An approach to membrane protein structure without crystals

Paul L. Sorgen*†, Yonglin Hu†‡, Lan Guan‡, H. Ronald Kaback‡§, and Mark E. Girvin*§

*Biochemistry Department, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461; and ‡Howard Hughes Medical Institute, Departments of Physiology and Microbiology, Immunology, and Molecular Genetics, Molecular Biology Institute, University of California, Los Angeles, CA 90095-1662

Contributed by H. Ronald Kaback, September 11, 2002

The lactose permease of *Escherichia coli* catalyzes coupled translocation of galactosides and H^+ across the cell membrane. It is the best-characterized member of the Major Facilitator Superfamily, a related group of membrane proteins with 12 transmembrane domains that mediate transport of various substrates across cell membranes. Despite decades of effort and their functional importance in all kingdoms of life, no high-resolution structures have been solved for any member of this family. However, extensive biochemical, genetic, and biophysical studies on lactose permease have established its transmembrane topology, secondary structure, and numerous interhelical contacts. Here we demonstrate that this information is sufficient to calculate a structural model at the level of helix packing or better.

bioenergetics | membrane transport | lactose permease | site-directed thiol cross-linking | engineered divalent metal-binding sites

The lactose permease (LacY) is a 12-transmembrane helix bundle that transduces free energy stored in an electrochemical H^+ gradient into a concentration gradient of galactosides or vice versa (i.e., galactoside/H^+ symport). Because LacY and similar membrane proteins are resistant to traditional means of structural analysis, alternative approaches have been developed to discern the overall 3D fold. LacY has been the test bed for most of the development of these methods, including mapping transmembrane segments by *phoA* fusions, protein insertion into loops and deletion analysis, accessibility of natural or uniquely engineered Cys residues to membrane-impermeant reagents, determining secondary structure by CD, laser Raman and Fourier transform IR spectroscopy, and identifying long-range contacts through second-site suppressers, thiol cross-linking, excimer fluorescence, engineered Mn(II) binding sites, site-directed electron paramagnetic resonance, and discontinuous mAb epitope mapping (1–3).

Data obtained with these methods provide an overview of the organization of the protein. This information together with the observation that only six residues play an irreplaceable role in the mechanism, as well as detailed characterization of mutants in these six residues, has led to a proposed mechanism for LacY that is consistent with a large body of evidence (3). These data should also be useful in deriving structural models of LacY at the level of helix packing or better with respect to the positioning of the irreplaceable residues.

Distance constraint-based torsion-angle dynamics structure calculation methods (4) are the logical choice for such efforts. Even the earliest reports stress that NMR structure calculation methods are general solutions for generating families of structural models that fit any type of constraint data. Here, we show that by using helical backbone constraints for the transmembrane segments with long-range constraints derived from thiol cross-linking and engineered Mn(II) binding sites, 3D structural models of LacY can be calculated that satisfy all constraints and converge to a single family of models with 1.1 Å rms deviations for the backbone atoms of the transmembrane segments.

Materials and Methods

Constraints. The boundaries of the transmembrane helices were approximated to be residues: 11–33, 47–67, 73–91, 111–133, 138–159, 168–188, 222–247, 261–278, 291–308, 315–335, 346–372, and 381–397 (5). Helical backbone H bonds were implemented as distance constraint ranges of 1.60–2.05 Å between O_i and NH_{i+4}, and 2.2–3.0 Å between O_i and N_{i+4} for the residues in these segments. Helical torsion-angle constraints of $-60 \pm 5°$ for *phi* and $-40 \pm 5°$ for *psi* were used for the same residues. The long-range distance constraints were derived from Cys cross-linking results (6–19) and from Mn(II)-binding studies with *bis*-His mutants (20–24). For the four different cross-link types (direct disulfide, 1,3 propanediyl *bis*-methanethiosulfonate, *N,N′-o*-phenylenedimaleimide-, *N,N′-p*-phenylenedimaleimide-, or 1,6-*bis*-maleimidohexane-mediated) target C^β to C^β distances of 3–5, 3–6, 2–6, 2–10, or 6–16 Å, respectively, were used. For Gly residues, constraints were to their C^α, and the distances were increased by 1.5 Å. Many pairs of Cys residues formed cross-links with more than one reagent. Distances for these pairs were interpolated from the reactivity profiles. For engineered Mn(II)-binding sites, target C^β-to-C^β distances of 8.9–9.4 Å were used. In total, 99 long-range constraints were defined. When lactose was included, the center of the galactopyranosyl ring was constrained to be within 5 Å of Trp-151, the C-1 position within 5 Å of the Cys-148 C^β, the center of the glucopyranosyl ring to be within 5 Å of the Ala-122 C^β (25), the C-4 OH group of the galactopyranosyl ring to be within 3.3 Å of the Arg-144 NH1 or NH2, and the C-3 OH group to be within 3.3 Å of Glu-269.

Structure Calculation. Structural models were calculated from a fully extended starting conformation by torsion-angle dynamics-simulated annealing methods (4) by using the CNS program (26). For each of 250 independent simulations, the distance and angle constraints were gradually imposed over a 40-ps equilibration at 100,000 K. The simulation temperature was increased to 150,000 K, and annealed by cooling to 0 K over 200 ps. All models were refined by Cartesian dynamics annealing (2,000 to 0 K, over 15 ps), and the 25 models that best fit the constraints were refined further with a second Cartesian dynamics annealing run followed by energy minimization.

Results and Discussion

A typical protein structure calculation from NMR data begins with a fully extended polypeptide chain and uses local distance and angle constraints based on backbone nuclear Overhauser effects and coupling constants, along with long-range distance constraints from nuclear Overhauser effects to draw the protein into its tertiary fold over the course of a high-temperature torsion-angle dynamics-simulated annealing procedure. A sim-

Abbreviation: LacY, lactose permease.

†P.L.S. and Y.H. contributed equally to this work.

§To whom correspondence may be addressed. E-mail: girvin@aecom.yu.edu or ronaldk@hhmi.ucla.edu.

BIOCHEMISTRY

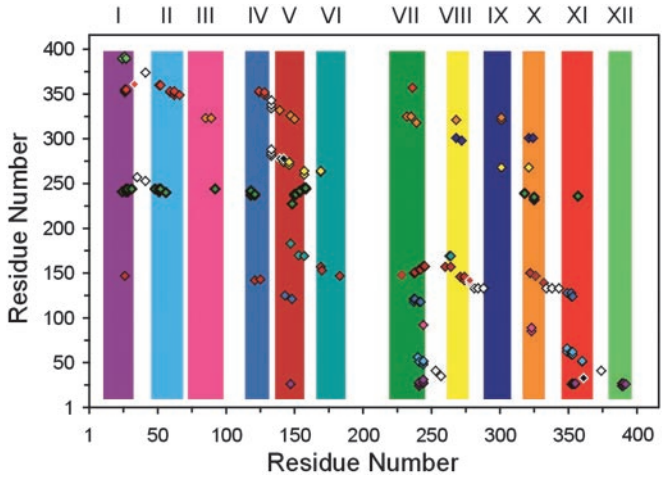

Fig. 1. Cross-linking constraints used for calculation of LacY structural models. Each vertical color strip represents a transmembrane helix (helix number above). Symbols represent a cross-link with colors corresponding to the second helix involved. White diamonds with black borders are cross-links between loop residues. Colored diamonds with white borders are cross-links to a loop residue from the corresponding helix. Black diamonds with white borders identify cross-links from loop residues.

ilar method is used here for LacY, but local helical H-bonding and angle constraints were created for the transmembrane helices, and long-range distance constraints were derived from the length of the reagent that gives optimal thiol cross-linking for pairs of engineered Cys residues or with the six irreplaceable residues, from distances derived from Mn(II)-binding studies with *bis*-His replacement mutants. A summary of the constraints is shown in Fig. 1. The lengths of the long-range constraints range from 2 to 22 Å, and the number of constraints per helix ranges from 3 for helix III to 30 for helix VII. Simulations were run both with and without lactose as substrate. An extremely high initial simulation temperature (150,000 K) was needed for successful annealing of the models, but the parameters are otherwise typical for NMR protein structure calculations.

The resulting structural models that satisfy all constraints (to within 0.4 Å) fall into a single family (Fig. 2). The constraints define the structure relatively well, with pair-wise rms deviations for the backbone atoms of all transmembrane helices of 1.1 Å. The backbone is not altered in any measurable way by inclusion of ligand in the calculations. Variations in the average displacements along the individual helices are shown in Fig. 3. The loops were included in the structure calculations to provide additional covalent constraints between the ends of the helices, but because no constraints are available for secondary structure and few for long-range contacts, the loops are not included in the coordinates deposited (Protein Data Bank ID code 1M2U). Constraints are also lacking for the N-terminal half of helix I. In the models calculated initially, the N terminus exited the helical bundle at what would be the middle of the membrane. Because this topology conflicts with biochemical data (5), helix I was constrained to be transmembrane in all subsequent calculations.

The 12 transmembrane helices of LacY form a nearly circular bundle ≈48 Å in cross section (Fig. 2*B*), which is generally consistent with other measurements (27, 28). Helices IV, V, VII, and X are near the center of the bundle, removed from significant contact with the lipid bilayer. The remaining helices pack around this core, with helices I and XII packing next to each other. The transmembrane helices are not arranged sequentially, but form clusters with helices I–III, IV–VI, VII–X, and XI–XII nearest to each other. A more symmetric model, based on low-sequence homology between the two halves of LacY and other sequence comparisons, has been proposed (29) in which each helix in the two halves of the protein occupy symmetry-related positions about a central cavity. Although the structures calculated here show hints of such an arrangement (e.g., nearly symmetric positions for helix II and VII and helix IV and X), it seems that significant divergence has occurred from a hypothetical fused homodimer, with the two halves of LacY interdigitating with each other. Also, many of the experimentally derived distance constraints are inconsistent with the earlier model (29). Instead of a central cavity, two deep clefts accessible from the outer surface are observed surrounded by helices V, VII, X, and VIII, and II, III, X, and VII, and an internal cavity between helices I, IV, V, and VII that is most accessible from the

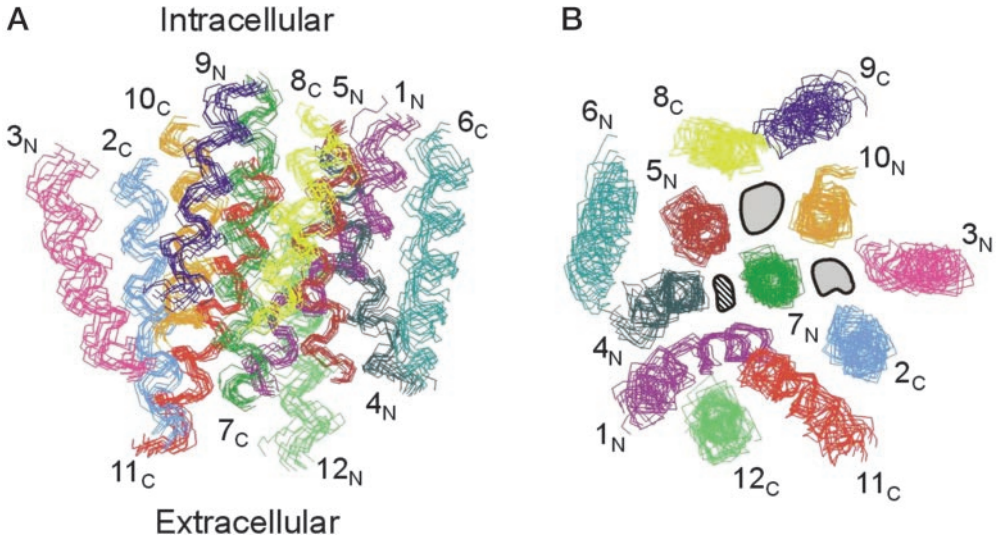

Fig. 2. Structural models of LacY. Side (*A*) and cytoplasmic (*B*) views of backbone traces of the 10 LacY conformers that best fit the constraints are aligned by superimposing the backbone atoms of the transmembrane helices (loop segments removed). Numbers identify each transmembrane helix, with subscripts indicating the N or C termini. The colors of the helices correspond to those of Fig. 1. Two cavities accessible from the outer surface are indicated diagrammatically by shaded ovals; the cavity accessible from the internal surface is indicated by a striped oval. The figure was prepared by using the program MOLMOL (43).

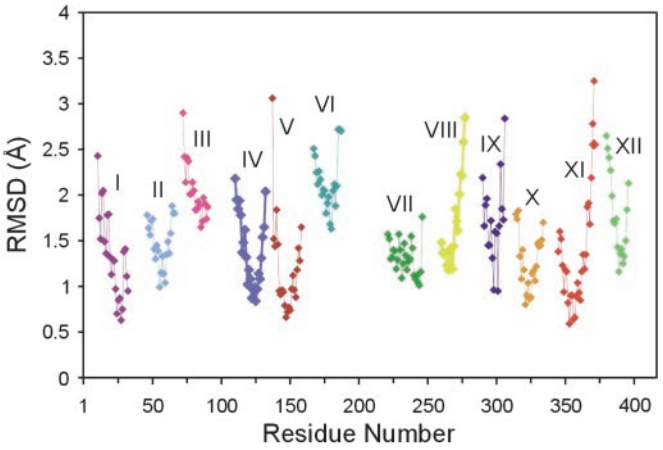

Fig. 3. Average global backbone displacements for the transmembrane helices of the LacY structural models. The helices are identified by number, and colors correspond to Fig. 1. RMSD, rms deviations.

cytoplasmic surface is observed. Although a 2D projection map of the oxalate transporter OxlT (30) indicates that each helix in the two halves of the protein occupy symmetry-related positions, like LacY, neither the Na^+/H^+ antiporter NhaA (31) nor the Na^+/sugar symporter MelB (32) exhibit such symmetry.

The calculated models are entirely consistent with features typical of membrane proteins, and with the large variety of biochemical data obtained for LacY. As expected, the most hydrophobic surfaces of the transmembrane helices face the surrounding lipid, whereas the protein core is more polar. The lipid-exposed faces include the clusters of residues most tolerant to mutation or alkylation of Cys-replacement mutants (2). A recent study identifies residues exposed on the periphery of the 12-helix bundle comprising LacY (33), and these residues are located at the periphery in the structural models. The calculated packing arrangement also predicts the interresidue distances estimated by other techniques listed above that were not used as constraints in the calculations reported here. Finally, the clefts observed between transmembrane helices explain the surprising accessibility of the protein backbone to water, as judged by the high rate of backbone amide H/D exchange (34, 35), the accessibility of engineered Cys residues located at or near the cytoplasmic face of the membrane to modification by water-soluble thiol reagents from the periplasmic surface (36–39), and accessibility of galactosides to the binding site from both sides of the membrane (40).

Evidence for the role of the six irreplaceable residues in the symport mechanism, as well as their spatial relationships, has been reviewed (3). A carboxyl group at position 126 (helix IV) and a guanidinium at position 144 (helix V) are obligatory for ligand binding, and the two positions are within close proximity (Fig. 4). A carboxyl group at position 269 (helix VIII) is also obligatory for ligand binding, and this residue is positioned between His-322 (helix X) and Arg-144 (helix V). Recent experiments (A. B. Weinglass, J. P. Whitelegge, Y.H., K. F. Faull, and H.R.K., unpublished work) show that Glu-269 interacts with ligand, probably with the C-3 OH of the galactopyranosyl ring. This residue likely plays an important role in H^+ translocation as well. Arg-302 (helix IX) and Glu-325 (helix X), which are also in close proximity, clearly play a critical role in H^+ translocation with Arg-302 facilitating deprotonation of Glu-325 after release of substrate on the inner surface of the membrane when LacY relaxes back to the ground state. Also shown are the salt bridge between Asp-237 (helix VII) and Lys-358 (helix XI), which does not play a direct role in the symport mechanism, but is important for membrane insertion, as well as the salt bridge between Asp-240 (helix VII) and Lys-319 (helix X), which may play an indirect role in the mechanism (3), but is not obligatory.

A more detailed view of binding site interactions in LacY based on biochemical observations is shown in Fig. 5 with lactose as ligand. The galactopyranosyl ring contains all of the determinants for specificity, and the C-4 OH is most important (41). Galactose is the most specific substrate for LacY, but has very low affinity ($K_d \approx 30$ mM); however, various adducts at the 1-position can increase affinity by more than 3 orders of magnitude (42). Cys-148 (helix V) interacts weakly and hydrophobically with the galactosyl moiety of substrate, and Ala-122 (helix IV) is in close proximity to the nongalactosyl moiety (25). Although LacY with W151F or W151Y transports lactose almost as well as WT, the mutants exhibit about a 100- or a 50-fold decrease in affinity, respectively (L.G. and H.R.K., unpublished work), indicating the Trp-151 stacks with the hydrophobic face of the galactopyranosyl ring, placing it at a right angle with helix IV and abutting Cys-148 near the 1 position. In this orientation, the C-4 OH can H-bond directly with either NH1 or NH2 of

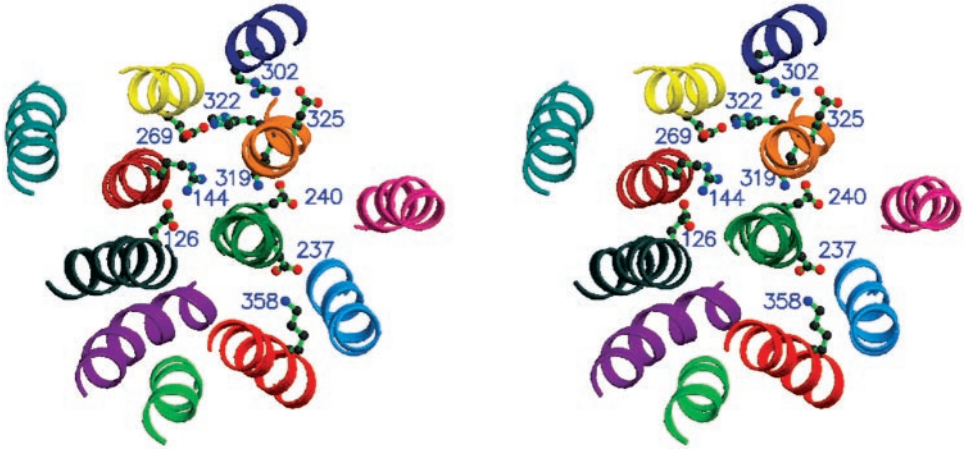

Fig. 4. Stereoview of LacY helix packing showing important positions. Glu-126 (helix IV), Arg-144 (helix V), Glu-269 (helix VIII), Arg-302 (helix IX), and His-322 and Glu-325 (helix X) are labeled. Also shown are the salt bridges between Asp-237 (helix VII) and Lys-358 (helix XI) and between Asp-240 (helix VII) and Lys-319 (helix X). The transmembrane helices are colored as in Fig. 1. Prepared by using MOLSCRIPT (44) and RASTER3D (45).

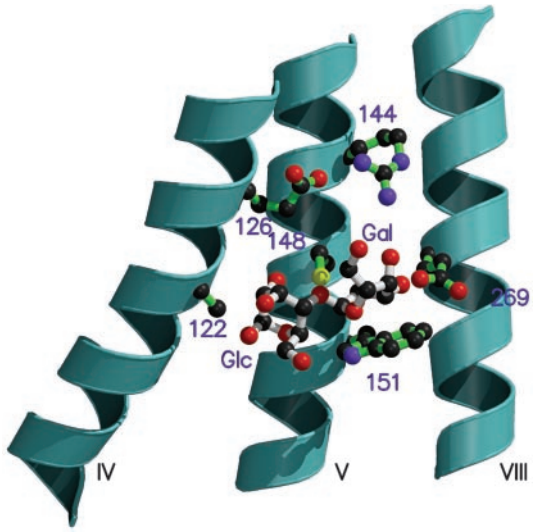

Fig. 5. Binding-site interactions in LacY. The model shown is based on biochemical observations, and lactose is shown as ligand. See text for details.

Arg-144. Because the C-3 OH is close to Glu-269 (helix VIII) but at an angle, it is reasonable to suggest that a water molecule may mediate this interaction.

The method described here has enabled us to calculate structural models for LacY, a member of the Major Facilitator Superfamily that has been used extensively for the development of nontraditional methods for obtaining structure/function that are now widely used to study membrane proteins. The structures are in complete accord with a variety of biochemical and biophysical measurements that have been made on LacY. Compared with traditional structural methods such as x-ray crystallography, NMR, and cryo-electron microscopy, the resolution achieved seems at least as good as that of cryo-electron microscopy, particularly with respect to bypassing the problem of identifying individual helices and the positioning of the residues essential for the mechanism. Resolution varies with the number and distribution of the measured constraints, but can be improved iteratively by more measurements on the less-defined portions of the structure or on regions of functional significance. Distance measurements in the presence of substrate would also allow ligand-induced conformational changes to be calculated.

We are particularly indebted to Jianhua Wu for initiating thiol cross-linking experiments with LacY and for most of the measurements that led to the constraints used and to Kirsten and Heinrich Jung, John Voss, Molly He, Qingda Wang, Christopher Wolin, Wei Zhang, and Min Zhao for other constraints. This work was supported in part by National Institutes of Health Grants DK51131:06 (to H.R.K.), GM55371 (to M.E.G.), and F32 GM20504 (to P.L.S.).

1. Kaback, H. R. & Wu, J. (1997) *Q. Rev. Biophys.* **30,** 333–364.
2. Frillingos, S., Sahin-Tóth, M., Wu, J. & Kaback, H. R. (1998) *FASEB J.* **12,** 1281–1299.
3. Kaback, H. R., Sahin-Tóth, M. & Weinglass, A. B. (2001) *Nat. Rev. Mol. Cell. Biol.* **2,** 610–622.
4. Stein, E. G., Rice, L. M. & Brünger, A. T. (1997) *J. Magn. Reson.* **124,** 154–164.
5. Wolin, C. D. & Kaback, H. R. (2001) *Biochemistry* **40,** 1996–2003.
6. Wu, J. & Kaback, H. R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 14498–14502.
7. Wu, J. & Kaback, H. R. (1997) *J. Mol. Biol.* **270,** 285–293.
8. Sun, J., Kemp, C. R. & Kaback, H. R. (1998) *Biochemistry* **37,** 8020–8026.
9. Wu, J., Hardy, D. & Kaback, H. R. (1998) *J. Mol. Biol.* **282,** 959–967.
10. Wu, J., Hardy, D. & Kaback, H. R. (1998) *Biochemistry* **37,** 15785–15790.
11. Wu, J., Hardy, D. & Kaback, H. R. (1999) *Biochemistry* **38,** 2320–2325.
12. Wu, J., Hardy, D. & Kaback, H. R. (1999) *Biochemistry* **38,** 1715–1720.
13. Wang, Q. & Kaback, H. R. (1998) *Biochemistry* **38,** 3120–3126.
14. Wang, Q., Voss, J., Hubbell, W. L. & Kaback, H. R. (1998) *Biochemistry* **37,** 4910–4915.
15. Wang, Q. & Kaback, H. R. (1999) *J. Mol. Biol.* **291,** 683–692.
16. Wang, Q. & Kaback, H. R. (1999) *Biochemistry* **38,** 16777–16782.
17. Kwaw, I., Sun, J. & Kaback, H. R. (2000) *Biochemistry* **39,** 3134–3140.
18. Guan, L., Weinglass, A. B. & Kaback, H. R. (2001) *J. Mol. Biol.* **312,** 69–77.
19. Zhang, W., Guan, L. & Kaback, H. R. (2002) *J. Mol. Biol.* **315,** 53–62.
20. Jung, K., Voss, J., He, M., Hubbell, W. L. & Kaback, H. R. (1995) *Biochemistry* **34,** 6272–6277.
21. He, M. M., Voss, J., Hubbell, W. L. & Kaback, H. R. (1995) *Biochemistry* **34,** 15661–15666.
22. He, M. M., Voss, J., Hubbell, W. L. & Kaback, H. R. (1995) *Biochemistry* **34,** 15667–15670.
23. He, M. M., Voss, J., Hubbell, W. L. & Kaback, H. R. (1997) *Biochemistry* **36,** 13682–13687.
24. Zhao, M., Zen, K. C., Hubbell, W. L. & Kaback, H. R. (1999) *Biochemistry* **38,** 7407–7412.
25. Guan, L., Sahin-Tóth, M. & Kaback, H. R. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 6613–6618.
26. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kuntsleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
27. Costello, M. J., Escaig, J., Matsushita, K., Viitanen, P. V., Menick, D. R. & Kaback, H. R. (1987) *J. Biol. Chem.* **262,** 17072–17082.
28. Zhuang, J., Privé, G. G., Verner, G. E., Ringler, P., Kaback, H. R. & Engel, A. (1999) *J. Struct. Biol.* **125,** 63–75.
29. Green, A. L., Anderson, E. J. & Brooker, R. J. (2000) *J. Biol. Chem.* **275,** 23240–23246.
30. Heymann, J. A., Sarker, R., Hirai, T., Shi, D., Milne, J. L., Maloney, P. C. & Subramaniam, S. (2001) *EMBO J.* **20,** 4408–4413.
31. Williams, K. A. (2000) *Nature* **403,** 112–115.
32. Hacksell, I., Rigaud, J. L., Purhonen, P., Pourcher, T., Hebert, H. & Leblanc, G. (2002) *EMBO J.* **21,** 3569–3574.
33. Guan, L., Murphy, F. D. & Kaback, H. R. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 3475–3480.
34. Le Coutre, J., Kaback, H. R., Patel, C. K., Heginbotham, L. & Miller, C. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6114–6117.
35. Patzlaff, J. S., Moeller, J. A., Barry, B. A. & Brooker, R. J. (1998) *Biochemistry* **37,** 15363–15375.
36. Venkatesan, P., Liu, Z., Hu, Y. & Kaback, H. R. (2000) *Biochemistry* **39,** 10649–10655.
37. Venkatesan, P., Kwaw, I., Hu, Y. & Kaback, H. R. (2000) *Biochemistry* **39,** 10641–10648.
38. Venkatesan, P., Hu, Y. & Kaback, H. R. (2000) *Biochemistry* **39,** 10656–10661.
39. Kwaw, I., Zen, K.-C., Hu, Y. & Kaback, H. R. (2001) *Biochemistry* **40,** 10491–10499.
40. Frillingos, S. & Kaback, H. R. (1996) *Biochemistry* **35,** 3950–3956.
41. Sahin-Tóth, M., Lawrence, M. C., Nishio, T. & Kaback, H. R. (2001) *Biochemistry* **43,** 13015–13019.
42. Sahin-Tóth, M., Gunawan, P., Lawrence, M. C., Toyokuni, T. & Kaback, H. R. (2002) *Biochemistry*, in press.
43. Koradi, R., Billeter, M. & Wüthrich, K. (1996) *J. Mol. Graphics* **14,** 51–55.
44. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
45. Merritt, E. A. & Bacon, D. J. (1977) *Methods Enzymol.* **277,** 504–524.