Identification of core amino acids stabilizing rhodopsin

A. J. Rader*, Gülsüm Anderson†, Basak Isin*, H. Gobind Khorana‡, Ivet Bahar*, and Judith Klein-Seetharaman*†‡§

*Center for Computational Biology and Bioinformatics, Department of Molecular Biology and Biochemistry, and †Department of Pharmacology, School of Medicine, University of Pittsburgh, 200 Lothrop Street, Pittsburgh, PA 15261; and ‡Departments of Biology and Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by H. Gobind Khorana, March 1, 2004

Rhodopsin is the only G protein-coupled receptor (GPCR) whose 3D structure is known; therefore, it serves as a prototype for studies of the GPCR family of proteins. Rhodopsin dysfunction has been linked to misfolding, caused by chemical modifications that affect the naturally occurring disulfide bond between C110 and C187. Here, we identify the structural elements that stabilize rhodopsin by computational analysis of the rhodopsin structure and comparison with data from previous *in vitro* mutational studies. We simulate the thermal unfolding of rhodopsin by breaking the native-state hydrogen bonds sequentially in the order of their relative strength, using the recently developed Floppy Inclusion and Rigid Substructure Topography (FIRST) method [Jacobs, D. J., Rader, A. J., Kuhn, L. A. & Thorpe, M. F. (2001) *Proteins* 44, 150–165]. Residues most stable under thermal denaturation are part of a core, which is assumed to be important for the formation and stability of folded rhodopsin. This core includes the C110—C187 disulfide bond at the center of residues forming the interface between the transmembrane and the extracellular domains near the retinal binding pocket. Fast mode analysis of rhodopsin using the Gaussian network model also identifies the disulfide bond and the retinal ligand binding pocket to be the most rigid region in rhodopsin. Experiments confirm that 90% of the amino acids predicted by the FIRST method to be part of the core cause misfolding upon mutation. The observed high degree of conservation (78.9%) of this disulfide bond across all GPCR classes suggests that it is critical for the stability and function of GPCRs.

network models | membrane protein | folding | G protein-coupled receptor | simulation

Rhodopsin is the only member of the G protein-coupled receptors (GPCRs), the largest family of cell-surface receptors, whose 3D structure is known (1). The signature motif of the GPCR family is a bundle of seven-transmembrane (TM) helices connected by polypeptide loops that form the cytoplasmic (CP) and the extracellular (EC) domains on opposite sides of the TM domain (Fig. 1). GPCRs perform extremely diverse and vital functions that include responses to light, odor, taste, neurotransmitters, hormones, and a variety of other signals (2). Whereas, in rhodopsin and related visual pigments, the ligand 11-*cis*-retinal (RET) is covalently bound to the apoproteins (opsins), all other GPCRs occur in the ligand-free form, and subsequent binding of appropriate ligand(s) results in their activation. There is a wide variation in the nature of the ligands and their binding modes such as direct binding to the TM domain, the EC domain, or both.

Based on pharmacological specificity and sequence conservation, GPCRs are divided into eight classes (3). Although there is no sequence homology between GPCRs in different classes, the seven-TM helix motif is conserved throughout, and all GPCRs share a common topology. They can be grouped into three main classes: receptors related to rhodopsins (class A), secretin receptors (class B), and the metabotropic neurotransmitter receptors (class C). Of these, class A, the largest class, contains >1,200 distinct members and >7,000 putative members (4) and has been studied the most.

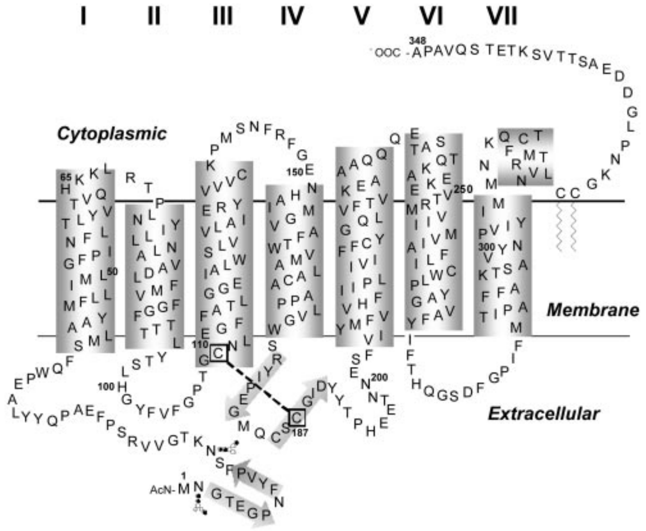

Fig. 1. Secondary structures in bovine rhodopsin. The seven-TM helices are shown by numbered gray boxes, and β-strands are shown by arrows. The respective residue ranges of these TM helices are as follows: I, 35–60; II, 71–100; III, 107–137; IV, 151–173; V, 200–225; VI, 247–277; VII, 286–306; VIII, 310–324. The dashed line indicates the C110—C187 disulfide bond located at the interface between the TM and EC domains.

A large number of naturally occurring point mutations has been characterized in rhodopsin (5, 6). The majority of these mutations are associated with *Retinitis pigmentosa* (RP), a disease that ultimately leads to photoreceptor degradation and consequent loss of vision. RP mutations are found in each of the three structural domains of rhodopsin. However, most of the mutations are found in the TM and EC domains (5–9). Mutations in the EC domain cause partial to complete misfolding, misfolding being defined as the loss of ability to bind RET (10–14). Studies of both naturally occurring RP and designed mutations in the EC domain showed that misfolding involved the formation of a nonnative disulfide bond between C185 and C187 instead of the naturally occurring C110—C187 bond (15). Furthermore, studies of RP mutations in the TM domain of rhodopsin showed that they also cause misfolding by formation of an abnormal disulfide bond (16, 17), identified by mass spectrometric analysis to be between C185 and C187 (18). The abnormal disulfide bond was the same regardless of whether the mutations were in the TM or EC domain. This finding suggested

Abbreviations: GPCR, G protein-coupled receptor; RET, retinal; EC, extracellular; TM, transmembrane; CP, cytoplasmic; GNM, Gaussian network model; FIRST, Floppy Inclusion and Rigid Substructure Topography; PDB, Protein Data Bank.

§To whom correspondence should be addressed. E-mail: jks33@pitt.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0401429101

that packing of the helices in the TM domain and folding to a tertiary structure in the EC domain are coupled (19). However, the molecular details of the relationship between the disulfide bond and the coupling between the structures in the TM and EC domains are not known.

The recently published 2.6-Å resolution crystal structure of rhodopsin [Protein Data Bank (PDB) ID code 1L9H] (20) provides an opportunity to identify by computational techniques the regions critical for folding. We simulated the unfolding of rhodopsin using the Floppy Inclusion and Rigid Substructure Topography (FIRST) method (21) to identify the amino acids that form the structural stability core. Additionally, we applied the mode decomposition of the Gaussian network model (GNM) (22, 23) to investigate residues that play a key role in maintaining the folded state. These computational methods generate predictions about the potential flexibility and mobility of each amino acid and, from these, predict sites that are crucial for folding and stability (24–27). The methods are applied here to the crystal structure of rhodopsin, and the resulting predictions are compared with experimental mutational data. The predicted core includes amino acids at the interface between the TM and EC domains, providing a molecular explanation for the experimentally observed importance of the structural coupling between EC and TM domains for folding and stability of rhodopsin.

Methods

Simulated Unfolding Using FIRST Software. FIRST (21) uses techniques from graph theory to analyze and quantify the rigidity or flexibility of proteins. Given the atomic coordinates of a protein, each bond is identified as either flexible (free to rotate) or rigid (nonrotatable) according to geometric criteria. For rhodopsin, all-atom calculations were performed on molecule "A" from PDB entry 1L9H, including all ligands and buried water molecules. Buried water molecules were defined by PROACT (28), and polar hydrogen positions were optimized for hydrogen bonding by WHATIF (29). FIRST software is accessible at http://firstweb.asu.edu.

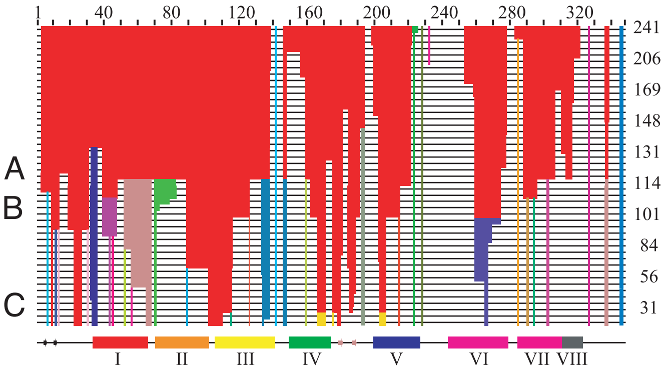

Fig. 2. Simulated thermal denaturation plot for rhodopsin (PDB ID code 1l9h). Each line in this hydrogen bond dilution plot depicts which residues are rigid and flexible with a certain set of hydrogen bonds present. Along the right ordinate are indicated the numbers of present hydrogen bonds for some steps. Thin black lines represent residues with a flexible backbone, and each colored block identifies which rigid cluster a residue belongs to. As one moves down the hydrogen bond dilution plot, hydrogen bonds are removed one at a time based on energy. Lines are shown only when there is a change in the backbone rigid clusters. The PDB-defined helical regions are shown as colored blocks numbered I–VIII, and the four β-strands are indicated by arrows along the line immediately below the dilution plot. Removing hydrogen bonds according to energy is analogous to thermal denaturation and hence simulates protein unfolding (26, 32). Lines representative of a native-like structure, transition state, and folding core are labeled A, B, and C, respectively. See Fig. 3 and the text for the definitions of the transition state and folding core.

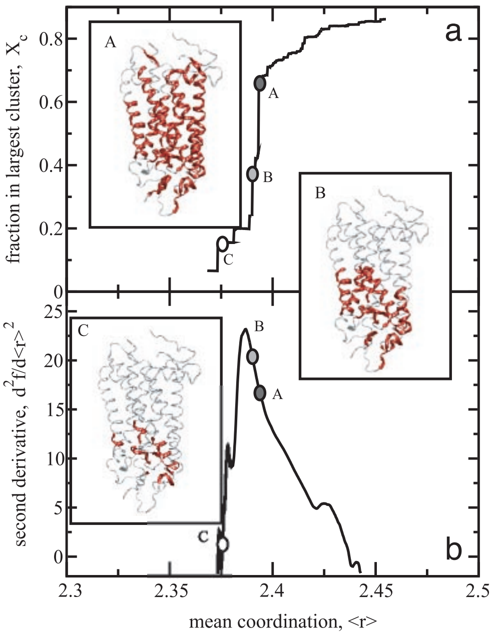

Fig. 3. Simulated unfolding of rhodopsin. (*a*) Order parameter. (*b*) Specific heat-like curve. The fraction of the number of atoms participating in the largest cluster, X_c, as a function of the mean coordination number, $<r>$, serves as the order parameter for this system. The peak in the second derivative of the number of floppy modes with respect to $<r>$ is used to identify the transition state (B). *Insets* show 3D images of the structure along the unfolding pathways, corresponding to positions A–C from Fig. 2. The largest rigid cluster, shown by red ribbons, decreases as the protein unfolds from the native-like state (A) through the transition state (B) to the folding core (C). These and other 3D images were created by using VMD (54).

Protein Dynamics Using the GNM. In the GNM (22, 30), the protein is modeled as an elastic network composed of beads and springs connecting all interacting residues. A connectivity, or Kirchhoff, matrix (Γ) is constructed, which identifies the residues that are within a certain cutoff distance with respect to one another. The native state of the protein, taken from the PDB, is used to identify contacts typically within a cutoff distance of r_c = 7 Å. The collective modes of motion are uniquely defined by the particular topology of contacts. The eigenvalues of Γ represent the frequencies of the N-1 non-zero GNM modes, and the eigenvectors describe the distribution of residue mobilities corresponding to each mode. The inverse of Γ describes the correlations between residue fluctuations near the native state (22). The diagonal elements of the inverse Kirchhoff matrix scale with the mean-square fluctuations, which can be expressed as a weighted sum over all modes (23).

Statistical Analysis of a Conserved Disulfide Bond in GPCRs. Cysteine residue pairs in GPCRs that could potentially form a disulfide bond structurally "equivalent" to the C110—C187 disulfide bond in rhodopsin were detected as follows. TM helix predictions, amino acid sequence alignments, and snake-like plots, similar to Fig. 1, generated by VISEUR software (31) were inspected for each of the 2,962 GPCR sequences in the GPCR database (3) available as of June 2002. Cysteine residues were assumed to be equivalent to C110 if they were located within the lower (EC side) half of helix III or upper half of the first EC loop, and equivalent to C187 if they were within the predicted second EC loop region or the first turn of either helix IV or V. Each receptor sequence was then classified as having the disulfide bond if it contained both equivalent cysteines (as in Fig. 1) or as lacking the disulfide bond if either one or both cysteines were absent.

Results

Simulated Thermal Unfolding. Thermal denaturation is simulated with FIRST by breaking hydrogen bonds in the protein one at a time and recalculating the rigid and flexible regions at each step (26, 32). Fig. 2 shows the resulting dilution profile (26) for rhodopsin. Each line gives a schematic representation of the distribution of the rigid (thick, colored) and flexible (thin, black) regions along the sequence as the structure gradually unfolds. Each colored region represents a different rigid cluster, which can include sequentially distant but spatially close segments because the rigid regions are a consequence of the underlying 3D structure. The number of hydrogen bonds present is indicated for some of the lines in Fig. 2. Often hydrogen bonds can be removed from rigid regions without causing any change in the rigid or flexible regions, and such inconsequential steps are omitted from Fig. 2. For a given dilution profile, the transition state (B) is defined by the peak in the second derivative of the number of "floppy modes" with respect to $<r>$, the mean coordination or average number of neighbors for each atom. These floppy modes quantify the available degrees of freedom that are associated with dihedral rotations. A previous study found a universal unfolding transition for proteins at $<r> = 2.405 \pm 0.015$ (32). The folding core (C) is defined as the last (lowest) line in the dilution profile in which at least three residues of two or more secondary structures are part of the same rigid cluster. This definition has been shown to correlate well with slow exchanging folding cores indicated by hydrogen/deuterium exchange experiments on proteins (26, 27).

Scanning down the dilution profile in Fig. 2, the reduction in red-colored regions corresponds to a loss in rigidity as hydrogen bonds are removed. A–C indicate specific points along the unfolding pathway. Even in the initial state (the first line of the denaturation profile) there are flexible regions, such as the third CP loop, the N-terminal part of helix VI, and the C-tail. Between this mostly rigid state and the native-like state (A) the loss of rigidity is gradual. The first regions to gain flexibility are the CP ends of the TM helices IV–VI. From state A through the transition state (B) to the stability core (C) there is a more dramatic loss of rigidity. These three highlighted points in the simulated unfolding pathway (A–C) are mapped onto the 3D structure of rhodopsin and are shown in Fig. 3. Fig. 3 also indicates where these points lie along the curves corresponding to the order parameter (*a*) and specific heat-like quantity (*b*) (32). Both quantities are plotted against the mean coordination number, $<r>$, to indicate the catastrophic change in rigidity as the protein is "unfolded."

The core (C) consists of parts of TM helices III–V, the β-sheet in the second EC loop region, and parts of the first and second EC loops. Specifically, it includes residues 9, 10, 22–27, 102–116, 166–171, 175–180, 185–188, 203–207, and 211. Both C110 and C187 forming the disulfide bond are part of this core and remain mutually rigid even in the final line of the dilution plot. The potential unfolding pathway of rhodopsin presented in Fig. 2 suggests that the EC ends of helices III–V form the most stable region in the protein because they remain mutually rigid until all hydrogen bonds have been removed. Much of the region identified as the stability core in Fig. 2 overlaps with the RET binding region (within 4.5 Å) including E113, S186, M207, and H211. Repeating the simulated denaturation with all protein–RET interactions removed produced no significant change in the unfolding pathway (data not shown). This finding suggests that the stability core may also serve as a rigid docking site for RET binding during the folding process.

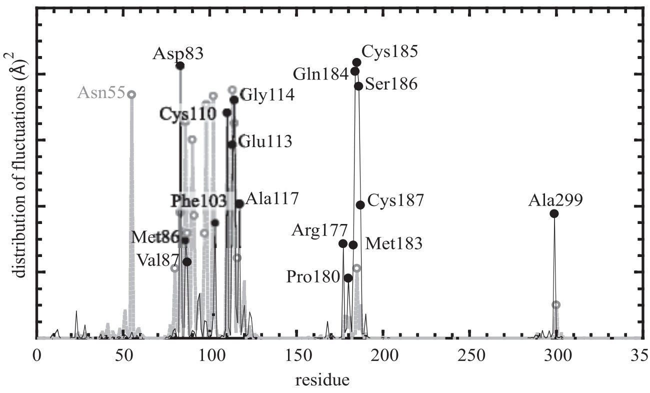

Fig. 4. Distribution of fluctuations in the high frequency modes of GNM. The solid curve with filled circles displays the mode shape averaged over the fastest 10 GNM modes calculated with $r_c = 10$ Å. The peaks indicate the residues that are most likely to participate in the folding nucleus of rhodopsin. Sixteen residues with values above a threshold of 6 N^{-1}, where $N = 348$ residues, are labeled. The dashed gray line refers to the results with $r_c = 7$ Å, which indicate more localized fluctuations. Only the peak at N55 is labeled. Additional peaks (not labeled) are at A80, D83, F91, T97, S98, and Y102.

Gaussian Network Fast Mode Peak Residues in Rhodopsin. The slow modes extracted by mode decomposition of the GNM dynamics (or by conventional normal mode analysis of equilibrium structures) typically determine the global (or essential) motions associated with biologically relevant functions (22, 24). Although the high frequency end of the spectrum is usually viewed as "uninteresting" in normal mode analysis, the peaks in the fastest GNM modes identify the residues that maintain structural integrity by resisting conformational motion (25). Previous studies have shown that these residues correlate with experimentally determined folding nuclei (24, 25, 27). Fig. 4 plots the mode shapes as averages of the fastest 10 modes for two cases: $r_c = 10$ Å (black solid traces) and $r_c = 7$ Å (gray dotted traces). The results are influenced by the range of interactions considered in the analysis because the high frequency modes are highly localized. Increasing r_c from 7 to 10 Å better captures the cooperative interactions that stabilize the folding core by identifying the residues that participate in slightly larger local clusters. These 16 peak residues are D83, M86–V87, **F103**, **C110**, **E113–G114**, A117, **R177**, **P180**, M183, Q184, **C185**, **S186**, **C187**, and A299. Included in this set are the critical disulfide bond cysteine residues C110 and C187; nine residues identified by both FIRST and GNM (listed in italics here and shown by cyan tubes in Fig. 5*B*); and other residues in TM helices II and III. The correlation between a majority of these residues and the core residues identified by FIRST suggests that these residues are structurally important for stabilizing the folded state.

Experimental Validation of the Folding Core. To test the relevance of these computational results, the residues predicted to be critical for folding by FIRST and GNM analysis were compared to mutational data bearing on the folding of rhodopsin. A data set was extracted from the literature containing point mutations and deletions of four or fewer residues. If the mutations or deletions allowed normal binding of RET, they were considered to be correctly folded. If the mutants did not bind RET, they were considered completely misfolded. Mutants between these extremes were grouped together as partially misfolded (14). The deficiency of RET binding is not a direct indicator for misfolding, and it would be preferable to use direct evidence for misfolding such as mistrafficking (33). However, there exists good correlation between deficiency of RET binding and mistrafficking in which both types of data are available (9, 12).

Table 1. Experimental data

Number of residues experimentally studied	164
Fraction observed not to affect folding	91/164
Fraction observed to cause misfolding	73/164
Fraction that cause complete misfolding	40/73

Because opsin is thermally less stable than rhodopsin (34), the term misfolding is used here to collectively describe misfolded and RET binding deficient mutants. See Tables 1 and 2 to compare the predicted folding core and the experimental data (10–14, 16, 17, 35–38). Mutagenesis experiments have been performed for 164 residues, 91 of which did not affect folding, whereas the remaining 73 caused misfolding (40 complete and 33 partial). The computational results are presented for three sets: 45 stability core residues identified by FIRST, 16 GNM fast mode peaks, and 52 residues from the union of these two sets. For the 39 of these 52 residues with experimental data, each computational method correctly predicts >78% to cause misfolding (complete and partial). Thirty-one of the 34 (91%) FIRST stability core residues are known to cause misfolding. The observed high degree of correlation is remarkable. Additionally, 26 of the above 31 residues (83.4%) are known to cause complete misfolding. These data suggest that the residues identified by FIRST are significant for initiation of proper folding.

Statistical Analysis of a Conserved Disulfide Bond in GPCRs Equivalent to That in Rhodopsin. To assess the significance of the findings from GNM and FIRST analysis of rhodopsin for the GPCR family, the occurrence of disulfide bonds in GPCRs that may be structurally equivalent to the C110—C187 disulfide bond in rhodopsin was determined (see *Methods*). Table 3 indicates the degree of conservation calculated for each GPCR class. For all but the final two classes (putative and orphans) there is >87% conservation of a pair of cysteines potentially forming a disulfide bond. Even including these two classes, a large majority (78.9%) of all GPCR sequences have both analogous cysteine residues. Among the 2210 sequences comprising the well established GPCR classes A–C, >92% contain both cysteine residues. Of the 626 sequences that were lacking either one or both of the cysteines, 184 lacked the C187 equivalent, 220 lacked the C110 equivalent, and 222 lacked both cysteines. Within the classes that display a high degree of conservation, often a few subclasses lack the putative disulfide bond. Two subclasses of class A, cannabis and lysosphingolipid receptors, have 0% conservation, suggesting that these subclasses have evolved an alternative mechanism for stability. The overall high conservation of these two cysteines suggests their important role for GPCR structure and function in general.

Discussion

Residues Important for Rhodopsin Stability. In this article, we identify the amino acids important for rhodopsin stability by analysis of the crystal structure using two computational techniques, FIRST and GNM. The sets of stable residues obtained by the two methods differ to some extent, although both methods assume that information about the folding process is encoded in the native conformation. However, the GNM adopts an elastic network description and provides insights about the functional motions near the native state. The FIRST simulated unfolding method systematically alters the network to mimic unfolding and reanalyzes the structure at each step. The GNM fast mode shapes are not as robust as the slow mode shapes, and core residues predicted from fast mode peaks are usually interpreted in conjunction with complementary data from experiments or more detailed simulations. Fig. 5 summarizes the results obtained with the two methods. The most stable residues found by FIRST and GNM are shown in red and green, respectively. Both methods combined identify 52 residues. Nine residues are predicted to form the stability core by both methods, shown in Fig. 5*A* in cyan and in yellow for the cysteines forming the disulfide bond. The local environment for these core residues is shown in Fig. 5*B*. They are positioned at the TM–EC interface of the RET binding pocket, surrounded by the remaining core residues predicted by FIRST or GNM. The importance of these folding core amino acids is strongly validated by experimental evidence. In particular, the disulfide bond plays a critical role in the folding and stability of rhodopsin (15–18, 39), and Fig. 5 shows that it serves as an anchor for many interactions between RET in the TM domain and the tertiary structure in the EC domain (20).

Quantitative analysis of the degree of conservation of a conserved disulfide bond between cysteines in positions equivalent to C110 and C187 in rhodopsin shows that 78.9% of all GPCR sequences including orphan and putative receptor sequences and 87% of characterized GPCRs have this disulfide bond. This confirms earlier suggestions that this disulfide bond is conserved among the majority of GPCR (40). In contrast, the cysteine at position 185 that forms a wrong disulfide bond with C187 in misfolded rhodopsin (see the Introduction) is not conserved (5% of class A GPCR sequences have a cysteine at this position, and the majority of these sequences are closely related to rhodopsin). Combining the high degree of conservation of C110 and C187 with the observed stabilizing role in rhodopsin, it is likely that this disulfide bond couples the EC and TM domains for GPCRs in general.

Other residues that are 80–100% conserved by multiple sequence alignments of the seven-TM helices in GPCRs (41, 42) include N55 in helix I; D83 in helix II; C110 and E134–R135–Y136 in helix III; W161 in helix IV; Y223 in helix V; F261, W265, and P267 in helix VI; and P303 and Y306 in helix VII. The peaks N55, D83, and C110 in Fig. 4 are consistent with the conserved residues in helices I–III, suggesting that the conservation of these residues originates, at least in part, from stability/folding requirements. Functional criteria also play an important role. It is conceivable that the interactions at the global hinge site are finely tuned and conserved to comply with the mechanical constraints imposed by the collective dynamics of the protein. Residues A80 and A299 are reported to be 60–80% conserved (42). Other peaks observed in Fig. 4 include residues not expected to be generally conserved in the GPCR family, namely those that coordinate RET. This includes the Schiff base counterion E113 and its close neighbors in helix III (T97–S98, L112, G114, F116–A117, and G120) and the second EC loop, including

Table 2. Computational predictions

	Method		
	FIRST	GNM	Both
No. of key residues theoretically identified	45	16	52
Fraction also studied by experiments	34/45	14/16	39/52
Fraction correctly observed to cause misfolding	31/34	11/14	35/39
Fraction of them that cause complete misfolding	26/31	5/11	26/35

Table 3. GPCR disulfide bond conservation

Receptors		C110—C187 disulfide bond	
Classes	Sequences	% present	% absent
A	1,972	92.4	7.6
B	163	97.6	2.4
C	75	91.5	8.5
D	24	87.5	12.5
E	5	100.0	0.0
Frizzled/smoothened	68	97.1	2.9
Putative	317	17.0	83.0
Orphans	338	41.4	58.6
Total	2,962	78.9	21.1

the fourth EC β-strand (P180, M183–Q184–C185–S186–C187), and Y102–F103 in the first EC loop. The tight packing near the RET binding site is presumably instrumental for the efficient transmission of light-induced conformational changes to the CP surface.

The CP ends of the helices are found by FIRST analysis to be composed of small, independently rigid groups, some of which are present in the transition state and form tertiary contacts that persist even at later stages of unfolding. Two examples of such tertiary contacts shown in pink and blue in Fig. 5*A* contain residues known to be highly conserved in GPCR class A, including the D(E)RY motif residues E134–R135–Y136. These rigid clusters are separated by flexible regions from the largest rigid cluster, suggesting that the CP domain retains some flexibility despite being structured into small, rigid elements. This type of flexibility may be functionally required to undergo the conformational changes that are recognized by other molecules in the signaling cascade, such as the G protein (43).

Relation to Folding of Membrane Proteins. The FIRST simulated unfolding procedure is a method to identify likely unfolding pathways. Most importantly, the core of structural stability identified as the last persisting cluster of mutually rigid bonds has been shown to correlate extraordinarily well with experimentally determined folding cores in a large number of soluble proteins (26, 27). The experimental determination of folding pathways in membrane proteins has lagged behind that of soluble proteins because the presence of lipids or detergents complicate both the experimental setup and the interpretation of folding studies of membrane proteins (44, 45). Additionally, the higher stability of TM helices compared with soluble helices often prevents their denaturation (46). Only two proteins have been refolded from a completely unfolded state: bacteriorhodopsin (47), which also contains seven-TM helices, and OmpA (48), a β-barrel protein. Based on the finding that native bacteriorhodopsin can be refolded *in vitro* from fragments corresponding to individual α-helices, a two-stage hypothesis has been proposed that states that membrane proteins fold by first forming independently stable α-helices, followed by association of these helices to form the native structure (49).

Whereas the simulated unfolding of rhodopsin does not directly determine a folding pathway, the high level of correlation between the predicted core residues and the experimental misfolding mutations suggests that the thermal denaturation profile presented in Fig. 2 provides insights about a highly

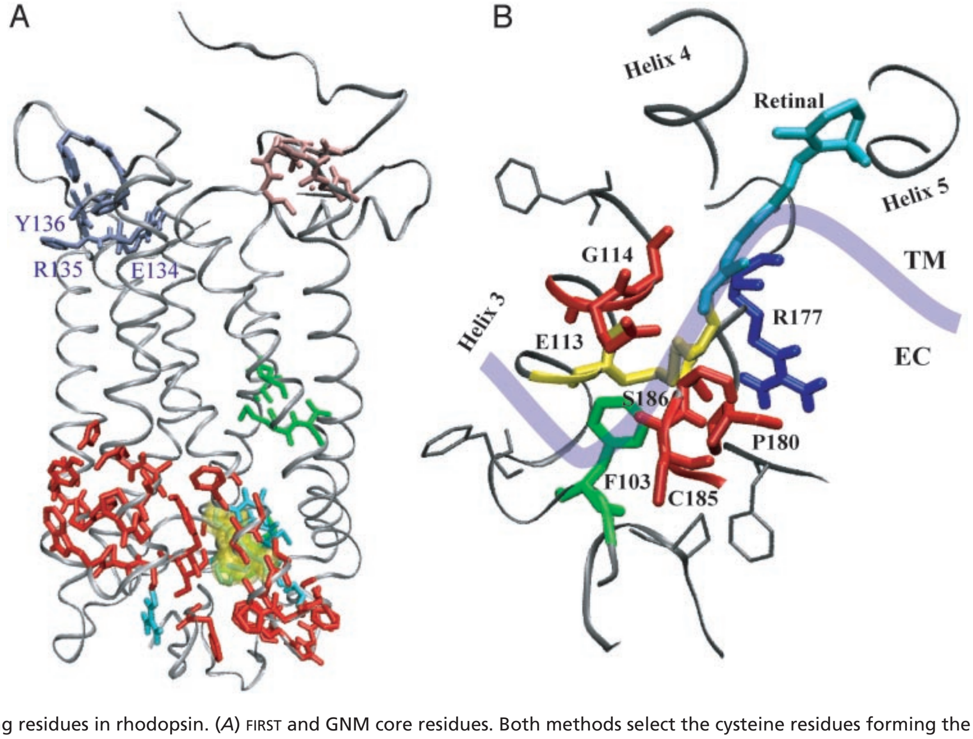

Fig. 5. Critical folding residues in rhodopsin. (*A*) FIRST and GNM core residues. Both methods select the cysteine residues forming the critical disulfide bond, shown by space-filling yellow spheres. The other seven residues found in common by these methods are shown in cyan. The remaining seven GNM fast mode peak residues are shown in green, and the remaining FIRST core residues from line C in Fig. 2 are shown in red. The two next largest rigid clusters, from FIRST analysis, are shown in violet and pink to lie at the CP end of the TM region. (*B*) Local neighborhood of the most stable residues. The nine residues in common between FIRST and GNM fast mode peaks (cyan and yellow in *A*) are shown by thick, colored sticks. At the center of this cluster are the cysteine residues (C110 and C187 in yellow) that form the critical disulfide bond. F103 (green), five of the common residues (E113, G114, P180, C185, and S186 in red), and the RET chromophore (Retinal, cyan) are located within 4 Å of this disulfide bond and span the TM–EC interface (suggested by the thick blue curve). The other residue in common between these methods, R177 (blue), demonstrates how this local stability is propagated across side-chain interactions. Side chains for a few of the 45 FIRST folding core residues are shown by thin sticks to orient the reader.

probable potential folding pathway. The early stages of this pathway (lowest lines) are proposed to involve the C110—C187 disulfide bond and the mutually rigid residues predicted by FIRST as the folding core. This folding core lies at the EC–TM domain interface and substantiates the hypothesis that the TM and EC domains might be structurally coupled. Retracing the pathway backwards in Fig. 2, one can see that rigidity of the folding core (C) spreads to envelop most of the EC end of the TM helices (B) followed by the rest of these helices. Thus, our model emphasizes the contribution of loops, not captured by the two-stage model, to be important for the folding of rhodopsin. The two-stage hypothesis was proposed based on studies on bacteriorhodopsin, but, in contrast to bacteriorhodopsin, rhodopsin cannot be fully denatured. Furthermore, rhodopsin helix fragments do not associate *in vitro*, and only a subset of fragments do so *in vivo* (50, 51). However, the importance of loops for folding has been observed recently even for bacteriorhodopsin (52, 53). Thus, much remains to be discovered about the underlying principles governing the folding of helical membrane proteins, and further experimental studies are needed.

This work was supported by National Institute of General Medical Sciences National Programs of Excellence in Biomedical Computing Planning Grant 065805-01A1, National Science Foundation Information Technology Research Grant 0225636, the Humboldt Foundation/Zukunftsinvestitionsprogramm der Bundesregierung Deutschland, and National Eye Institute (National Institutes of Health) Grant EY011716.

1. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., *et al*. (2000) *Science* **289,** 739–745.
2. Gether, U. (2000) *Endocr. Rev.* **21,** 90–113.
3. Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F. E. & Vriend, G. (2003) *Nucleic Acids Res.* **31,** 294–297.
4. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002) *Nucleic Acids Res.* **30,** 276–280.
5. Berson, E. L. (1993) *Invest. Ophthalmol. Visual Sci.* **34,** 1659–1676.
6. Dryja, T. P., Berson, E. L., Rao, V. R. & Oprian, D. D. (1993) *Nat. Genet.* **4,** 280–283.
7. Macke, J. P., Davenport, C. M., Jacobson, S. G., Hennessey, J. C., Gonzalez-Fernandez, F., Conway, B. P., Heckenlively, J., Palmer, R., Maumenee, I. H., Sieving, P., *et al*. (1993) *Am. J. Hum. Genet.* **53,** 80–89.
8. Inglehearn, C. F., Keen, T. J., Bashir, R., Jay, M., Fitzke, F., Bird, A. C., Crombie, A. & Bhattacharya, S. (1992) *Hum. Mol. Genet.* **1,** 41–45.
9. Sung, C. H., Davenport, C. M. & Nathans, J. (1993) *J. Biol. Chem.* **268,** 26645–46649.
10. Anukanth, A. & Khorana, H. G. (1994) *J. Biol. Chem.* **269,** 19738–19744.
11. Doi, T., Molday, R. S. & Khorana, H. G. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4991–4995.
12. Kaushal, S. & Khorana, H. G. (1994) *Biochemistry* **33,** 6121–6128.
13. Liu, X., Garriga, P. & Khorana, H. G. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 4554–4559.
14. Ridge, K. D., Lu, Z., Liu, X. & Khorana, H. G. (1995) *Biochemistry* **34,** 3261–3267.
15. Hwa, J., Reeves, P. J., Klein-Seetharaman, J., Davidson, F. & Khorana, H. G. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 1932–1935.
16. Garriga, P., Liu, X. & Khorana, H. G. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 4560–4564.
17. Hwa, J., Garriga, P., Liu, X. & Khorana, H. G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10571–10576.
18. Hwa, J., Klein-Seetharaman, J. & Khorana, H. G. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4872–4876.
19. Khorana, H. G. (2000) *J. Biomol. Struct. Dyn.* **11,** 1–16.
20. Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E. M. & Shichida, Y. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 5982–5987.
21. Jacobs, D. J., Rader, A. J., Kuhn, L. A. & Thorpe, M. F. (2001) *Proteins* **44,** 150–165.
22. Bahar, I., Atilgan, A. R. & Erman, B. (1997) *Folding Des.* **2,** 173–181.
23. Haliloglu, T., Bahar, I. & Erman, B. (1997) *Phys. Rev. Lett.* **79,** 3090–3093.
24. Bahar, I., Atilgan, A. R., Demirel, M. C. & Erman, B. (1998) *Phys. Rev. Lett.* **80,** 2733–2736.
25. Demirel, M. C., Atilgan, A. R., Jernigan, R. L., Erman, B. & Bahar, I. (1998) *Protein Sci.* **7,** 2522–2532.
26. Hespenheide, B. M., Rader, A. J., Thorpe, M. F. & Kuhn, L. A. (2002) *J. Mol. Graphics Model.* **21,** 195–207.
27. Rader, A. J. & Bahar, I. (2004) *Polymer* **45,** 659–668.
28. Williams, M. A., Goodfellow, J. M. & Thornton, J. M. (1994) *Protein Sci.* **3,** 1224–1235.
29. Vriend, G. (1990) *J. Mol. Graphics* **8,** 52–56.
30. Bahar, I. (1999) *Rev. Chem. Eng.* **15,** 319–349
31. Campagne, F., Jestin, R., Reversat, J. L., Bernassau, J. M. & Maigret, B. (1999) *J. Comput. Aided Mol. Des.* **13,** 625–643.
32. Rader, A. J., Hespenheide, B. M., Kuhn, L. A. & Thorpe, M. F. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 3540–3545.
33. Sanders, C. R. & Myers, J. K. (2004) *Annu. Rev. Biophys. Biomol. Struct.* **33,** 25–51.
34. Reeves, P. J., Hwa, J. & Khorana, H. G. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 1927–1931.
35. Karnik, S. S., Sakmar, T. P., Chen, H. B. & Khorana, H. G. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8459–8463.
36. Nakayama, T. A. & Khorana, H. G. (1991) *J. Biol. Chem.* **266,** 4269–4275.
37. Kaushal, S., Ridge, K. D. & Khorana, H. G. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 4024–4028.
38. Andres, A., Garriga, P. & Manyosa, J. (2003) *Biochem. Biophys. Res. Commun.* **303,** 294–301.
39. Kono, M., Yu, H. & Oprian, D. D. (1998) *Biochemistry* **37,** 1302–1305.
40. Sakmar, T. P. (2002) *Curr. Opin. Cell. Biol.* **14,** 189–195.
41. Baldwin, J. M., Schertler, G. F. X. & Unger, V. M. (1997) *J. Mol. Biol.* **272,** 144–164.
42. Mirzadegan, T., Benko, G., Filipek, S. & Palczewski, K. (2003) *Biochemistry* **42,** 2759–2767.
43. Klein-Seetharaman, J. (2002) *Chembiochem* **3,** 981–986.
44. White, S. H. & Wimley, W. C. (1999) *Annu. Rev. Biophys. Biomol. Struct.* **28,** 319–365.
45. Brown, M. F. (1994) *Chem. Phys. Lipids* **73,** 159–180.
46. Booth, P. J., Templer, R. H., Curran, A. R. & Allen, S. J. (2001) *Biochem. Soc. Trans.* **29,** 408–413.
47. London, E. & Khorana, H. G. (1982) *J. Biol. Chem.* **257,** 7003–7011.
48. Surrey, T. & Jahnig, F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7457–7461.
49. Popot, J. L. & Engelman, D. M. (1990) *Biochemistry* **29,** 4031–4037.
50. Ridge, K. D., Lee, S. S. & Yao, L. L. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3204–3208.
51. Ridge, K. D. & Abdulaev, N. G. (2000) *Methods Enzymol.* **315,** 59–70.
52. Allen, S. J., Kim, J. M., Khorana, H. G., Lu, H. & Booth, P. J. (2001) *J. Mol. Biol.* **308,** 423–435.
53. Kim, J. M., Booth, P. J., Allen, S. J. & Khorana, H. G. (2001) *J. Mol. Biol.* **308,** 409–422.
54. Humphrey, W., Dalke, A. & Schulten, K. (1996) *J. Mol. Graphics* **14,** 33–38.