tion apparatus is poised to proceed. In the case of trans cleavage, the requirement for two precisely aligned protomers could serve to restrain inappropriate cleavages, as proposed for FLP (*17*). A second potential role of the flexible catalytic loop may be to accommodate strand transfer during synapsis. Movement of the protein-DNA complex may be required for access of the invading strand to the catalytic center, and this movement might be facilitated by adjustment of the active site loop.

The structure of λ Int c170 with its flexibly tethered tyrosine nucleophile also suggests a mechanism to explain another variable property of Int family recombinases. Some members of the Int family are, like λ Int, quite fastidious in their requirement for DNA-DNA homology within the overlap regions of two recombining partners (*46*). Other Int family recombinases, such as the Tn1545 and Tn916 transposases, are more relaxed in their response to heterologies between recombining partners (*47*, *48*). This difference might be due in part to the variety of amino acids occupying this segment of Int family recombinases (Fig. 2B). Alterations in the delivery arm of Tyr^{342} could be expected to significantly influence such parameters as the rate, reversibility, and specificity of the DNA cleavage and ligation reactions.

REFERENCES AND NOTES

1. J. M. Leong *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **49**, 707 (1984).
2. K. E. Abremski and R. H. Hoess, *Protein Eng.* **5**, 87 (1992).
3. G. W. Blakely and D. J. Sherratt, *Mol. Microbiol.* **20**, 234 (1996).
4. A. Landy, *Annu. Rev. Biochem.* **58**, 913 (1989).
5. N. L. Craig, *Annu. Rev. Genet.* **22**, 77 (1988).
6. P. D. Sadowski, *FASEB* **7**, 760 (1993).
7. W. M. Stark, M. R. Boocock, D. J. Sherratt, *Trends Genet.* **8**, 432 (1992).
8. A. Landy, *Curr. Opinion Genet. Dev.* **3**, 699 (1993).
9. Y. Kikuchi and H. A. Nash, *J. Biol. Chem.* **253**, 7149 (1978).
10. L. Moitoso de Vargas, C. A. Pargellis, N. M. Hasan, E. W. Bushman, A. Landy, *Cell* **54**, 923 (1988).
11. R. S. Tirumalai, E. Healey, A. Landy, in preparation.
12. C. A. Pargellis, S. E. Nunes-Düby, L. Moitoso de Vargas, A. Landy, *J. Biol. Chem.* **263**, 7678 (1988).
13. R. L. Parsons, P. V. Prasad, R. M. Harshey, M. Jayaram, *Mol. Cell. Biol.* **8**, 3303 (1988).
14. Y. W. Han, R. I. Gumport, J. F. Gardner, *J. Mol. Biol.* **235**, 908 (1994).
15. L. K. Arciszewska and D. J. Sherratt, *EMBO J.* **14**, 2112 (1995).
16. G. W. Blakely, A. O. Davidson, D. J. Sherratt, *J. Mol. Biol.* **265**, 30 (1997).
17. J. Chen, J. Lee, M. Jayaram, *Cell* **69**, 647 (1992).
18. S. E. Nunes-Düby *et al.*, *EMBO J.* **13**, 4421 (1994).
19. L. Holm and C. Sander, *J. Mol. Biol.* **233**, 123 (1993).
20. F. Dyda *et al.*, *Science* **266**, 1981 (1994).
21. G. Bujacz *et al.*, *J. Mol. Biol.* **253**, 333 (1995).
22. P. Rice and K. Mizuuchi, *Cell* **82**, 209 (1995).
23. M. R. Sanderson *et al.*, *ibid.* **63**, 1323 (1990).
24. C. D. Lima, J. C. Wang, A. Mondragon, *Nature* **367**, 138 (1994).
25. J. M. Berger, S. J. Gamblin, S. C. Harrison, J. C. Wang, *ibid.* **379**, 225 (1996).
26. G. J. Barton, *Methods Enzymol.* **183**, 403 (1990).
27. S. Nunes-Düby, unpublished results.
28. L. Dorgai, E. Yagil, R. A. Weisberg, *J. Mol. Biol.* **252**, 178 (1995).
29. E. Yagil, L. Dorgai, R. Weisberg, *ibid.*, p. 163.
30. C. J. E. Schwartz and P. D. Sadowski, *ibid.* **205**, 647 (1989).
31. B. R. Evans *et al.*, *J. Biol. Chem.* **265**, 18504 (1990).
32. M. G. Rossman, A. Liljas, C. Branden, L. J. Banaszak, in *The Enzymes* (Academic Press, New York, 1975), pp. 61–102.
33. J. Zheng *et al.*, *Acta Crystallogr.* **D49**, 362 (1993).
34. H. L. DeBondt *et al.*, *Nature* **363**, 595 (1993).
35. S. A. Weston, A. Lahm, D. Suck, *J. Mol. Biol.* **226**, 1237 (1992).
36. P. Rice and T. A. Steitz, *Structure* **2**, 371 (1994).
37. W. Yang and T. A. Steitz, *Cell* **82**, 193 (1995).
38. J. Lee, I. Whang, M. Jayaram, *EMBO J.* **13**, 5346 (1994).
39. Y. W. Han, R. I. Gumport, J. F. Gardner, *ibid.* **12**, 4577 (1993).
40. A. C. Shaikh and P. D. Sadowski, *J. Biol. Chem.* **272**, 5695 (1997).
41. X. Qian and M. M. Cox, *Genes Dev.* **9**, 2053 (1995).
42. W. M. Stark and M. R. Boocock, *Trends Genet.* **11**, 121 (1995); M. Jayaram and J. Lee, *ibid.*, p. 432.
43. G. B. Panigrahi and P. D. Sadowski, *J. Biol. Chem.* **269**, 10940 (1994).
44. W. Ross and A. Landy, *Cell* **33**, 261 (1983).
45. R. S. Spolar and T. Record, Jr., *Science* **263**, 777 (1994).
46. R. A. Weisberg, L. W. Enquist, C. Foeller, A. Landy, *J. Mol. Biol.* **170**, 319 (1983).
47. M. G. Caparon and J. R. Scott, *Cell* **59**, 1027 (1989).
48. P. Trieu-Cuot, C. Poyart-Salmeron, C. Carlier, P. Courvalin, *Mol. Microbiol.* **8**, 179 (1994).
49. Z. Otwinowski, *Data Collection and Processing* (SERC Daresbury Laboratory, Warrington, UK, 1993), pp. 56–62.
50. T. C. Terwilliger and D. Eisenberg, *Acta Crystallogr.* **A39**, 813 (1983).
51. K. D. Cowtan and P. Main, *ibid.* **D49**, 148 (1993).
52. K. Zhang and P. Main, *ibid.* **A46**, 41 (1990).
53. T. A. Jones, in *Molecular Replacement* (SERC Caresbury Laboratory, Warrington, UK, 1992), pp. 91–105.
54. T. A. Jones, *O-The Manual* (Uppsala, Sweden, 1992).
55. R. J. Read, *Acta Crystallogr.* **A42**, 140 (1986).
56. A. T. Brunger, *X-PLOR Version 3.1, A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
57. A. T. Brunger, *Nature* **355**, 472 (1992).
58. S. V. Evans, *J. Mol. Graphics* **11**, 134 (1990).
59. G. J. Barton, *Protein Eng.* **6**, 37 (1993).

60. We thank E. Healey for purified protein, S. Nunes-Düby for help with sequence alignments, R. Sweet for assistance with data collection at beamline X-12C, National Synchrotron Light Source (Upton, NY), T. Oliveira for technical assistance, J. Boyles for assistance with manuscript preparation, and J. Cheah, S. Doublié, S. Nunes-Düby and other members of our research groups for their assistance and comments. Supported by the Lucille P. Markey Charitable Trust (TE) and NIH grants AI13544 and GM33928 (AL) and a Howard Hughes Medical Institute predoctoral fellowship (HJK). The coordinates have been deposited in the Brookhaven Protein Data Base with accession number 1AE9.

22 November 1996; accepted 25 February 1997

A Transmembrane Helix Dimer: Structure and Implications

Kevin R. MacKenzie, James H. Prestegard, Donald M. Engelman*

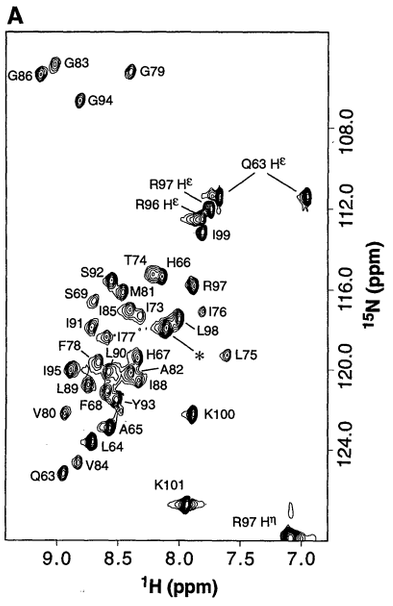
The three-dimensional structure of the dimeric transmembrane domain of glycophorin A (GpA) was determined by solution nuclear magnetic resonance spectroscopy of a 40-residue peptide solubilized in aqueous detergent micelles. The GpA membrane-spanning α helices cross at an angle of −40 degrees and form a small but well-packed interface that lacks intermonomer hydrogen bonds. The structure provides an explanation for the previously characterized sequence dependence of GpA dimerization and demonstrates that van der Waals interactions alone can mediate stable and specific associations between transmembrane helices.

Glycophorin A, the primary sialoglycoprotein of human erythrocyte membranes, forms noncovalent dimers by sequence-specific, reversible association of its single hydrophobic membrane-spanning domain (*1*–*3*). We have determined the dimeric structure of GpAtm, a peptide corresponding to residues 62 to 101 of human GpA, by heteronuclear nuclear magnetic resonance (NMR) methods (Fig. 1). Nuclear Overhauser effect (NOE) distance restraints and *J* coupling–derived dihedral restraints were used to generate a family of 20 structures (*4*). Residues 62 to 71, which lie outside the membrane (*5*) and are unnecessary for dimerization (*2*), are poorly defined by this procedure, but within the transmembrane region the structures differ from the average by only 0.4 Å [root mean square (rms)] for backbone and 0.8 Å (rms) for all nonhydrogen atoms. The well-defined interface between the hydrophobic helices allows us to describe the physical basis for helix-helix association, the second stage of the proposed "two-stage model" for membrane protein folding (*6*).

Residues 73 to 96 of GpA form α helices that interact at a −40° (right-handed) crossing angle in a symmetric GpAtm dimer (Fig. 2). Side chain and backbone atoms of seven residues make favorable van der Waals inter-

K. R. MacKenzie and D. M. Engelman, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520–8114, USA.
J. H. Prestegard, Department of Chemistry, Yale University, New Haven, CT 06520–8107, USA.

*To whom correspondence should be addressed.



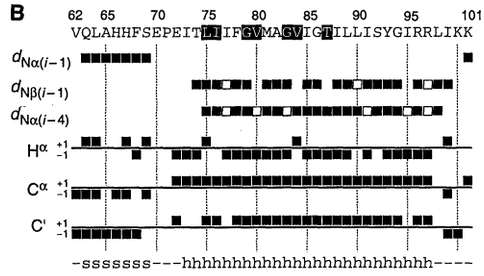


Fig. 1. (**A**) ^{1}H-^{15}N heteronuclear single-quantum coherence spectrum (*17*) of GpAtm in 5% dodecylphosphocholine at 500 MHz. Each backbone amide proton except those of Glu^{70} and Glu^{72} gives rise to a single correlation identified by residue number; the signals for Thr^{87} and Arg^{96} overlap and are indicated by an asterisk. Correlations arising from side chain atoms are labeled with residue number and atom name. (**B**) NOE and chemical shift data identify secondary structure of the GpAtm peptide. NOE correlations indicative of α-helical [$d_{N\alpha(i-4)}$ and $d_{N\beta(i-1)}$] and β-strand [$d_{N\alpha(i-1)}$] configurations are denoted by filled boxes; correlations that cannot be unambiguously identified due to chemical shift degeneracies are denoted by open boxes. Chemical shift index analysis (*18*) was applied to the resonances of GpAtm: for H^{α} resonances, scores of +1 indicate a β strand, whereas scores of −1 indicate a region of α helix; for the C^{α} and carbonyl resonances, the scoring is reversed. Resonances with statistically insignificant secondary shifts received scores of 0. The consensus secondary structure is indicated at the bottom (h, helix; s, sheet). Abbreviations for the amino acid residues are as follows: A, Ala; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; and Y, Tyr.

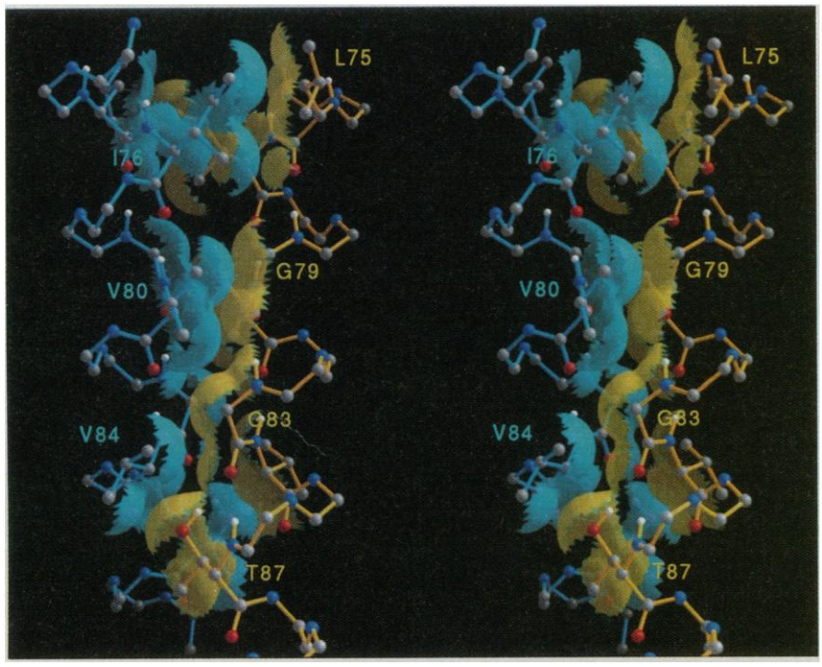


Fig. 2. Stereo view of the GpAtm dimer interface with the dyad axis aligned vertically. Main chain atoms of residues 73 to 88 and side chain atoms of interfacial residues are shown (gray, carbon; blue, nitrogen; red, oxygen; white, polar hydrogen). Molecular surfaces of each monomer, drawn at radii corresponding to the van der Waals hard sphere limit, are shown for those portions of the surfaces that lie within 1.4 Å of the other monomer. The detailed fit between the yellow "groove" and blue "ridge" residues in the front half of the dimer interface can be recognized by considering the complementarity of the semitransparent surface representations. Figures 2 and 3 were generated with the program GRASP (*19*).

actions at the dimer interface. Extensive backbone-backbone contact between the helices is made possible by two glycines separated by four residues: The side chains of Val^{80} and Val^{84} form a ridge that packs against the groove created by Gly^{79} and Gly^{83} of the opposite monomer. This groove is extended by the side chains of Leu^{75} and Thr^{87}, and the side chain of Ile^{76} extends the ridge. The symmetry-related Thr^{87} β methines and γ methyls interact with one another across the dyad. The Gly^{83} carbonyl oxygen forms a bifurcated hydrogen bond with the Thr^{87} hydroxyl and amide protons of its own monomer, as observed in many soluble proteins (*7*), and the Thr^{87} γ oxygen packs with the backbone and side chain of Val^{84}. GpA dimerizes by burying only a small interaction surface (400 $Å^2$) and without forming intermonomer hydrogen bonds or salt bridges.

The preponderance of glycines and β-branched amino acids in the GpA transmembrane domain contrasts markedly with the low frequencies at which these residues are found in α helices of soluble proteins (*8*). Theoretical considerations and experimental evidence (*6*, *9*) suggest that the stabilization energy provided by backbone hydrogen bonds within a low dielectric, hydrophobic environment can drive GpA helix formation despite the high proportion of residues with poor helix-forming tendencies (*10*). Interestingly, the GpAtm structure shows that the properties of these "unfavorable" residues can be used to generate specific interactions between transmembrane monomers. Because only one χ_1 rotamer is accessible to the β-branched amino acids isoleucine, valine, or threonine within an α helix (*11*), the positions of most GpA interfacial atoms within one monomer are determined by helix formation, and two largely preformed surfaces associate to form the dimer. This results in only a small loss of protein conformational entropy upon dimerization.

The closely packed dimer interface of GpAtm has side chain torsion angles displaced only slightly from the ideal values of the rotamers seen in α helices (*11*, *12*), indicating that the side chains are not strained by their participation in the interface. The enthalpic penalty associated with generating the side chain configurations needed to make good intermonomer packing interactions is therefore minimal. Although helix-lipid enthalpic interactions and lipid entropy presumably affect the monomer-dimer equilibrium, the relative smoothness of the dimer interface suggests that lipid chains would interact at least as favorably with the interfacial residues as with the other faces of monomeric GpA helices.

Mutagenesis studies of GpA identified a dimerization motif composed of Leu^{75}, Ile^{76}, Gly^{79}, Val^{80}, Gly^{83}, Val^{84}, and Thr^{87} (*3*,

13). In our structure, these residues contribute essentially all the atoms that participate in intermonomer van der Waals interactions, reinforcing the conclusions of the mutagenesis work and providing a structural rationale for the general ability of this motif to induce dimerization (*13*). The model proposed on the basis of simulated annealing searches (*14*) has the same features as the structure presented here; the difference for backbone atoms of residues 74 to 91 is only 0.8 Å (rms), showing that a combination of computational and mutagenesis methods can provide accurate models of transmembrane helix interactions. The excellent agreement between the conclusions of the indirect structural methods previously applied to GpA and the direct structure determination presented here should encourage further application and development of indirect approaches to the structural characterization of membrane proteins.

Modeling of disruptive hydrophobic substitutions (*3*) in the GpAtm structure through use of side chain rotamers commonly found in α helices (*11*) reveals that more than 90% of these mutations alter packing interactions between monomers. Steric clashes are found to be more disruptive than changes that eliminate favorable interactions, and clashes with backbone atoms are more disruptive than clashes with side chain atoms. For instance, the Gly^{79}→Ala substitution, which causes slight clashes that might be alleviated by small side chain movements, is mildly disruptive, whereas the Gly^{83}→Ala substitution results in a methyl-carbonyl oxygen clash and is completely disruptive (Fig. 3). The disruptive effects of substitutions Leu^{75}→Ile and Ile^{76}→Ala and the ability of these two nondimerizing proteins to form a heterodimeric complex (*3*) are consistent with the GpAtm structure, as are the results of alanine insertion mutagenesis (*15*). The qualitative agreement between mutagenesis data and steric considerations based on the wild-type structure reinforces the idea that helix-helix packing is the main determinant of GpA stability.

The GpAtm structure presented here demonstrates that van der Waals interactions alone can stabilize specific dimerization of membrane helices, but a full understanding of the structural basis for transmembrane helix association will require the analysis of many systems of varying size and complexity. Approaches such as the use of deuterated protein and quadruple resonance pulse sequences (*16*) and the availability of higher fields should make structure determination by solution NMR applicable to many detergent-solubilized single- and multispanning helical membrane protein systems.

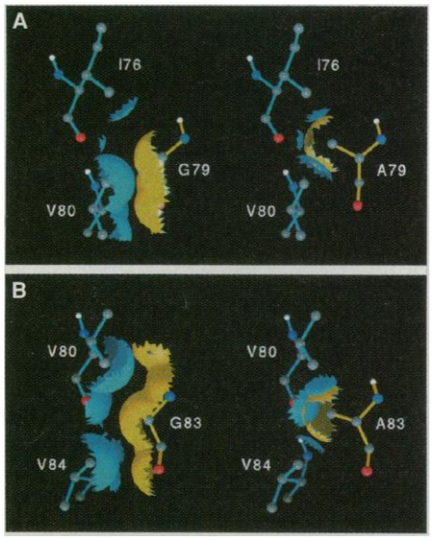


Fig. 3. Steric clashes generated by alanine substitutions at positions Gly^{79} and Gly^{83}. The left-hand panels portray the favorable van der Waals interactions in the wild-type structure through surface representations as in Fig. 2, whereas the right-hand panels depict the steric clashes incurred upon replacing glycine by alanine. (**A**) Replacement of Gly^{79} with Ala causes moderate methyl-methyl clashes; this sequence variant shows reduced amounts of dimer (*3*). (**B**) Replacement of Gly^{83} with Ala causes methyl-carbonyl and methyl-methyl clashes; this sequence variant shows no dimer (*3*). All other identified hydrophobic substitutions at these two sites completely disrupt dimerization (*3*).

REFERENCES AND NOTES

1. B.-J. Bormann, W. J. Knowles, V. T. Marchesi, *J. Biol. Chem.* **264**, 4033 (1989).
2. M. A. Lemmon *et al.*, *ibid.* **267**, 7683 (1992).
3. M. A. Lemmon, J. M. Flanagan, H. R. Treutlein, J. Zhang, D. M. Engelman, *Biochemistry* **31**, 12719 (1992).
4. The ^{15}N,^{13}C-labeled peptide was derived by proteolysis of a chimeric protein expressed in *Escherichia coli* grown in media containing ^{13}C-enriched glucose and ^{15}N-enriched ammonium chloride as sole C and N sources. Lyophilized peptide-detergent complex was dissolved in 90% H_2O–10% D_2O (pH 6.0) to a final concentration of 3.5 mM GpAtm monomer and 5% d_{38}-dodecylphosphocholine. NMR experiments were carried out at 40°C on either a GE Omega 500 spectrometer equipped with a Bruker triple-resonance triple-axis pulsed-field gradient probe or a Varian UNITY+600 spectrometer equipped with a triple-resonance, *z*-gradient probe. We achieved the sequential assignment of the backbone and side chain resonances at 500 MHz using through-bond correlations derived from three-dimensional (3D) HNCA [L. E. Kay, M. Ikura, R. Tschudin, A. Bax, *J. Magn. Reson.* **89**, 496 (1990)] and HCCH-total correlation spectroscopy [A. Bax, G. M. Clore, A. M. Gronenborn, *ibid.* **88**, 425 (1990)] experiments. Side chain dihedral restraints were derived from quantitative *J* couplings measured by means of 2D spin-echo difference constant-time heteronuclear single-quantum coherence spectra [G. W. Vuister, A. C. Wang, A. Bax, *J. Am. Chem. Soc.* **115**, 5334 (1993); S. Grzesiek, G. W. Vuister, A. Bax, *J. Biomol. NMR* **3**, 487 (1993); K. R. MacKenzie, J. H. Prestegard, D. M. Engelman, *ibid.* **7**, 256 (1996)]. Interproton distance restraints were derived from 3D ^{15}N- or ^{13}C- separated NOE spectra [E. R. P. Zuiderweg and S. W. Fesik, *Biochemistry* **28**, 6150 (1989); S. M. Pascal *et al.*, *J. Magn. Reson. Ser. B* **103** 197 (1994)] acquired at 600 MHz with mixing times of 20 ms. Six pairs of protons whose minimum intramonomer separations were found to exceed 5.5 Å due to the constraints of α-helical backbone geometry and *J* coupling–derived side-chain dihedral information, but which nevertheless exhibited intense NOE correlations, were determined as making intermonomer contacts. All other NOE restraints were treated as having arisen from either intra- or intermonomer cross-relaxation [M. Nilges, *Proteins* **17**, 297 (1993)]. A family of 20 structures was calculated with the program X-PLOR [A. T. Brünger, *X-PLOR Version 3.1, User's Manual*, Yale University (1992)] and a protocol combining distance geometry and simulated annealing [M. Nilges, G. M. Clore, A. M. Gronenborn, *FEBS Lett.* **229**, 317 (1988)]. No symmetry terms were used. A purely repulsive potential was used to limit the closest approach of nonbonded atoms; no attractive van der Waals or electrostatic terms were used. The 20 structures contain no bond length violations greater than 0.025 Å, no bond angle violations greater than 5°, no NOE distance restraint violations greater than 0.5 Å, and no dihedral violations greater than 5°. The average rms deviations for bond lengths, bond angles, impropers, NOE distances, and dihedrals were 0.002 Å, 0.379°, 0.239°, 0.070 Å, and 0.38°, respectively. Residues 74 to 91 of the family of structures are subject to an average of eight experimental restraints per residue and exhibit rms deviations from the mean structure of 0.75 Å for all nonhydrogen atoms and 0.40 Å for backbone atoms; the nonhydrogen atoms of the seven residues forming the dimerization interface have an rmsd of 0.50 Å. The structures have an average intermonomer Lennard-Jones (L-J) energy of −25 ± 7 kcal/mol; this L-J energy did not form part of the mixed target function. The experimental restraints and the atomic positions of the family of structures have been deposited at the Brookhaven Protein Databank (accession number 1AFO).
5. A. H. Ross, R. Radhakrishnan, R. Robson, H. G. Khorana, *J. Biol. Chem.* **257**, 4152 (1982); N. Challou, E. Goormaghtigh, V. Cabiaux, K. Conrath, J. M. Ruysschaert, *Biochemistry* **33**, 6902 (1994).
6. J.-L. Popot and D. M. Engelman, *Biochemistry* **29**, 4031 (1990).
7. T. M. Gray and B. W. Matthews, *J. Mol. Biol.* **175**, 75 (1984).
8. P. Y. Chou and G. D. Fasman, *Annu. Rev. Biochem.* **47**, 251 (1978).
9. S. C. Li and C. M. Deber, *J. Biol. Chem.* **208**, 22975 (1993).
10. K. T. O'Neil and W. F. DeGrado, *Science* **250**, 646 (1990).
11. R. L. Dunbrack and M. Karplus, *J. Mol.Biol.* **230**, 543 (1993).
12. J. W. Ponder and F. M. Richards, *ibid.* **193**, 775 (1987).
13. M. A. Lemmon, H. R. Treutlein, P. D. Adams, A. T. Brünger, D. M. Engelman, *Nature Struct. Biol.* **1**, 157 (1994).
14. P. D. Adams, D. M. Engelman, A.T. Brünger, *Proteins* **26**, 257 (1996).
15. I. Mingarro, P. Whitley, M. A. Lemmon, G. von Heijne, *Protein Sci.* **5**, 1339 (1996).
16. S. Grzesiek, J. Anglister, H. Ren, A. Bax, *J. Am. Chem. Soc.* **115**, 4369 (1993); T. Yamazaki, W. Lee, C. H. Arrowsmith, D. R. Muhandiram, L. E. Kay, *ibid.* **116**, 11655 (1994).
17. L. E. Kay, P. Keifer, T. Saarinen, *ibid.* **114**, 10663 (1992).
18. D. S. Wishart and B. D. Sykes, *J. Biomol. NMR* **4**, 171 (1994); _____ and F. M. Richards, *Biochemistry* **31**, 1647 (1992).
19. A. Nicholls, K. A. Sharp, B. Honig, *Proteins* **11**, 281 (1991).

20. We thank R. B. Hill, J. R. Tolman, I. T. Arkin, and M. A. Lemmon for assistance and encouragement and D. Jeruzalmi, J. L. Popot, C. E. Rogge, W. P. Russ, and S. C. Stallings for critical evaluations of the manuscript and figures. Pulse sequences provided by L. E. Kay are gratefully acknowledged. Supported by NIH grant P01 GM54160 and NSF grant MCB-9406983.

13 August 1996; accepted 4 February 1997