Stability of Loops in the Structure of Lactose Permease†

Michael Bennett,‡ James A. Yeagle,‡,§ Mark Maciejewski,|| James Ocampo,‡ and Philip L. Yeagle*,‡

Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269, and Department of Biochemistry, University of Connecticut Health Science Center, Farmington, Connecticut 06030

Received May 18, 2004; Revised Manuscript Received August 9, 2004

ABSTRACT: Structural analysis of peptide fragments has provided useful information on the secondary structure of integral membrane proteins built from a helical bundle (up to seven transmembrane segments). Comparison of those results to recent X-ray crystallographic results showed agreement between the structures of the fragments and the structures of the intact proteins. Lactose permease of *Escherichia coli* (lac Y) offers an opportunity to test that hypothesis on a substantially larger integral membrane protein. Lac Y contains a bundle of 12 transmembrane segments connected by 11 loops. Eleven segments, each corresponding to one of the loops in this protein, were studied. Five of these segments form defined structures in solution as determined by multidimensional nuclear magnetic resonance. Four peptides form turns, and one peptide reveals the end of one of the transmembrane helices. These results suggest that some loops in helical bundles are stabilized by short-range interactions, particularly in smaller bundles, and such intrinsically stable loops may contribute to protein stability and influence the pathway of folding. Greater conformational flexibility may be found in large integral membrane proteins.

Lactose permease (lac Y) of *Escherichia coli* is an integral membrane protein (for a review, see ref *1*). This protein is a symport, responsible for coupling the transport of galactosides with hydrogen ion. A stoichiometry of 1:1 (galactoside:H^+) is observed. The protein functionally and structurally acts as a monomer (*2*).

Lac Y was predicted from hydrophobicity plots to consist of a bundle of 12 transmembrane helices (TM) connected by loops of varying length. This hypothesis has been tested by a variety of biochemical experiments, and the results have supported the model (*3*). Furthermore, these investigations have proposed important details on the boundaries between the transmembrane helices and the connecting loops, providing further insight into the overall structure of this protein (*4*). Recently, an X-ray crystal structure was reported for a mutant lac Y (*5*). The crystal structure shows a bundle of 12 transmembrane helices, with which the previous studies on structure are largely consistent.

Recent experiments on smaller integral membrane proteins, also consisting of helical bundles, have revealed that the transmembrane helices (*6*−*17*) and the turns (*18*−*25*) behave as subdomains of the overall protein. For example, peptide fragments corresponding to the individual helices of bacteriorhodopsin (*26*) and bovine rhodopsin (*27*) are helical in solution, as determined by multidimensional nuclear magnetic resonance (NMR),[1] and are consistent with the corresponding region of the crystal structure. Likewise, peptides corresponding to putative turns in these proteins form turns in solution independent of the remainder of the protein.

These results led to the hypothesis that useful information about secondary structure (helices and turns) can be obtained for membrane proteins from structural studies on fragments. In this report, we test this hypothesis on the large helical bundle of lac Y. We find that peptide fragments from five of the turns of this 12 TM membrane protein behave similarly to peptide fragments from smaller 7 TM membrane proteins and report on the secondary structure of the protein. Since X-ray crystal structures of membrane proteins are still relatively rare and turns are often disordered in the structures, and since turns are difficult to predict with computer modeling, it is important to have alternative means to obtain this structural information. The stability of many loops from membrane proteins suggests that loops could help to guide protein folding. The majority of the turns in lac Y, however, do not form stable structures as separate fragments, a result that differs significantly from previous studies on smaller helical bundles. These latter observations suggest that greater conformational flexibility may be found in large integral membrane proteins, or at least those involved in active transport.

MATERIALS AND METHODS

Peptide Synthesis. The following peptides were synthesized by solid-phase synthesis at Genemed (San Francisco, CA): MGAYFPFFPIWLHDINHISKSDTGIIFAAISL, loop 1−2, 32mer, lpt1−2; QPLFGLLSDKLGLRKYLLWIITG, loop 2−3, 23mer, lpt2−3; FIFIFGPLLQYNILVGSIVGGIYLGF, loop 3−4, 26mer, lpt3−4; VEAFIEKVSRRS-

† This work was supported by a grant from the NIH to P.L.Y. (GM65250).

* Correspondence should be addressed to this author. E-mail: yeagle@uconnvm.uconn.edu. Phone: 860-486-5154. Fax: 860-486-4331.

‡ University of Connecticut.

§ Present address: St. Olaf College, Northfield, MN 55057.

|| University of Connecticut Health Science Center.

[1] Abbreviations: CD, circular dichroism; DPC, dodecylphosphocholine; NMR, nuclear magnetic resonance; rmsd, root mean square deviation; SDS, sodium dodecyl sulfate.

10.1021/bi049000s CCC: $27.50
Published on Web 09/18/2004

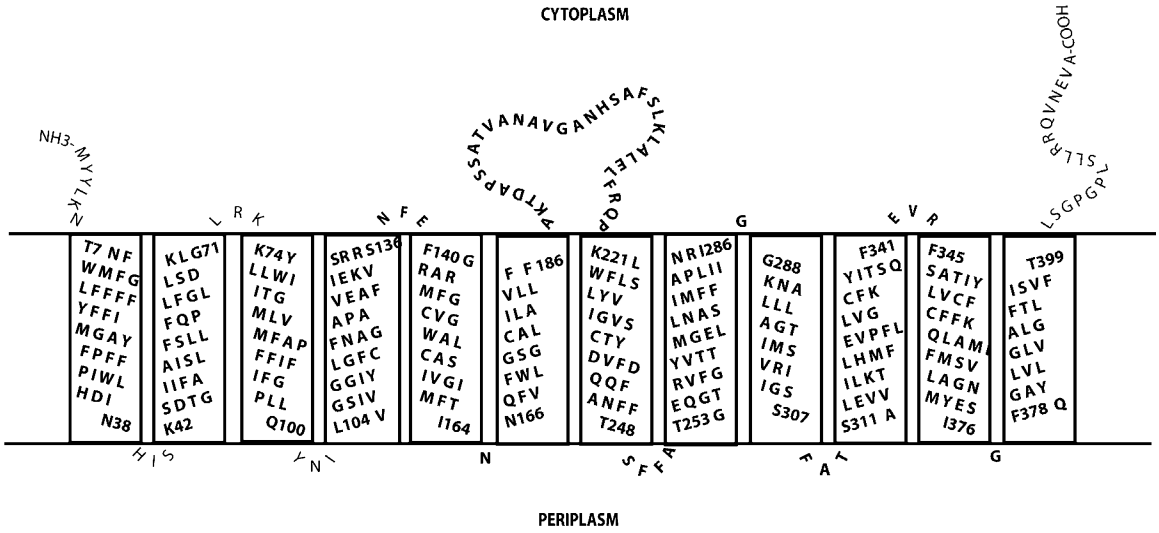

FIGURE 1: Schematic representation of the primary structure of lac Y. The peptide sequences examined in this paper can be identified in this schematic as encompassing each of the loop regions. Figure adapted from ref *1*. The residues indicated as located in the loops have been so identified by direct analysis of the X-ray crystal structure of the protein, and this figure differs in this manner from the original figure.

NFEFGRARMFGCVG, loop 4−5, 26mer, lpt4−5; IVIMFTINNQFVFWLGSG, loop 5−6, 18mer, lpt5−6; LAVLLFFAKTDAPSSATVANAVGANHSAFSLKLALELFRQPKL, loop 6−7, 43mer, lpt6−7; SCTYDVFDQQFANFFTSFFATGEQGTRVFGY, loop 7−8, 31mer, lpt7−8; SIMFFAPLIINRIGGKNALLLAGTI, loop 8−9, 25mer, lpt8−9; AGTIMSVRIIGSSFATSALEVVILKTLHM, loop 9−10, 29mer, lpt9−10; VGCFKYITSQFEVRFSATIYLV, loop 10−11, 22mer, lpt10−11; SVLAGNMYESIGFQGAYLVLGL, loop 11−12, 22mer, lpt11−12. The peptides were designed prior to the publication of the X-ray crystal structure on the basis of the previous biochemical experiments. The sequences were selected to be roughly centered on the turns. The underlined residues are those actually found in the turns from the crystal structure. The original design therefore succeeded in locating the turn roughly in the middle of the peptide. Other design considerations included overall length; previous experiments revealed that if the peptide was too short, it might be unstructured. Stable turns have been seen in peptides as short as 16 residues, but since longer peptides usually produced better results, all but one of these lac Y peptides was designed with 22 or more residues. Another factor is that too many hydrophobic residues cause problems in synthesis and solubility, which limited the total length of the peptides that could be used. Finally, for these loop peptides, it has been found to be helpful to include enough residues on either side of the actual turns to potentially form one or two turns of the attached helices. Where successful, this can lead to a helix−turn−helix motif in the experimental structure, which provides considerable information on not only residues in a turn but also the ends of the transmembrane helices.

These peptides were analyzed by mass spectrometry and showed the correct mass for the sequence and a purity greater than 90%, which is adequate for the NMR measurements. Figure 1 shows the primary structure of lac Y in which these sequences can be located. These peptides correspond to the sequences for all of the loops of lac Y.

NMR Spectroscopy. All NMR spectra were recorded on a Varian INOVA-600 spectrometer at 30 °C in neat DMSO. DMSO was used because these peptides are poorly soluble, or insoluble, in water, DMSO does not preferentially stabilize any particular secondary structure, and structures of fragments in DMSO from other integral membrane proteins agreed well with corresponding regions of the X-ray crystal structure (*26*, *28*). Standard pulse sequences and phase cycling were employed to record double quantum filtered (DQF) COSY and NOESY [data were collected with 400 ms mixing times (*29*)]. Previous work with other similarly sized peptides at mixing times of 150, 250, and 400 ms showed no evidence of spin diffusion, and 400 ms showed the most useful interactions in the NOESY. All spectra were accumulated in a phase-sensitive manner using time-proportional phase incrementation for quadrature detection in F_1. Chemical shifts were referenced to the residual protons in the DMSO-d_6. Sequence-specific assignments were obtained using standard approaches, and the results are listed in Tables 1−5.

Structure Refinement. The sequence-specific assignment of the ^{1}H NMR spectrum for each peptide was carried out using standard methods employing FELIX (MSI, Inc.). Assigned NOE cross-peaks were segmented using a statistical segmentation function and characterized as strong, medium, and weak, corresponding to upper bound distance range constraints of 2.7, 3.5, and 5.0 Å, respectively. Lower bounds between nonbonded atoms were set to the sum of their van der Waals radii (approximately 1.8 Å). Pseudoatom corrections were added to interproton distance restraints where necessary (*30*). Distance geometry calculations were carried out using the program DIANA (*31*) within the SYBYL 6.6 package (Tripos Software Inc., St. Louis, MO). First-generation DIANA structures, 150 in total, were calculated. Energy refinement calculations (restrained minimizations/dynamics) were carried out on the best distance geometry structures using the SYBYL program implementing the Kollman all-atom force field. Structures were also obtained using simulated annealing on the peptides with the distance constraints obtained from the NOESY data. Simulated annealing was done with the Kollman all-atom force field and Kollman charges within Sybyl. The molecule was heated to 1000 K for 1000 fs followed by cooling to 200 K during

Table 1: Chemical Shift Table for lpt45

residue	αH	NH	βH	other
Val1	3.58		2	2.21, 2.25
Glu2	4.32	8.46	1.7, 1.83	
Ala3	4.25	8.03	1.12	
Phe4	4.56	8.07	2.76, 2.96	
Ile5	4.16	7.91	1.66	0.78, 1.04
Glu6	4.22	8.04	1.81, 1.91	2.14
Lys7	4.288	8.02	1.59, 1.63	1.24, 1.29, 1.46, 2.7
Val8	4.21	7.74	1.91	0.78, 0.8
Ser9	4.3	8.06	3.53	5
Arg10	4.18	8.02	1.43, 1.61	3.04, 7.43
Arg11	4.29	7.97	1.45, 1.64	3.06, 7.38
Ser12	4.28	7.95	3.47, 3.53	5.07
Asn13	4.51	8.16	2.32, 2.49	
Phe14	4.42	7.97	2.69, 2.91	
Glu15	4.19	8.06	1.66, 1.81	2.13
Phe16	4.48	7.93	2.78, 3.01	
Gly17	3.71, 3.75	8.25		
Arg18	4.3	7.96	1.46, 1.65	3.05
Ala19	4.22	8.08	1.18	
Arg20	4.17	8.029	1.45, 1.6	3.05
Met21	4.28	7.8	1.69, 1.83	2.32
Phe22	4.5	8.04	2.76, 3.01	
Gly23	3.66, 3.81	8.27		
Cys24	4.3	8	2.7, 3.05	
Val25	4.177	7.9	1.96	0.83
Gly26	3.69	8.31		

Table 2: Chemical Shift Table for lpt56

residue	αH	NH	βH	other
Ile1			1.75	
Val2	4.24	8.29	1.87	
Ile3	4.13	8.09	1.69	1.05, 1.36
Met4	4.32	7.93	1.67, 1.80	2.33
Phe5	4.67	7.93	2.79, 3.03	
Thr6	4.30	8.02	1.01, 4.00	
Ile7	4.22	7.67	1.72	1.08, 1.41
Asn8	4.53	8.13	2.40, 2.58	7.37
Asn9	4.48	7.99	2.42	
Gln10	4.09	7.98	1.67, 1.83	2.02
Phe11	4.46	7.95	2.81, 2.95	
Val12	4.07	7.58	1.89	0.71
Phe13	4.52	7.92	2.77, 2.95	
Trp14	4.55	8.03	2.96, 3.12	7.12, 7.55
Leu15	4.30	7.99	1.44	0.81
Gly16	3.74	7.90		
Ser17	4.34	7.87	3.59	
Gly18	3.74	8.14		

1500 fs. Ten consecutive cycles were calculated. These calculations were performed on a Silicon Graphics R10000 computer.

Structural Analysis. The structures of the loop peptides were compared with the native protein structure, since after this project was largely complete the crystal structure of one form of lac permease was reported (*5*). VMD (*32*) was used to overlay the structures of the loops on the crystal structure and to calculate the rmsd of the fit using the backbone atoms.

Circular Dichroism (CD). CD spectra were obtained on a Jasco J715 spectropolarimeter at 45 °C (to enhance the optical clarity of the samples) in 2 mm path length cuvettes and corrected by subtracting a detergent blank. Attempts to deconvolute the spectra were not successful because the basis set for the deconvolution programs is from proteins, not from peptides. Each CD sample was prepared using 1.0 mg of the peptide in 700 μL of a mixed detergent solution containing 136 mM lauryl sulfate/45 mM octyl glucoside in 10 mM potassium phosphate buffer, pH 7.0.

Table 3: Chemical Shift Table for lpt89

residue	αH	NH	βH	other
Ser1	4.21	8.18		
Ile2	4.32	8.73	1.76	0.83
Met3	4.4	8.2	1.8, 1.9	2.44
Phe4	4.62	8.02	2.85, 3.07	7.3
Phe5	4.67	8.28	2.87, 3.09	7.23, 7.34
Ala6	4.63	8.39	1.31	
Pro7	4.45		1.91, 1.97	2.1, 3.6
Leu8	4.36	8.09	1.55, 1.62	1.7
Ile9	4.29	7.83	1.82	0.45, 0.89, 1.17
Ile10	4.26	7.93	1.8	0.53, 0.9, 1.18
Asn11	4.66	8.24	2.53, 2.64	7.08, 7.56
Arg12	4.36	8.12	1.59, 1.67	1.57, 1.82, 3.18, 7.57
Ile13	4.26	7.92	1.87	0.54, 0.95, 1.17
Gly14	3.85	8.24		
Gly15	3.9	8.18		
Lys16	4.39	8.2	1.61, 1.75	1.4, 2.86
Asn17	4.615	8.34	2.54, 2.68	2.54, 2.68
Ala18	4.26	8.16	1.31	
Leu19	4.32	8.09	1.55, 1.62	0.93, 0.99
Leu20	4.37	7.82	1.58, 1.67	0.93, 0.99
Leu21	4.38	7.86	1.56, 1.7	0.93, 0.99
Ala22	4.34	8.01	1.33	
Gly23	3.82, 3.94	8.23		
Thr24	4.41	7.78	4.04	1.15
Ile25	4.3	7.99	1.89	0.51, 0.97, 1.28

Table 4: Chemical Shift Table for lpt10−11

residue	αH	NH	βH	other
Val1	3.73	8.17	2.16	
Gly2	3.84	8.71	4.04	
Cys3	4.52	8.21	2.73, 2.81	2.38
Phe4	4.61	8.29	2.85, 3.07	7.32
Lys5	4.34	8.22	1.61, 1.71	1.35, 1.57, 2.83, 7.78
Tyr6	4.63	8.01	2.78, 3.01	7.10
Ile7	4.39	8.15	1.86	0.93, 1.18
Thr8	4.42	7.93	4.12	1.14, 5.10
Ser9	4.43	7.88	3.64, 3.70	5.14
Gln10	4.26	8.17	1.75, 1.93	2.14
Phe11	4.59	8.08	2.86, 3.10	7.32
Glu12	4.42	8.23	1.85, 2.00	2.35
Val13	4.28	7.80	2.01	0.89
Arg14	4.38	8.12	1.53, 1.70	1.11, 3.15, 7.50
Phe15	4.72	8.03	2.88, 3.14	7.30
Ser16	4.42	8.28	3.68	
Ala17	4.50	8.17	1.34	
Thr18	4.35	7.97	4.04	1.09, 5.04
Ile19	4.28	7.69	1.79	0.82, 1.35
Tyr20	4.56	8.00	2.76, 2.96	7.09
Leu21	4.50	8.07	1.54, 1.70	0.93
Val22	4.22	7.93	2.15	0.99

RESULTS

Five of the peptides corresponding to loops of lac Y form stable structures in solution. Structures were determined for each of these peptides from NMR as described in Materials and Methods. The details for each of these five peptides are described in the following.

Lpt3−4 corresponds to residues 91−116 of lac Y. A total of 105 intraresidue, 41 sequential, and 18 long-range constraints were observed in the spectrum. The density of constraints is presented in Figure 2. The pattern of constraints (Figure 3A) shows long-range constraints in the middle and few in the extremes of the sequence. The overlay of the five

Table 5: Chemical Shift Table for lpt34

residue	αH	NH	βH	other
Phe1	4.17		2.91, 3.01	7.21
Ile2	4.36	8.53	1.52, 1.78	0.89, 1.11
Phe3	4.72	8.35	2.85, 2.99	
Ile4	4.70	7.97	1.76	0.82, 1.12, 1.44
Phe5	4.74	8.16	2.90, 3.14	
Gly6	3.84, 3.94	8.24		
Pro7	4.44		1.90, 1.96	2.11, 3.54, 3.58
Leu8	4.33	8.12	1.56, 1.72	
Leu9	4.38	7.81	1.53, 1.65	
Gln10	4.28	7.95	1.92, 2.13	
Tyr11	4.54	7.93	2.75, 2.96	
Asn12	4.71	8.40	2.50, 2.64	
Ile13	4.29	7.88	1.87	0.91, 1.19, 1.51
Leu14	4.44	8.16	1.54	
Val15	4.21	7.73	2.06	0.92
Gly16	3.82, 3.89	8.21		
Ser17	4.52	7.96	3.64	
Ile18	4.32	8.04	1.87	0.91
Val19	4.24	7.84	2.07	0.94
Gly20	3.67, 3.78	7.97		
Gly21	3.79, 3.88	8.09		
Ile22		7.91	1.75	0.83, 1.08, 1.40
Tyr23	4.54	8.17	2.75, 2.96	
Leu24	4.33	7.81	1.54	
Gly25	3.67, 3.88	7.93	3.68	
Phe26		8.03	2.89, 3.12	

lowest energy DIANA structures is shown in Figure 4. The middle of the peptide is structured. The average pairwise backbone rmsd of these structures for residues 6−16 was 0.44. The total energy after minimization was 283 kcal/mol. All residues in the structured region lie within the acceptable regions of the Ramachandran plot. The sequence forms a turn between residues 99−106 of lac permease, while most residues outside this turn are disordered due to the lack of NOE constraints.

Lpt4−5 corresponds to residues 125−150, putatively a loop connecting helices 4 and 5 of lac permease. A total of 136 intraresidue, 30 sequential, and 14 long-range interactions were assigned to produce the final structure. The density of constraints is reduced in the middle where residues R10 and R11 could not be unambiguously assigned, and few long-range constraints could be found in the N-terminal half (Figure 2). The pattern of constraints (Figure 3B) suggests helicity at the C-terminal end of the peptide. The chemical shift index (*33*) also suggests helical tendency at the C-terminal end of the peptide (data not shown). The overlay of the five lowest energy DIANA structures is shown in Figure 4. The structure was minimized using the Powell method in Sybyl6.6. The rmsd for the backbone from the overlay of residues 16−26 is 0.45. The total energy after minimization was 295 kcal/mol. All residues except Phe14 of the peptide lie within the acceptable regions of the Ramachandran plot. The final structure for lpt4−5 is disordered from the N-terminal end until residue 136 of lac Y (residue 12 of the peptide) and then helical from F140 through the end of the peptide (G150).

Lpt5−6 corresponds to residues 157−175 of the extramembraneous segment 5 of lac permease. A total of 93 intraresidue, 49 sequential, and 31 long-range interactions were assigned to produce the final structure. The density of constraints (Figure 2) is fairly uniform over the whole peptide. The pattern of constraints shows some helical tendency outside the turn region (Figure 3C). The overlay of the last five simulated annealing cycles is shown in Figure 4. The average pairwise rmsd of the backbone for the entire molecule is 0.86. The total energy after minimization was 224 kcal/mol. All residues except Ile7 and Phe11 of the peptide lie inside acceptable regions of the Ramachandran plot. The structure forms a turn between residues T163 and F168 of lac permease, leaving N164, N165, and Q166 (residues 8−10 of the peptide) at the apex of the turn.

Lpt8−9 corresponds to residues 274−298 of lac permease, a putative extramembraneous loop. A total of 198 intraresidue, 79 sequential, and 17 long-range interactions were assigned to produce the final structure. Long-range constraints are primarily located in the region of the turn with a low density of constraints in the amino terminus and carboxyl terminus (Figures 2 and 3D). The overlay of the five lowest energy DIANA structures is shown in Figure 4. The average pairwise rmsd of the backbone for residues 7−12 is 0.07. The total energy after minimization was 295 kcal/mol. All residues except Leu8 and Met9 in the structured region lie within acceptable regions of the Ramachandran plot. The sequence forms a turn between residues 280 and 286 of lac permease, with most residues outside this turn being disordered due to the lack of NOE constraints.

Lpt10−11 corresponds to residues 331−352 of lac Y. A total of 117 intraresidue, 49 sequential, and 15 long-range interactions were observed in the spectrum. The distribution of long-range constraints is shown in Figure 2. The connectivity plot of the constraints shows that the peptide is structured only between residues 6−13 (Figure 3E). The overlay of the five lowest energy DIANA structures is shown in Figure 4. The average pairwise backbone rmsd of these structures for residues 6−13 was 1.87. The total energy after minimization was 64 kcal/mol. All residues in the structured region are within the acceptable regions of the Ramachandran plot. The sequence forms a turn between residues 336−343 of lac permease. Most residues outside this turn are disordered due to the lack of NOE constraints. Superposition of this peptide structure on the X-ray crystal structure produces an rmsd of 2.9. This is substantially higher than the other peptides, as is the rmsd among the family of structures for this sequence, and may be the result of too few distance constraints from the NOESY data for this peptide.

Of the remaining peptides, none showed enough long-range NOE contacts to define structure in any region of the peptide. Therefore, these loop peptides have been designated as disordered.

After most of this structural work was complete, the X-ray crystal structure of the protein was reported. This enables a comparison of the loop structures from the fragments with the structures in the native protein by a superposition of the ordered regions of the five loop fragments, lpt3−4, lpt4−5, lpt5−6, lpt8−9, and lpt10−11, on the crystal structure of lac permease. Table 6 lists the rmsd of these overlays as well as those of other loops from smaller helical bundles.

CD spectra were obtained of each of the peptides in a mixed detergent micelle system to further explore structural features of these loop peptides. The results are shown in Figure 5. The turns which yielded stable structures in the NMR measurements in DMSO, lpt5−6, lpt8−9, and lpt10−11, showed class B CD spectra characteristic of type II β-turn structures, with minima at 218−220 nm and maxima at 200 nm, and the CD spectrum of lpt3−4 showed contributions

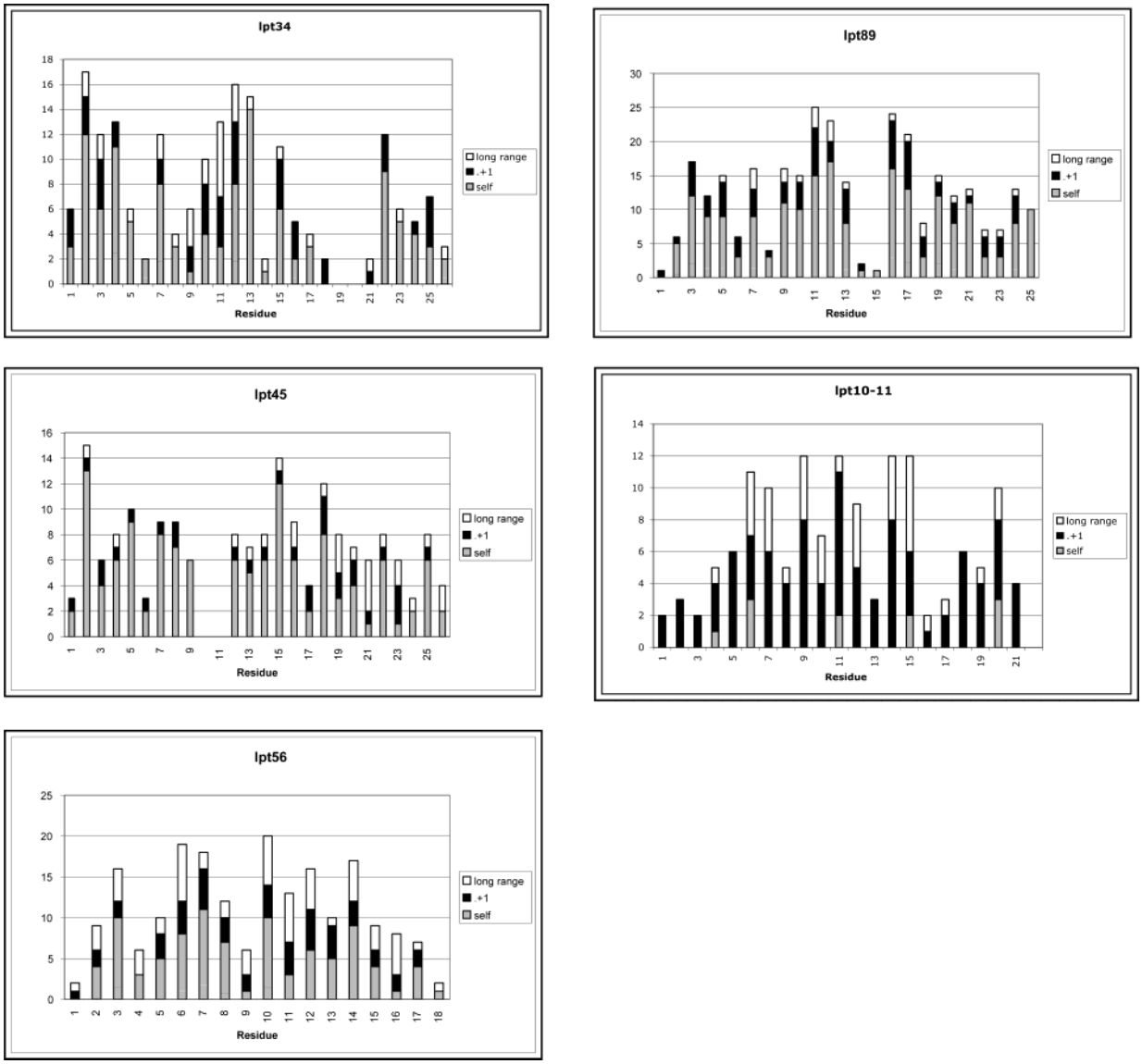

FIGURE 2: Constraint plots showing the number of intraresidue, sequential, and long-range interactions obtained from the NOESY experiments for each residue in each peptide. Shown in hatched boxes are the intraresidue interactions, above which are the sequential interactions in black and the long-range interactions in clear.

Table 6: rmsd Values for the Overlay of the Loop NMR Structures with the Crystal Structure

loop	residues	rmsd
rhodopsin first cytoplasmic loop[a]	60−70	1.0
rhodopsin second cytoplasmic loop	138−150[a]	2.4
lac Y loop 3−4	100−105	1.9
lac Y loop 4−5	140−150	1.4
lac Y loop 5−6	163−168	1.4
lac Y loop 8−9	280−286	2.1
lac Y loop 10−11	337−343	2.9

[a] Reference *21*.

from both β-turns and α-helix. The turns that were unstructured in DMSO in the NMR measurements, with the exception of lpt11−12, exhibited CD spectra characteristic of α-helix with minima at 208 and 222 nm. CD spectra cannot be obtained in DMSO because of interference from absorbance of the solvent. We were unable to obtain usable NMR data from the samples containing detergent. Samples made with just octyl β-glucoside scattered too much light for CD measurements.

DISCUSSION

Lac Y is a bundle of 12 transmembrane helices. We have examined the hypothesis in the introduction by determining the structures in solution of fragments of the protein that correspond to turns of lac Y. Peptides corresponding to all 11 of the potential helix−turn−helix motifs in lac Y were examined. Some stable structure was found in five of these loop peptides: four peptides formed a turn in solution, and the fifth formed a helix on one end while the remainder was disordered.

Values of 1.4−2.1 (with one exception) are obtained for the rmsd of the superposition of these peptide structures on the X-ray crystal structure of lac Y. These values compare favorably with previous comparisons of superpositions of fragment structures on native crystal structures for rhodopsin, as can be seen in the two entries for rhodopsin in the table. The rhodopsin fragments correspond to the first and second cytoplasmic loops; part of the third cytoplasmic loop is disordered in the crystal structure of rhodopsin so that this comparison cannot be made for that loop. Previous studies on bacteriorhodopsin from *Halobacterium salinarium* showed

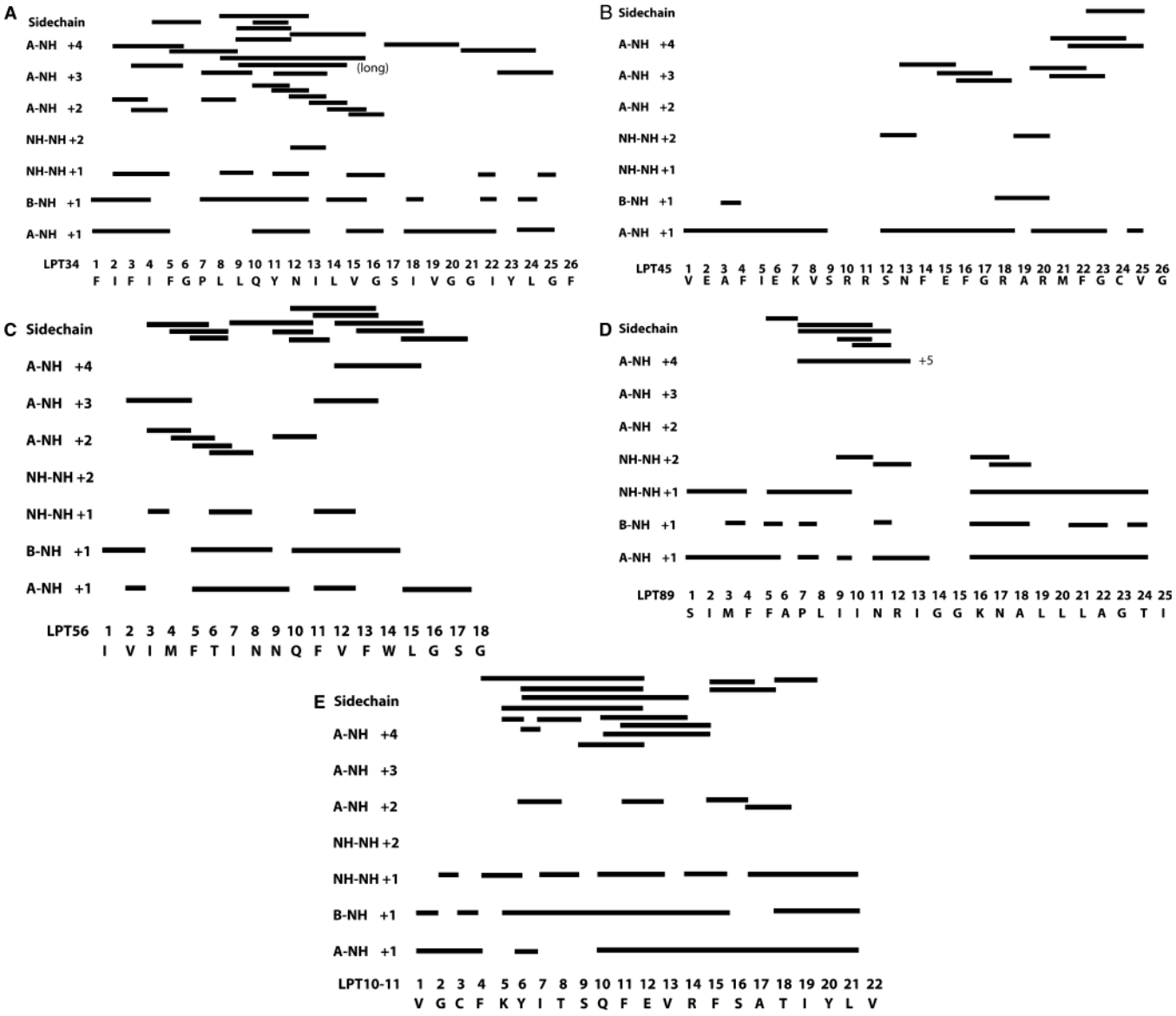

FIGURE 3: Connectivity plots showing the interactions observed in the NOE constraints obtained for each peptide: (A) lpt34 connectivities; (B) lpt45 connectivities; (C) lpt56 connectivities; (D) lpt89 connectivities; (E) lpt10−11 connectivities.

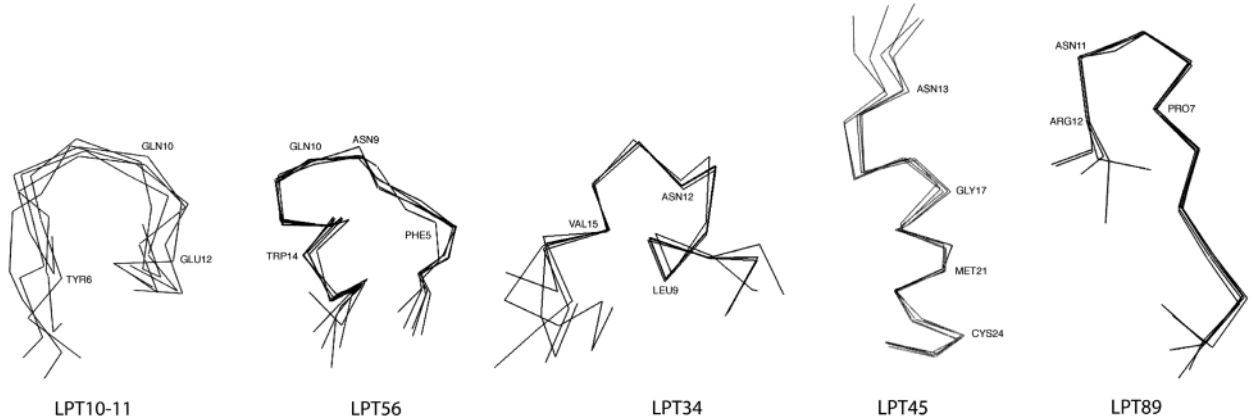

FIGURE 4: Overlay of the family of solution structures. (A) Overlay of residues 6−13 of the five lowest energy lpt10−11 DIANA structures. (B) Overlay of the last 5 of 20 cycles of simulated annealing of lpt56. (C) Overlay of residues 6−16 of the five lowest energy lpt34 DIANA structures. (D) Overlay of residues 16−26 of the five lowest energy lpt45 DIANA structures. (E) Overlay of residues 7−12 of the five lowest energy lpt89 DIANA structures.

rmsd's for superposition of the loops onto the crystal structure of 2.0−2.5 (*26*), consistent with the results for lac Y and rhodopsin.

Studies on loops from other membrane proteins have also shown stable structures for peptides encompassing the sequences of the turns. These include the second cytoplasmic

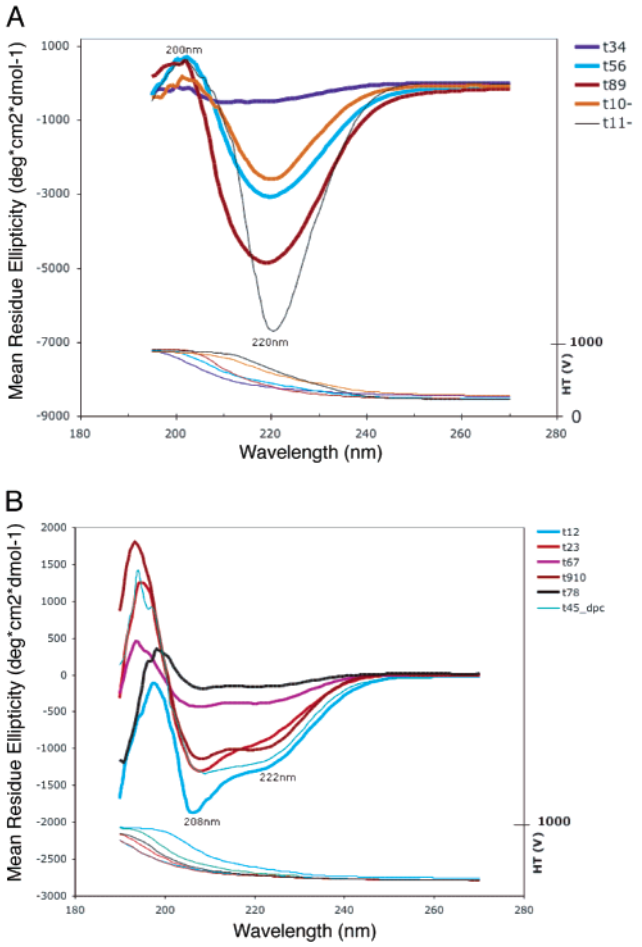

FIGURE 5: CD spectra of the loop peptides in SDS/OG micelles. (A) The CD from turns stable in DMSO, including lpt 3−4, lpt5−6, lpt8−9, and lpt10−11. (B) The CD of the remaining turn peptides which did not form stable structures in DMSO. The voltage on the photomultiplier is shown at the bottom of each panel.

loop of the α2A adrenergic receptor (*34*), the third intracellular loop of the human cannabinoid 1 receptor (*35*), a loop from the anion transporter of erythrocytes (*36*), the first extracellular loop (*37*) and the third cytoplasmic loop of the rat angiotensin II AT1A receptor (*18*), a loop from the turkey erythrocyte adrenoreceptor (*23*), the third cytoplasmic loop (*20*) and the first extracellular loop of the PTH receptor (*17*), loops of EmrE (*38*), and the third extracellular loop of the neurokinin 1 receptor (*39*).

These results collectively from studies on 34 loops of integral membrane proteins (consisting of helical bundles) provide considerable information on the residues that form the turn and, in favorable cases, the end points of transmembrane helices. While predictions have proven quite successful in locating the transmembrane helices within the primary sequence of a membrane protein (for example, ref *40*), predictions have been less successful in predicting the above kinds of information about turns. Therefore, structural studies on peptides spanning the turn regions of membrane proteins provide useful structural information on the membrane proteins from which the peptide fragments were derived when no high-resolution structures are available.

Several factors have been examined in an attempt to understand what stabilizes turns from membrane proteins. Jones et al. suggested from an analysis of sequences of membrane proteins that there is a higher incidence of a particular subset of amino acids in turns (*41*). Thus amino acids such as R, K, H, Q, and E occurred more frequently in turns on the portion of the protein facing the cell cytoplasm than in transmembrane helices where they were quite rare. Likewise, for turns on the opposite face of the protein, E, D, Q, H, N, and P were particularly prevalent. The turns of lac Y do not adhere to this pattern uniformly. One might then hypothesize that the presence of residues that are rarely found in turns might destabilize the structure of the turn. However, examination of the turns in lac Y revealed no support for this hypothesis. We do not find that overall hydrophobicity or net charge correlates with stability/instability of the turns, either. We find no evidence of motifs in the sequence that would fully explain the relative stability of the loops, even though the stability is expected to be sequence-specific. For example, while two of the loops that form stable structures in DMSO contain the hydrophobic stable sequence described previously, the other two loops that form stable structures do not (*42*).

It is interesting, however, to note that, of the 23 membrane protein loops from helical bundles that we have examined, the vast majority are not soluble in water and are therefore surprisingly hydrophobic. This observation suggests that the surfaces of lac Y (and bacteriorhodopsin and rhodopsin) are relatively hydrophobic.

The length of the peptide used to encompass the residues in a turn can influence the stability of the structure obtained. For example, the original work on the third cytoplasmic loop of rhodopsin used a 22mer, which formed only a short structured region. Upon lengthening this peptide to a 26mer, a full helix−turn−helix structure was observed. However, the issue of length is not simple. Thus peptides as short as 16 and 17 residues form stable structures of turns in solution. However, more generally it has been found that longer peptides are superior for studying localized secondary structure.

To better understand the issue of stability/instability of the turns, CD spectra were obtained of the peptides corresponding to turns of lac Y in the presence of negatively charged detergent micelles. CD allowed the investigation of secondary structure in the presence of this membrane mimetic where the quality of the NMR data that could be obtained was too poor to permit structure determination. Two classes of CD spectra were obtained: one consistent with turn structures and the other dominated by α-helix. Interestingly a strong correlation was observed between the class of CD spectra from the peptides in the presence of detergent micelles and stability/instability of the turn structure in DMSO as determined by NMR. Peptides corresponding to loops lpt3−4, lpt5−6, lpt8−9, and lpt10−11 produced CD spectra consistent with β-turns or a combination of β-turn and α-helix. These were the same peptides that formed stable turns in DMSO as determined by the NMR analysis. Six of the remaining seven peptides that were unstable in DMSO gave rise to CD spectra dominated by α-helix in the presence of negatively charged micelles. These data show a self-consistent pattern, in that the turns identified as relatively stable by NMR are stable as turns in micellar suspensions as well, and those found to be disordered in DMSO by NMR do not show evidence of turns in micellar suspensions. However, these data also collectively reveal that while many turn peptides in DMSO form structures consis-

tent with the X-ray crystal structure of the protein, detergent micelles introduce other factors, such as a negatively charged interface, that may lead to formation of non-native structures.

Not all loop peptides and transmembrane helix peptides form stable structures. However, we have not observed stable structures of peptide fragments of helices or loops that disagreed with the known secondary structure of the protein when examined in DMSO. Therefore, to date, the observation of a stable structure from a fragment, whether loop or helix, has reflected the secondary structure of the native protein.

These results suggest broader implications for membrane protein structure. One is that useful information about loop structure in helical bundles can be obtained from solution structural studies of fragments encompassing the loop regions. Since loops are difficult to predict in computer modeling, and since X-ray crystal structures of helical bundles often have high *B* factors for loop regions, the availability of other sources of information about loop structures is important.

Another implication is that some loops may contribute, albeit modestly, to the overall stability of the protein structure. The relative stability of loops may influence the overall stability of membrane proteins built around helical bundles (*43*, *44*). In both bacteriorhodopsin and rhodopsin, the loops have been implicated in the thermal stability of the protein, and all of the loops of these two proteins show stable structures in DMSO. These data collectively support the hypothesis that loop sequences that form stable loops independent of the rest of the protein likely contribute to the overall stability of the protein because of the propensity of the loop sequences to form loops.

Loops may contribute to the folding of membrane proteins as well (*45*). In the cases that the amino acid sequence drives loop formation, initial folding of the loop may help to direct the folding pathway of the protein, limiting sampling of folding pathways to those that correctly form the loop. In those cases where sequence does not drive loop formation, helix−helix interactions likely dominate the structural organization of the helical bundle during the folding process.

In contrast to previous work on bacteriorhodopsin and rhodopsin that showed stable structures from all loop fragments (*26*, *46*), the data on the loops from lac Y portray a more complicated picture. One important difference is that both bacteriorhodopsin and rhodopsin are smaller proteins than lac Y, with 7 transmembrane helices rather than the 12 transmembrane helices present in lac Y. It is possible that larger proteins are more conformationally flexible or that larger proteins involved in active transport are more conformationally flexible.

The stability or instability of the loops of a membrane protein may also play a role in function. Rigid-body movements of helices have been implicated in function of some membrane proteins. When helices are connected by short stable loops, the movement of one helix is directly communicated to the adjacent helix, forcing helices to move as partners. Such coordinated helix movements have been hypothesized for lac Y (*5*). In contrast, isolated rigid-body helix movements may be facilitated by flexible helix-connecting loops. Flexible loops can provide the conformational motility necessary to allow the tilting or deformation of one helix somewhat independently from the adjacent helix. Lac Y must enfold a sugar molecule in its protein core and release it to the opposite side of the membrane. This could be facilitated by flexible helix movements, which are facilitated by flexible loops. Therefore, the stability or instability of loops connecting transmembrane helices may be important to the molecular mechanism of transport.

In summary, this work suggests that studies of peptide fragments of helical bundles may be able to provide useful information on membrane proteins for which no X-ray crystal structure is available. While not a substitute for X-ray crystallography, nevertheless in cases where secondary structure is observed in the fragment, it is possible to identify which residues form the turn. The factors stabilizing loops are likely sequence-specific, but clear motifs are currently lacking. Longer peptides tend to favor stable loops, but specific lengths that favor stability cannot yet be identified. The loops of integral membrane proteins that have been examined to date are surprisingly hydrophobic. Since the majority of such loops that have been examined to date are stabilized by short-range interactions, loops may play a role in membrane protein folding. The instability of some loops might play a role in membrane protein function.

REFERENCES

1. Kaback, H. R., Sahin-Toth, M., and Weinglass, A. B. (2001) The kamikaze approach to membrane transport, *Nat. Rev. Mol. Cell Biol. 2*, 610−620.
2. Sahin-Toth, M., Lawrence, M. C., and Kaback, H. R. (1994) Properties of permease dimer, a fusion protein containing two lactose permease molecules from *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A. 91*, 5421−5425.
3. Kaback, H. R., and Wu, J. (1999) What to do while awaiting crystals of a membrane transport protein and thereafter, *Acc. Chem. Res. 32*, 805−813.
4. Wolin, C. D., and Kaback, H. R. (2001) Functional estimation of loop-helix boundaries in the lactose permease of *Escherichia coli* by single amino acid deletion analysis, *Biochemistry 40*, 1996−2003.
5. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R., and Iwata, S. (2003) Structure and mechanism of the lactose permease of *Escherichia coli*, *Science 301*, 610−615.
6. Popot, J.-L., and Engelman, D. M. (2000) Helical membrane protein folding, stability, and evolution, *Annu. Rev. Biochem. 69*, 881−922.
7. Lemmon, M. A., Flanagan, J. M., Hunt, J. F., Adair, B. D., Bormann, B.-J., Dempsey, C. E., and Engelman, D. M. (1992) Glycophorin A dimerization is driven by specific interactions between transmembrane α-helices, *J. Biol. Chem. 267*, 7683−7689.
8. Hunt, J. F., Earnest, T. N., Bousche, O., Kalghatgi, K., Reilly, K., Horvath, C., Rothschild, K. J., and Engelman, D. M. (1997) A biophysical study of integral membrane protein folding, *Biochemistry 36*, 15156−15176.
9. Berlose, J., Convert, O., Brunissen, A., Chassaing, G., and Lavielle, S. (1994) Three-dimensional structure of the highly conserved seventh transmembrane domain of G-protein-coupled receptors, *FEBS Lett. 225*, 827−843.
10. Lomize, A. L., Pervushin, K. V., and Arseniev, A. S. (1992) Spatial structure of (34−65)bacterioopsin polypeptide in SDS micelles determined from nuclear magnetic resonance data, *J. Biomol. NMR 2*, 361−372.
11. Barsukov, I. L., Nolde, D. E., Lomize, A. L., and Arseniev, A. S. (1992) Three-dimensional structure of proteolytic fragment 163−231 of bacterioopsin determined from nuclear magnetic resonance data in solution, *Eur. J. Biochem. 206*, 665−672.
12. Pervushin, K. V., Orekhov, V. Y., Popov, A. I., Musina, L. Y., and Arseniev, A. S. (1994) Three-dimensional structure of (1−71)bacterioopsin solubilized in methanol/chloroform and SDS micelles determined by ^{15}N-^{1}H heteronuclear NMR spectroscopy, *Eur. J. Biochem. 219*, 571−583.

13. Chopra, A., Yeagle, P. L., Alderfer, J. A., and Albert, A. (2000) Solution structure of the sixth transmembrane helix of the G-protein coupled receptor, rhodopsin, *Biochim. Biophys. Acta 1463*, 1−5.
14. Yeagle, P. L., Danis, C., Choi, G., Alderfer, J. L., and Albert, A. D. (2000) Three-dimensional structure of the seventh transmembrane helical domain of the G-protein receptor, Rhodopsin, *Mol. Vision* (www.molvis.org/molvis/v6/a17/).
15. Arshava, B., Liu, S. F., Jiang, H., Breslav, M., Becker, J. M., and Naider, F. (1998) Structure of segments of a G protein-coupled receptor: CD and NMR analysis of the *Saccharomyces cerevisiae* tridecapeptide pheromone receptor, *Biopolymers 46*, 343−357.
16. Gargaro, A. R., Bloomberg, G. B., Dempsey, C. E., Murray, M., and Tanner, M. J. (1994) The solution structures of the first and second transmembrane-spanning segments of band 3, *Eur. J. Biochem. 221*, 445−454.
17. Piserchio, A., Bisello, A., Rosenblatt, M., Chorev, M., and Mierke, D. F. (2000) Characterization of parathyroid hormone/receptor interactions: structure of the first extracellular loop, *Biochemistry 39*, 8153−8160.
18. Franzoni, L., Nicastro, G., Pertinhez, T. A., Oliveira, E., Nakaie, C. R., Paiva, A. C., Schreier, S., and Spisni, A. (1999) Structure of two fragments of the third cytoplasmic loop of the rat angiotensin II AT1A receptor. Implications with respect to receptor activation and G-protein selection and coupling, *J. Biol. Chem. 274*, 227−235.
19. Katragadda, M., Alderfer, J. L., and Yeagle, P. L. (2000) Solution structure of the loops of bacteriorhodopsin closely resemble the crystal structure, *Biochim. Biophys. Acta 1466*, 1−6.
20. Mierke, D. F., Royo, M., Pelligrini, M., Sun, H., and Chorev, M. (1996) Third cytoplasmic loop of the PTH/PTHrP receptor, *J. Am. Chem. Soc. 118*, 8998−9004.
21. Yeagle, P. L., Alderfer, J. L., and Albert, A. D. (1997) The first and second cytoplasmic loops of the G-protein receptor, rhodopsin, independently form β-turns, *Biochemistry 36*, 3864−3869.
22. Yeagle, P. L., Salloum, A., Chopra, A., Bhawsar, N., Ali, L., Kuzmanovski, G., Alderfer, J. L., and Albert, A. D. (2000) Structures of the intradiskal loops and amino terminus of the G-protein receptor, rhodopsin, *J. Pept. Res. 55*, 455−465.
23. Jung, H., Windhaber, R., Palm, D., and Schnackerz, K. D. (1995) NMR and circular dichroism studies of synthetic peptides derived from the third intracellular loop of the beta-adrenoceptor, *FEBS Lett. 358*, 133−136.
24. Abdulaev, N. G., Ngo, T., Chen, R., Lu, Z., and Ridge, K. D. (2000) Functionally discrete mimics of light-activated rhodopsin identified through expression of soluble cytoplasmic domains, *J. Biol. Chem. 275*, 39354−39363.
25. Gelber, E. I., Kroeze, W. K., Willins, D. L., Gray, J. A., Sinar, C. A., Hyde, E. G., Gurevich, V., Benovic, J., and Roth, B. L. (1999) Structure and function of the third intracellular loop of the 5-hyroxytryptamine (2A) receptor: the third intracellular loop is alpha-helical and binds purified arrestins, *J. Neurochem. 72*, 2206−2214.
26. Katragadda, M., Alderfer, J. L., and Yeagle, P. L. (2001) Assembly of a polytopic membrane protein structure from the solution structures of overlapping peptide fragments of bacteriorhodopsin, *Biophys. J. 81*, 1029−1036.
27. Katragadda, M., Chopra, A., Bennett, M., Alderfer, J. L., Yeagle, P. L., and Albert, A. D. (2001) Structures of the transmembrane helices of the G-protein coupled receptor, rhodopsin, *J. Pept. Res. 58*, 79−89.
28. Choi, G., Landin, J., Galan, J. F., Birge, R. R., Albert, A. D., and Yeagle, P. L. (2002) Structural studies of metarhodopsin II, the activated form of the G-protein coupled receptor, rhodopsin, *Biochemistry 41*, 7318−7324.
29. Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton−proton cross-relaxation networks in biological macromolecules, *Biochem. Biophys. Res. Commun. 95*, 1−6.
30. Wüthrich, K., Billeter, M., and Braun, W. J. (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton proton distance constraints with nuclear magnetic resonance, *J. Mol. Biol. 169*, 949−961.
31. Guntert, P., Braunk, W., and Wüthrich, K. (1991) Efficient computation of three-dimensional protein structures in solution from NMR data using the program DIANA and the supporting programs CALIBA, HABAS, and GLOMSA, *Mol. Biol. 217*, 517−530.
32. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J. Mol. Graphics 14*, 33−38, 27−88.
33. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry 31*, 1647−1651.
34. Chung, D. A., Zuiderweg, E. R., Fowler, C. B., Soyer, O. S., Mosberg, H. I., and Neubig, R. R. (2002) NMR structure of the second intracellular loop of the alpha 2A adrenergic receptor: evidence for a novel cytoplasmic helix, *Biochemistry 41*, 3596−3604.
35. Ulfers, A. L., McMurry, J. L., Kendall, D. A., and Mierke, D. F. (2002) Structure of the third intracellular loop of the human cannabinoid 1 receptor, *Biochemistry 41*, 11344−11350.
36. Askin, D., Bloomberg, G. B., Chambers, E. J., and Tanner, M. J. A. (1998) NMR solution structure of a cytoplasmic surface loop of the human red cell anion transporter, band 3, *Biochemistry 37*, 11670−11678.
37. Nicastro, G., Peri, F., Franzoni, L., de Chiara, C., Sartor, G., and Spisni, A. (2003) Conformational features of a synthetic model of the first extracellular loop of the angiotensin II AT1A receptor, *J. Pept. Sci. 9*, 229−243.
38. Venkatraman, J., Nagana Gowda, G. A., and Balaram, P. (2002) Structural analysis of synthetic peptide fragments from EmrE, a multidrug resistance protein, in a membrane-mimetic environment, *Biochemistry 41*, 6631−6639.
39. Ulfers, A. L., Piserchio, A., and Mierke, D. F. (2002) Extracellular domains of the neurokinin-1 receptor: structural characterization and interactions with substance P, *Biopolymers 66*, 339−349.
40. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol. 157*, 105−132.
41. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology, *Biochemistry 33*, 3038−3049.
42. Muñoz, V., Blanco, F. J., and Serrano, L. (1995) The hydrophobic-stable motif and a role for loop residues in α-helix stability and protein folding, *Nat. Struct. Biol. 2*, 380−385.
43. Kahn, T. W., Sturtevant, J. M., and Engelman, D. M. (1992) Thermodynamic measurements of contributions of helix-connecting loops and of retinal to the stability of bacteriorhodopsin, *Biochemistry 31*, 8829−8839.
44. Landin, J. S., Katragadda, M., and Albert, A. D. (2001) Thermal destabilization of rhodopsin and opsin by proteolytic cleavage in bovine rod outer segment disk membranes, *Biochemistry 40*, 11176−11183.
45. Engelman, D. M., and Steitz, T. A. (1981) The spontaneous insertion of proteins into and across membranes; The helical hairpin hypothesis, *Cell 23*, 411−422.
46. Yeagle, P. L., Choi, G., and Albert, A. D. (2001) Studies on the structure of the G-protein coupled receptor rhodopsin including the putative G-protein binding site in unactivated and activated forms, *Biochemistry 40*, 11932−11937.

BI049000S