.8	4.0	5.3
1LN6[d]	4.5	4.9	4.0	4.6	3.5	4.7
a1	1.6[e]	4.7	6.4	6.1	4.4	3.9
b1	–	3.5[e]	7.2	4.5	3.9	4.5
c1	–	–	1.4[e]	7.8	5.6	7.8
a2	–	–	–	2.3[e]	4.2	5.5
b2	–	–	–	–	2.3[e]	4.2
c2	–	–	–	–	–	1.6[e]

[a] The current model structure, represented by the centre structure of each cluster (see text). The RMSD values of the backbone heavy atoms in the residues from 227 to 244 were calculated between pairs of centre structures.

[b] The crystal structures identified by their PDB codes. The RMSD values of the backbone heavy atoms in the residues from 227 to 235 and 241 to 244 were calculated between the crystal structure and the current model structures that are represented as the centre structures of the clusters.

[c] The C3 fragment structure from residues 231 to 252 analysed by NMR [12]. The RMSD value of the backbone heavy atoms in the residues from 231 to 244 was calculated between the fragment structure and the current model structures.

[d] The C3 fragment structure of the Rh* model [30]. The RMSD value of the backbone heavy atoms in the residues from 231 to 244 was calculated between the fragment structure and the current model structures.

[e] The average RMSD values of the backbone heavy atoms in the residues from 227 to 244 between the individual models in each cluster and the centre structure.

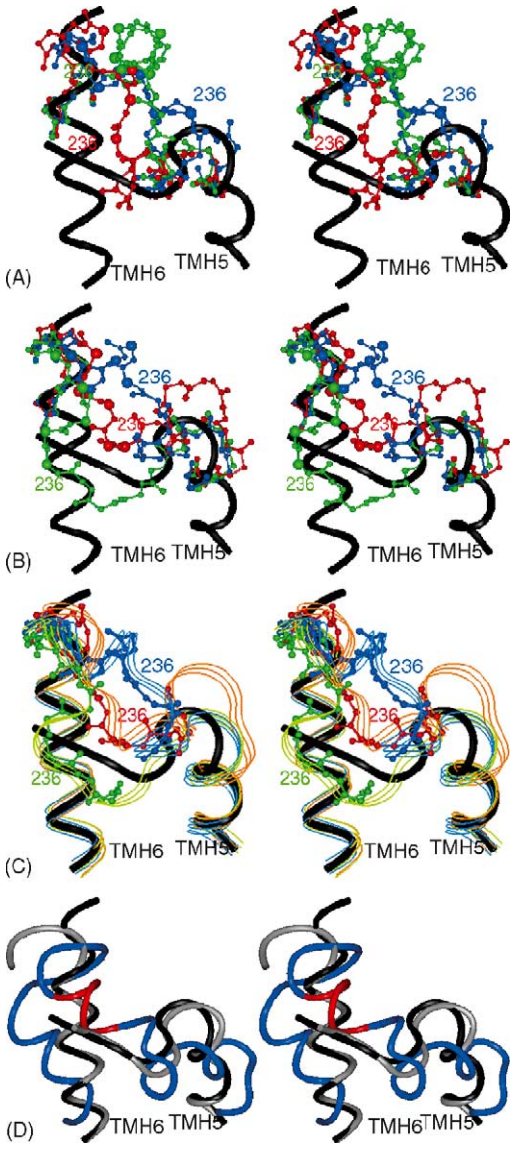

Fig. 5. Stereo drawings of the centre structures of the C3 loops for the six clusters in Fig. 3 3as the representative structures of the reweighted canonical ensembles at 310 K. Each large sphere indicates a C_α atom position of the flexible-C3 (residues: 236–240), and the C_α atom positions of the residue 236 were identified. All black pipe models are the crystal structures for Rh (PDB code, 1HZX), showing a part of the C3 and TMH5 (right) and a part of the C3 and TMH6 (left). (A) The red, blue, and green ball-and-stick models are the backbones of the centre structures of clusters a1, b1, and c1, respectively. Each stick represents a whole C3 loop (226 to 246). (B) The red, blue, and green ball-and-stick models are the backbones of the centre structures of clusters a2, b2, and c2, respectively. Each stick represents a whole C3 loop (226–246). (C) The orange, light blue, and light green ribbons are the backbones of the centre structures of clusters a2, b2, and c2, respectively. The red, blue, and green ball-and-stick models are the backbone of the native fragment structure having 11 residues (233–243), extracted from the database for clusters a2, b2, and c2, respectively. (D) The grey pipe model is the crystal structure for Rh (PDB code, 1F88). The blue-and-red pipe models (blue parts are residues 221–235 and 241–256, and red part is 236–240) show the model structure with the NMR information for Rh* (PDB code, 1LN6) by Choi et al. [30], where the backbone heavy atoms at residues from 221 to 235 and 241 to 256 were superimposed on those of the crystal structure, 1HZX.

similar to the crystal structures. In contrast, cluster b2 is highly helical in the entire C3. Interestingly, this cluster has a low RMSD value with the structure suggested by an NMR experiment [12].

Table 2 also shows the RMSD values among the conformations in the individual clusters. It is clear that the conformational varieties are the largest in cluster b1 (RMSD of 3.5 Å), which have few α-helical contents in C3. In contrast, cluster c2, which has no α-helical contents in C3, appears to have limited structural variation. The rigid-C3 of cluster c2 is the most similar to that of the crystal structure among all the clusters.

3.2. Characteristic features of pseudo-dihedral angles in Model II

In order to identify the local chain conformations that do not assume any typical secondary structure, we calculated the backbone pseudo-dihedral angles $\{\Theta^i\}$, which are formed by four successive C_α atoms ($C_\alpha^{i-1} - C_\alpha^i - C_\alpha^{i+1} - C_\alpha^{i+2}$) for clusters a2, b2, and c2 of Model II (see Fig. 6, from $i = 227$ to 244). Since the N- and C-termini of C3 assume α-helical conformations, frequencies around $\Theta^i = 210$ to 240° are high for residues 227–230 and 243–244.

Cluster a2 has a turn conformation around $i = 230$, 236, 239, and 242 with $\Theta^i = 30$, 60, 90, 300, 330, and 360°, by deforming the helical conformation. From $i = 231$ to 235, the chain conformation is imperfect helical, and at $i = 240$ and 241, there is a large variety in the local conformers of cluster a2. These conformations can also be seen in the average C3 structure of cluster a2, as indicated by the red model in Fig. 5(B). Cluster b2 has a turn at $i = 231$, 232, 236, and 242 with $\Theta^i = 30$, 60, 90, 330, and 360°. At other residues, helix-like conformations are abundant in cluster b2, as shown by the blue model in Fig. 5(B). Cluster c2 has successive turns at $i = 231$–234, and varieties of structures are observed only around $i = 240$–241.

3.3. Peptide fragment structures similar to flexible-C3 of Model II

To determine whether or not the local peptide structures in the conformational ensemble have been observed in native proteins, we examined a local fragment structure search. We looked for structures similar to those of the three centre structures for the C3 loop from residues 233 to 243 in clusters a2, b2, and c2, respectively, by a loop search program, FRGMNT [24]. Using Culledpdb [29], we extracted 2,324,306 fragments from 3,457 PDB files, where only non-homologous proteins, having amino-acid sequence homologies <30%, were examined.

For conformations similar to the clusters of Model II, we obtained hits of 44, 335, and 128 fragments as having conformations similar to clusters a2, b2, and c2, respectively, with RMSD values of <1.5 Å. Thus, those

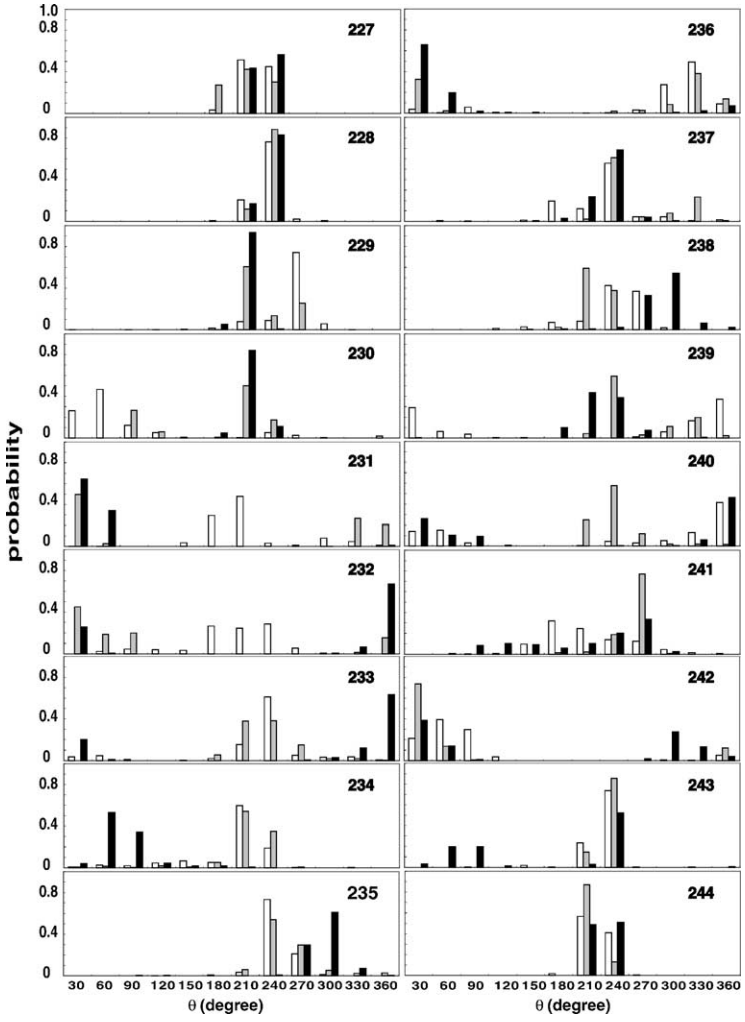

Fig. 6. The frequencies of the pseudo-dihedral angles $\{\Theta^i : C_\alpha^{i-1} - C_\alpha^i - C_\alpha^{i+1} - C_\alpha^{i+2}\}$ in each cluster of Model II for residue i from 227 to 244. White, grey, and black bars indicate the frequencies in clusters a2, b2, and c2, respectively.

cluster conformations are not rare in native proteins. For cluster a2 in particular, the backbone structure from residues 50 to 60 in a dihydrofolate reductase from *Haloferax volcanii* (PDB code, 1VDR B-chain: sequence ESMRDDLPGS) is found to be very similar to the RMSD of 1.00 Å. For clusters b2 and c2, the cell division protein FtsA at residues 370–380 (1E4F: INADEVANDPS) from *Thermotoga maritima* and the soybean lipoxygenase-1 at residues 767–777 (1YGE: RDNPHWTSDSK) are found to have conformations similar to RMSDs of 0.86 and 1.22 Å, respectively. The amino-acid sequence identity is low between these sequences in spite of the high similarity of their structures.

The three fragment conformations are also shown in Fig. 5(C) by superimposing them on the corresponding C3 cluster structures. It should be noted that all of those fragments found in the native proteins are located at the molecular surface, constructing loops. The average B-factors of the C_α atoms in those fragments from 1VDR, 1E4F, and 1YGE are 21.7, 28.5, and 26.0 Å^2, respectively.

3.4. Activated Rh conformation (Rh*)

In a recent NMR and modelling study, Choi et al. [30] reported that C3 of Rh and its activated form Rh* could have imperfect helical conformations (PDB code: 1LN6). Their model structure is also shown by the blue-and-red pipe models (blue parts are residues 221–235 and 241–256, and red part is 236–240) in Fig. 5(D), and the structural comparison is indicated in Table 2. Our current result does not contradict with their Rh* model [30], because one of the clusters, b2, among the conformational ensemble resembles their Rh* model (1LN6) as shown in Table 2.

Those models, including ours, assumed that the TMH conformations of Rh* do not change significantly from those in the ground state of Rh. Farrens et al. suggested that the distance between TMH3 and TMH6 grew upon formation of Rh* [31]. Such a movement should cause different interactions between C3 and the TMHs. Recently, it has been widely recognised that such a conformational change

is generally observed more or less when a protein accomplishes its specific function in the activated form [32,33]. If we have any reliable structural information for the TMHs of Rh^*, our method may provide more realistic loop model for Rh^* in addition to the ground state of Rh.

4. Conclusion

The present study provided the conformational ensemble for the C3 loop region of Rh, including several putative loop structures. In particular, three models with the surrounding loop structures, Model II, provided the putative C3 loop structures as the structural ensemble. The three representative cluster structures have native-like characters with their structural homologs in the crystal structure database. The active structure Rh^* may assume one of those loop structures as the interface with the G-protein. In fact, our ensemble covers the previously proposed Rh^* models. For structure-based drug design (SBDD) against GPCRs, our method could be powerful to sample many putative conformations of G-protein recognition loops, allowing a canonical ensemble to form at room temperature.[1]

Acknowledgement

We thank Prof. Philip L. Yeagle (University of Connecticut) for providing us the coordinates of the C3 fragment from his group's NMR study [12]. We thank Prof. Junichi Higo (Tokyo University of Pharmacy and Life Science), Drs. Tetsuji Okada (BIRC/AIST), Narutoshi Kamiya (BERI), Nobuyuki Nakajima (Institute for Protein Research, Osaka University), and Masataka Kuroda (Tanabe Seiyaku Co., Ltd.) for valuable discussions. This work was supported by grants from the New Energy and Industrial Technology Development Organisation (NEDO) and the Ministry of Economy, Trade and Industry (METI) of Japan.

References

[1] A. Wise, K. Gearing, S. Rees, Target validation of G-protein coupled receptors, Drug Discov. Today 7 (2002) 235–246.
[2] Y. Shichida, H. Imai, Visual pigment: G-protein-coupled receptor for light signals, Cell. Mol. Life Sci. 54 (1998) 1299–1315.
[3] T. Okada, O.P. Ernst, K. Palczewski, K.P. Hofmann, Activation of rhodopsin: new insights from structural and biochemical studies, Trends Biochem. Sci. 26 (2001) 318–324.
[4] R. Arimoto, O.G. Kisselev, G.M. Makara, G.R. Marshall, Rhodopsin–transducin interface: studies with conformationally constrained peptides, Biophys. J. 81 (2001) 3285–3293.
[5] K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, M. Miyano, Crystal structure of rhodopsin: a G protein-coupled receptor, Science 289 (2000) 739–745.
[6] D.C. Teller, T. Okada, C.A. Behnke, K. Palczewski, R.E. Stenkamp, Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs), Biochemistry 40 (2001) 7761–7772.
[7] T. Okada, Y. Fujiyoshi, M. Silow, J. Navarro, E.M. Landau, Y. Shichida, Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 5982–5987.
[8] T. Yamashita, A. Terakita, Y. Shichida, Distinct roles of the second and third cytoplasmic loops of bovine rhodopsin in G protein activation, J. Biol. Chem. 275 (2000) 34272–34279.
[9] K. Yang, D.L. Farrens, W.L. Hubbell, H.G. Khorana, Structure and function in rhodopsin: single cysteine substitution mutants in the cytoplasmic interhelical E–F loop region show position-specific effects in transducin activation, Biochemistry 35 (1996) 12464–12469.
[10] A.G. Krishna, S.T. Menon, T.J. Terry, T.P. Sakmar, Evidence that helix 8 of rhodopsin acts as a membrane-dependent conformational switch, Biochemistry 41 (2002) 8298–8309.
[11] W.J. Phillips, R.A. Cerione, A C-terminal peptide of bovine rhodopsin binds to the transducin alpha-subunit and facilitates its activation, Biophys. J. 299 (1994) 351–357.
[12] P.L. Yeagle, J.L. Alderfer, A.D. Albert, Structure of the third cytoplasmic loop of bovine rhodopsin, Biochemistry 34 (1995) 14621–14625.
[13] C. Altenbach, K. Yang, D.L. Farrens, Z.T. Farahbakhsh, H.G. Khorana, W.L. Hubbell, Structural features and light-dependent changes in the cytoplasmic interhelical E–F loop region of rhodopsin: a site-directed spin-labelling study, Biochemistry 35 (1996) 12470–12478.
[14] J.B. Heymann, M. Pfeiffer, V. Hildebrandt, H.R. Kaback, D. Fotiadis, B. Groot, A. Engel, D. Oesterhelt, D.J. Muller, Conformations of the rhodopsin third cytoplasmic loop grafted onto bacteriorhodopsin, Struct. Fold Des. 8 (2000) 643–653.
[15] N. Go, H.A. Scheraga, Ring closure and local conformational deformations of chain molecules, Macromolecules 3 (1970) 178–187.
[16] R.E. Bruccoleri, M. Karplus, Prediction of the folding of short polypeptide segments by uniform conformational sampling, Biopolymers 26 (1987) 137–168.
[17] A. Fiser, R.K. Do, A. Sali, Modeling of loops in protein structures, Protein Sci. 9 (2000) 1753–1773.
[18] H. Shirai, N. Nakajima, J. Higo, A. Kidera, H. Nakamura, Conformational sampling of CDR-H3 in antibodies by multicanonical molecular dynamics simulation, J. Mol. Biol. 278 (1998) 481–496.
[19] S.T. Kim, H. Shirai, N. Nakajima, J. Higo, H. Nakamura, Enhanced conformational diversity search of CDR-H3 in antibodies: role of the first CDR-H3 residue, Proteins 37 (1999) 683–696.
[20] B.A. Berg, T. Neuhaus, Multicanonical ensemble: a new approach to simulate first-order phase transitions, Phys. Rev. Lett. 68 (1992) 9–12.
[21] N. Nakajima, A. Kidera, H. Nakamura, Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides, J. Phys. Chem. B 101 (1997) 817–824.
[22] N. Kamiya, J. Higo, H. Nakamura, Conformational transition states of a beta-hairpin peptide between the ordered and disordered conformations in explicit water, Protein Sci. 11 (2002) 2297–2307.
[23] N. Nakajima, J. Higo, A. Kidera, H. Nakamura, Flexible docking of a ligand peptide to a receptor protein by multicanonical molecular dynamics simulation, Chem. Phys. Lett. 278 (1997) 297–301.
[24] H. Nakamura, K. Katayanagi, K. Morikawa, M. Ikehara, Structural models of ribonuclease H domains in reverse transcriptases from retroviruses, Nucl. Acids Res. 19 (1991) 1817–1823.
[25] S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case, An all atom force field for simulations of proteins and nucleic acids, J. Comput. Chem. 7 (1986) 230–252.
[26] Y. Fukunishi, Y. Mikami, H. Nakamura, The filling potential method: a method for estimating the free energy surface for protein–ligand docking, J. Phys. Chem. B 107 (2003) 13201–13210.
[27] R.M. Anderberg, Cluster Analysis for Applications, Academic Press, New York, 1973 , p. 139.

[1] Inquiries about the current model structures should be addressed to the corresponding author.

[28] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[29] U. Hobohm, M. Scharf, R. Schneider, C. Sander, Selection of representative protein data sets (http://www.fccc.edu/research/labs/dunbrack/culledpdb.html), Protein Sci. 1 (1992) 409–417.
[30] G. Choi, J. Landin, J.F. Galan, R.R. Birge, A.D. Albert, P.L. Yeagle, Structural studies of metarhodopsin II, the activated form of the G-protein coupled receptor, rhodopsin, Biochemistry 41 (2002) 7318–7324.
[31] D.L. Farrens, C. Altenbach, K. Yang, W.L. Hubbell, H.G. Khorana, Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin, Science 274 (1996) 768–770.
[32] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, W. Chiu, E.C. Garner, Z. Obradovic, Intrinsically disordered protein, J. Mol. Graph. Model. 19 (2001) 26–59.
[33] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm, J. Mol. Biol. 293 (1999) 321–331.
[34] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, J. Appl. Cryst. 24 (1997) 946–950.