Long-distance conformational changes in a protein engineered by modulated sequence duplication

Martin Sagermann*, Leslie Gay, and Brian W. Matthews†

Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, OR 97403-1229

Contributed by Brian W. Matthews, June 10, 2003

There are few, if any, known instances in which a biological signal is transmitted via a large conformational change through the body of a protein. We describe here a mutant of T4 lysozyme that was engineered to permit structural change at a distance. The design uses a tandem sequence repeat that makes it possible to transmit large-scale structural changes from one end of an α-helix to the other over a distance of 17–25 Å. The method should be of general applicability and may make it possible to introduce a mutation at one site in a protein that will induce large-scale changes in the structure at a spatially remote site.

Point mutants in proteins typically result in small changes in structure that are localized to the immediate vicinity of the substitution (1–3). For example, in the Ala-129 → Trp mutant of T4 lysozyme (4), which was specifically introduced to disrupt the structure, the average backbone shifts were ≈1.2 Å close to the mutation site and fell to 0.4 Å, which was the "noise" level, at a distance of ≈13 Å. Typically, changes in structure are less than this and dissipate more rapidly (e.g., see refs. 4 and 5). [Because we are considering only structural changes that occur within a single domain of a protein we do not include cases in which one domain moves relative to another, e.g., because of "hinge-bending" (6) or domain swapping (7).] Even insertion mutants that incorporate one or more amino acids within the polypeptide chain typically show rather modest changes in the structure of a protein (8–11). {The triple alanine insertion mutant 73[AAA] in T4 lysozyme does show somewhat larger-scale reorganization (11).} This may explain why there are few, if any, examples in which a biological signal is transmitted over a substantial distance (e.g., 20 Å) by a series of linked conformational changes within a protein. Such long-distance effects typically result from bodily movements of whole domains, whole subunits, or even whole proteins. The classical example is hemoglobin in which allosteric effects are mediated by changes in the alignments of the α- and β-chains, but with the conformations of the individual subunits largely preserved (12). Another common mechanism by which signals are transduced is ligand-induced dimerization (or oligomerization) of receptor molecules (13–15). Order-to-disorder transitions can also play a role in some cases (16).

We describe here a mutant of T4 lysozyme in which amino acid substitutions in one region of the protein result in large-scale conformational changes 17–25 Å away, these changes being transmitted through the body of the protein. The changes are facilitated by the incorporation of a tandemly duplicated sequence segment and illustrate a general approach that might be used to transmit long-distance signals through proteins.

The underlying idea behind the experimental design can be understood from consideration of Fig. 1. Fig. 1 shows the backbone structure of a mutant lysozyme, identified as L20, in which the sequence corresponding to helix B (residues Leu-39–Ile-50) is duplicated in tandem (17). In the resultant structure (Fig. 1) the "original" residues 39–50 (colored yellow in Fig. 1) form an α-helix virtually identical to that in the WT lysozyme structure. The "duplicated" residues (colored red in Fig. 1 and distinguished with the letter i for inserted) extend the N terminus of the helix B by approximately two turns before turning back to connect with the rest of the molecule at Ser-38.

At the N terminus of the extended helix B, i.e., in the vicinity of residues Leu-39i–Lys-43i (Fig. 1), the inserted residues are somewhat solvent exposed and form relatively weak interactions with each other and the rest of the helix (17). Likewise, in WT lysozyme the residues adjacent to the N terminus of helix B do not appear to participate in strong interactions (although Ser-38 has been shown to have a favorable helix-capping interaction) (18, 19) and to be important in helix stabilization (20).

In contrast, the residues adjacent to the C terminus of helix B (Arg-52–Asn-53–Thr-54–Asn-55) are involved in a number of specific stabilizing salt bridge, hydrogen-bonding, and hydrophobic interactions (21). It seems very likely that it is these stabilizing interactions that maintain the conformation of the C terminus of helix B in the tandem repeat structure and cause the helix to extend at its N terminus (Fig. 1). We also reasoned that the elimination of these favorable interactions at the C terminus of helix B might allow the helix to extend at its C terminus. Such an extension would cause a large change in conformation to switch from one end of the α-helix to the other.

To test this idea residues 52–55 were mutated to glycine. The absence of a side chain and the lack of conformational preferences of this amino acid not only abolishes all former electrostatic interactions but also avoids the introduction of additional helix-forming propensity that would be associated with a poly-alanine replacement. Because residues 51 and 56 are glycine in WT lysozyme this results in a string of six glycines. This mutant (L20pg, L20-polyglycine) was made in the context of the tandem duplication mutant L20 (Fig. 1).

Materials and Methods

Starting with the gene for the tandem repeat lysozyme L20 (17) mutant L20pg was created by standard PCR methods using the primers GAA TTA GAT AAA GCT ATT GGG GGT GGC GGT AAT GGT GTA ATT ACA AAA G and CTT TTG TAA TTA CAC CAT TAC CGC CAC CCC CAA TAG CTT TAT CTA ATT C. The product of the PCR was subsequently cloned into a pHS1403 expression vector by using the *Bam*HI and *Hin*dIII restriction sites. Upon expression most of the protein was found in inclusion bodies and was purified as follows (cf. refs. 11 and 22). After induction, cells were harvested and lysed by using the French press. After addition of DNase I the homogenate was centrifuged for 30 min at 37,000 × *g*. The resulting pellet was resuspended in 20 ml of 50 mM Tris·glycine buffer (pH 8.0) 50 mM NaCl, 10 mM EDTA, 1 tablet Boehringer protease inhibitor mixture, 1 mM DTT, and 10% Triton X-100 for 2 h at 4°C. The resuspension was then centrifuged at 37,000 × *g* for 20 min. The resulting pellet was subsequently resuspended in 10 vol of 2.5% β-octyl glucoside for 2 h at 4°C. A subsequent centrifugation at 37,000 × *g* for 2 h resulted in a white pellet that was

Abbreviation: L20pg, L20-polyglycine.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID code 1OYU).

*Present address: Department of Chemistry and Biochemistry, 1631 Physical Science North, University of California, Santa Barbara, CA 93106.

†To whom correspondence should be addressed. E-mail: brian@uoxray.uoregon.edu.

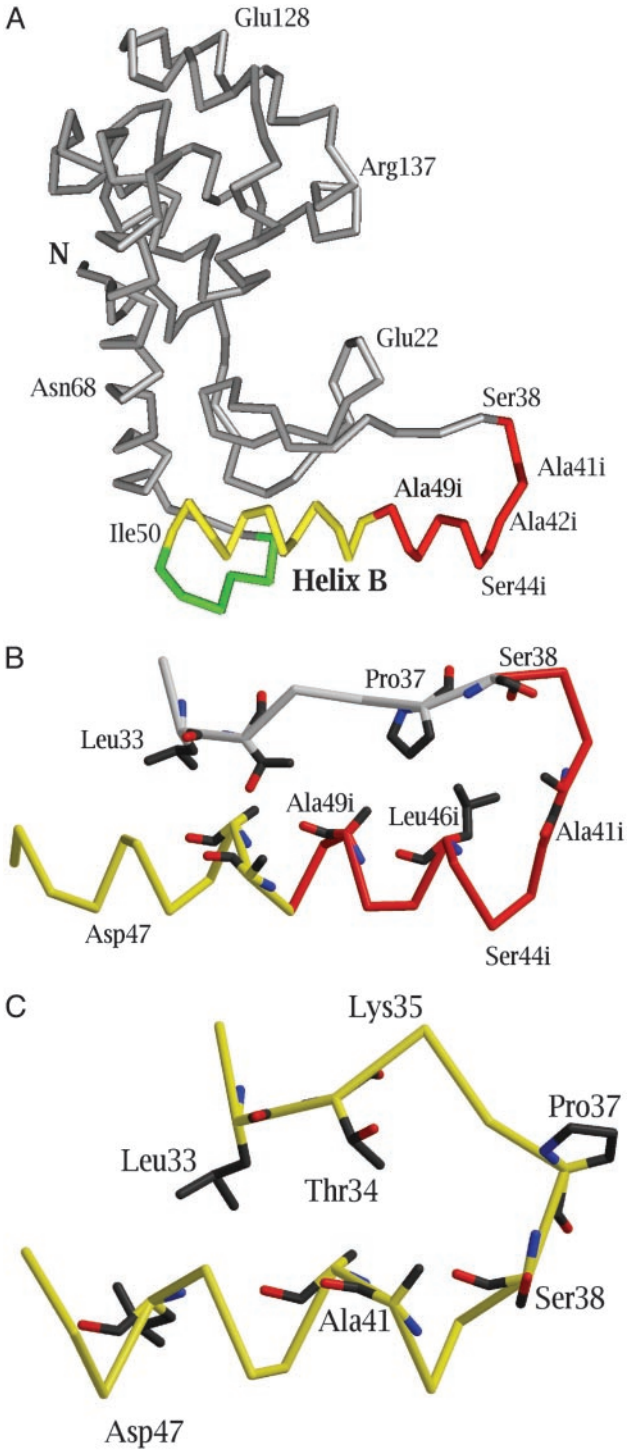

Fig. 1. (*A*) Backbone representation of the structure of T4 lysozyme mutant L20 in which residues 39–50 are duplicated in tandem (17). The original residues 39–50 (yellow) have a conformation similar to helix B of WT. The duplicated residues (red) extend helix B at its N terminus. Residues 51–56 are green. Their conformation is the same as WT and helps stabilize the conformation at the C terminus of helix B. *A* and Fig. 2 were created with PYMOL (Warren DeLano, DeLano Scientific, San Carlos, CA). (*B*) The details of the conformation of the N-terminal extension of helix B as seen in the structure of L20 (*A*). Most of the contacts within the extension are hydrophobic in nature and, for clarity, only the side chains that contribute to these interactions are shown. The color-coding of the backbone is the same as in *C*. [*B* and *C* and Figs. 3 and 4 were made with MOLSCRIPT and RASTER3D (27).] (*C*) Conformation at the N terminus of helix B in WT (compare *B*). As in *B*, only those side chains that make contacts within the loop are shown.

Table 1. X-ray data collection and refinement

Cell dimensions, Å	$a = b = 60.35$, $c = 213.90$
Space group	$P4_12_12$
Resolution, Å	2.5
Completeness, %	94.8
Unique reflections	12,441
R_{sym}, %	7.2
R_{cryst}, %	20.9
R_{free}, %	32.4
Δ_{bond}, Å	0.010
Δ_{angles}, °	1.6

resuspended and washed twice in deionized water. The washed pellet was then again resuspended in 20 ml glycine buffer and 4 M urea. Subsequently, phosphoric acid was added to pH 3.0. This solution was then centrifuged for 20 min at 37,000 × *g*. The supernatant was then dialyzed for 2 h in 50 mM citrate buffer (pH 3.0) and 10% glycerol and citrate buffer (pH 5.5) and 10% glycerol, respectively. At this point any aggregated material was centrifuged out, and the resulting supernatant dialyzed against citrate buffer (pH 3.0) and 10% glycerol and finally against 50 mM Tris (pH 7.5) and 100 mM NaCl. The protein was then concentrated by using Amicon and Centriprep protein concentration devices.

Crystals were grown at 4°C with 25% (wt/vol) polyethylene glycol 4000, 20% (vol/vol) isopropanol, 50 mM phosphate buffer (pH 5.5), and 0.2 mM ammonium acetate. The crystals grew to plates of ≈0.2 mm × 0.2 mm × 0.1 mm within 7–10 days. The structure was determined by molecular replacement (23) and refined by using TNT and CNS (24, 25, 28). Refinement statistics are given in Table 1. The coordinates have been deposited in the Protein Data Bank (ID code 1OYU).

Results

Mutant L20pg was constructed, purified, and crystallized, and its structure was determined by molecular replacement (see *Materials and Methods*). There are two molecules, A and B, in the asymmetric unit, and their structures are illustrated in Fig. 2 *A*–*C*. For both molecules the structure in the vicinity of the N terminus of helix B does not retain the conformation seen in the original tandem duplication mutant L20 (Fig. 1 *A* and *B*). Rather, the conformation in this region is virtually identical with WT (compare Figs. 2 *A* and *B* with 1*C*). It might also be noted that in the crystal there is a neighboring molecule close to the N terminus of helix B (Fig. 2*A*). If the N-terminal region were to extend (as in Fig. 1*A*), it would immediately result in steric clashes with this neighboring molecule.

At the C terminus of helix B the helical conformation is extended by approximately one turn. The electron density is somewhat clearer for molecule B (Fig. 2*B*) but even here the density between residues Ala-42 and Gly-56 is weak (Fig. 2*C*). The density for this loop region in molecule A is very ill defined. At the end of this segment, however, the electron density suggests that residues 56–59 retain the same conformation as WT, as does the following α-helix (helix C, residues 60–79) and the rest of the lysozyme molecule.

Discussion

The structural results show that the experimental design was successful. By removing the interactions that stabilize the loop at the C terminus of helix B the residues in the original helix (Leu-39–Ile-50) are replaced by the tandemly repeated residues. In the structure of L20 (Fig. 1*A*) there is an extension plus other structural changes at the N terminus of helix B. In contrast, in the structure of L20pg helix B extends at its C terminus, together with associated changes in the vicinity. In other words, the

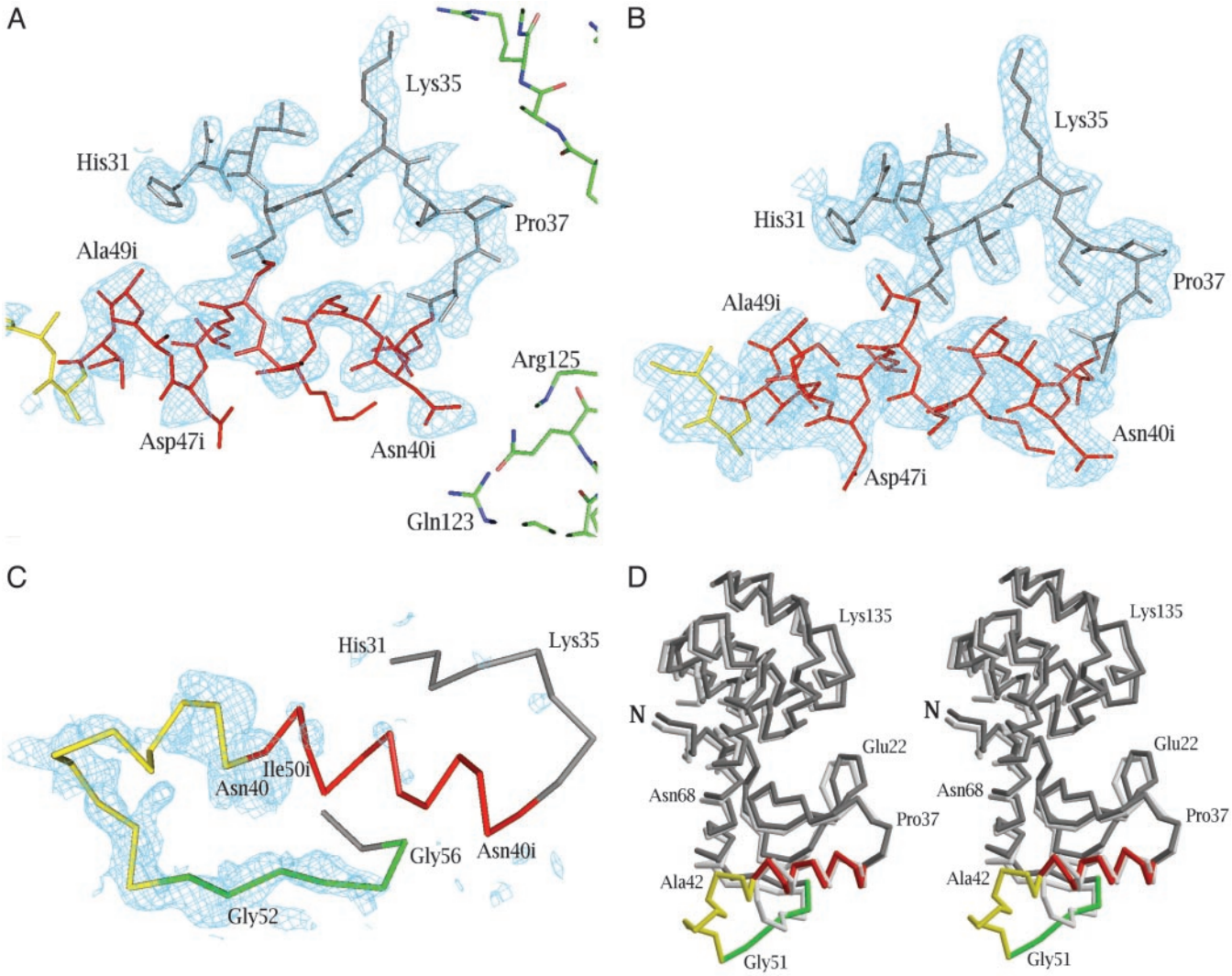

Fig. 2. (*A*) Simulated annealing omit electron density for molecule A of mutant L20pg in the vicinity of helix B. The region at the N terminus of the helix (near Pro-37 and Asn-40i) has a conformation virtually identical with WT. At the C terminus (near Ile-50i) the helix extends by approximately two turns at which point the electron density becomes poorly defined. The elements of structure shown at the right and colored in green are from a neighboring molecule in the crystal. Coefficients for the electron density are $F_o - F_c$, where F_o is the observed amplitudes for L20pg and F_c is calculated from the refined model of the structure with residues His-31–Ala-41 deleted. The resolution is 2.5 Å, and the map is contoured at 2.2 σ, where σ is the average density throughout the unit cell. (*B*) Simulated annealing omit electron density in the vicinity of helix B for molecule B of L20pg. All procedures are the same as in *A* except that the map is contoured at 2.5 σ. In this case the region in the vicinity of the N terminus of helix B is free of crystal contacts (compare *A*). (*C*) Simulated annealing omit map calculated at 2.5-Å resolution and contoured at 2.3 σ above the mean. Shown is the electron density of molecule B with residues Asn-40–Gly-54 omitted from the density calculation. Only the C^a backbone trace is shown with the same color coding as in *D*. The density for the omitted residues is weak, and no reliable side-chain density could be identified. The final polyalanine model of this loop was refined to an occupancy of 0.6. The loop structure is somewhat stabilized by a contacting symmetry related molecule (data not shown). The corresponding loop in molecule A is not detectable. (*D*) Superposition of the backbone structure of mutant L20pg on that of WT. In the mutant the tandem repeat is shown in red and yellow and the polyglycine region is shown in green. The remainder of the structure is in dark gray. The backbone structure of WT lysozyme is shown in light gray.

replacement of residues 52–55 by glycine, in the context of the tandem repeat structure, has caused a major change in structure to switch from one end of helix B to the other. The distance between the respective regions of structural change (Fig. 3) is ≈25–40 Å. The distance from the polyglycine substitutions to the remote region at which they cause conformational change is ≈20–30 Å.

The results clearly suggest that the conformation of the structure with the tandem duplication of the sequence corresponding to helix B is determined by competing interactions at the ends of the helix. If the stabilizing interactions are stronger at the N terminus of the helix, then the helix will extend at its C terminus. Conversely, if the stabilizing interactions are stronger in the vicinity of the C terminus then the extension will occur at the N terminus. This finding suggests that it might be possible to modulate these competing interactions to control the conformation of the molecule, provided the energy barrier between the two states is not too great.

In the present case, the loop structure that connects helix B with helix C in WT T4 lysozyme contains 10 residues. A number of these are involved in electrostatic and hydrophobic interactions. Arg-52, for example, forms a salt bridge with Glu-62 as well as hydrogen bonds with the backbone carbonyls of Asn-53 and Val-57 (Fig. 4). Truncation of the side chain of Arg-52 to valine results in a 7.1°C-decrease in the melting temperature (26), confirming that this residue contributes significantly to stability.

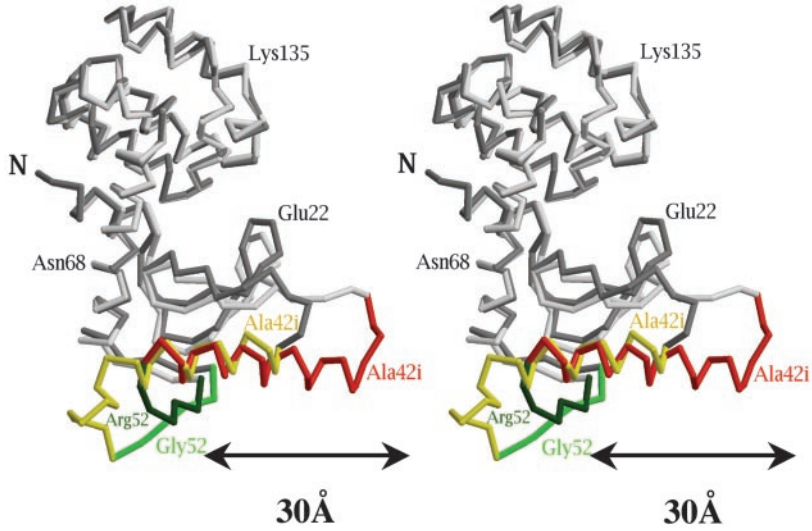

Fig. 3. Superposition of L20pg on L20. The two tandem repeat sequences are yellow and red, respectively. The corresponding C-terminal loop residues 51–56 are green and dark green. As representative examples, the alternative positions of residues 42 and 52 are labeled.

Arg-52 was one of four residues in L20pg that were replaced with glycine. If the arginine side chain was restored, and/or other residues in L20pg were substituted by glycine, it should be possible to restore stability and cause L20pg to switch such that conformation at the C terminus of helix B changes to that of WT. Conversely, there are also residues close to the N terminus of helix B that might be changed to stabilize or destabilize the conformation in this vicinity. For example, Ser-38 has been shown to be of particular importance (20) in stabilizing helix B. It is positioned at the amino terminus of the helix with its γ-hydroxyl-accepting hydrogen bonds from the amide groups in the first turn of the helix. When Ser-38 is replaced by alanine the enzyme is destabilized (20), whereas replacement with aspartic acid causes a slight increase in stability (18, 19). Thus, by making substitutions at one end of the helix or the other (e.g., at residue 38 at the N terminus or residue 52 at the C terminus) it should be possible to "fine-tune" the stability of one region relative to the other and switch between two alternative conformations essentially at will. Because some of the relevant interactions are electrostatic in nature (e.g., the salt bridge between Arg-52 and Glu-62) it might be possible to engineer a pH-dependent conformational switch.

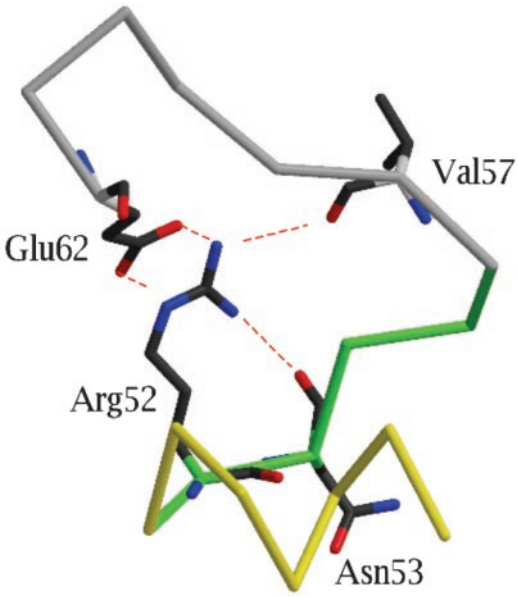

Fig. 4. Interactions that appear to stabilize the loop at the C terminus of helix B. Only the side chains that have core contacts within the loop are shown. The conformation shown occurs in WT lysozyme and L20. When residues 51–56 (green) are replaced with glycine a completely different conformation results.

The idea of using sequence duplication to induce a long-range conformational switch should, in principle, be applicable to other proteins. It also need not be restricted to α-helical regions. By incorporating a sequence duplication it facilitates the formation of two distinct conformational states. In the first conformational state the "parent" sequence mimics the WT structure, whereas the "duplicate" sequence is forced to adopt some other conformation. In the alternative conformational state the duplicate sequence mimics the WT, whereas the parent sequence has a different conformation. By adjusting the strength of interactions that occur at the beginning and the end of the duplicated region it may be possible to choose between the two conformational states. It may also be possible to use a single amino acid substitution at one site in a protein to induce a large-scale conformational change at a spatially remote site. In the case of T4 lysozyme, helix B is located on the surface of the protein. At each end of the helix there is a loop that is somewhat solvent-exposed. Presumably it is the presence of these relatively solvent-exposed regions that allow the helix to extend at either its N or C terminus (Fig. 3). Helices (or β-sheet strands) bounded by solvent-exposed loops are, however, quite common in proteins. This finding suggests that the procedure used here may be applicable in other situations.

We thank Dr. Walter Baase for helpful discussions and the staff at the Stanford Synchrotron Radiation Laboratory, Beamline 7.1, for their expert help. This work was supported in part by National Institutes of Health Grant GM21967 (to B.W.M.).

1. Matthews, B. W. (1987) *Biochemistry* **26,** 6885–6888.
2. Matthews, B. W. (1993) *Annu. Rev. Biochem.* **62,** 139–160.
3. Richards, F. M. & Lim, W. A. (1994) *Q. Rev. Biophys.* **26,** 423–498.
4. Liu, R., Baase, W. A. & Matthews, B. W. (2000) *J. Mol. Biol.* **295,** 127–145.
5. Dao-pin, S., Alber, T., Baase, W. A., Wozniak, J. A. & Matthews, B. W. (1991) *J. Mol. Biol.* **221,** 647–667.
6. Zhang, X.-J., Wozniak, J. A. & Matthews, B. W. (1995) *J. Mol. Biol.* **250,** 527–552.

7. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. (1995) *Protein Sci.* **4,** 2455–2468.
8. Heinz, D. W., Baase, W. A., Dahlquist, F. W. & Matthews, B. W. (1993) *Nature* **361,** 561–564.
9. Heinz, D. W., Baase, W. A., Zhang, X.-J., Blaber, M., Dahlquist, F. W. & Matthews, B. W. (1994) *J. Mol. Biol.* **236,** 869–886.
10. Keefe, L. J., Quirk, S., Gittis, A., Sondek, J. & Lattman, E. E. (1994) *Protein Sci.* **3,** 391–401.
11. Vetter, I. R., Baase, W. A., Heinz, D. W., Xiong, J.-P., Snow, S. & Matthews, B. W. (1996) *Protein Sci.* **5,** 2399–2415.
12. Perutz, M. F. (1989) *Q. Rev. Biophys.* **22,** 139–236.
13. de Vos, A. M., Ultsch, M. & Kossiakoff, A. A. (1992) *Science* **255,** 306–312.
14. Lemmon, M. A. & Schlessinger, J. (1994) *Trends Biochem. Sci.* **19,** 459–463.
15. Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H.-J., Walker, F., Frenkel, M. J., Hoyne, P. A., *et al.* (2002) *Cell* **110,** 763–773.
16. Spoerner, M., Herrmann, C., Vetter, I. R., Kalbitzer, H. R. & Wittinghofer, A. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4944–4949.
17. Sagermann, M., Baase, W. A. & Matthews, B. W. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 6078–6083.
18. Nicholson, H., Becktel, W. J. & Matthews, B. W. (1988) *Nature* **336,** 651–656.
19. Nicholson, H., Anderson, D. E., Dao-pin, S. & Matthews, B. W. (1991) *Biochemistry* **30**, 9816–9828.
20. Zhang, X.-J., Baase, W. A. & Matthews, B. W. (2002) *Biophys. Chem.* **101–102**, 43–56.
21. Weaver, L. H. & Matthews, B. W. (1987) *J. Mol. Biol.* **193,** 189–199.
22. Gassner, N. C., Baase, W. A., Lindstrom, J. D., Shoichet, B. K. & Matthews, B. W. (1997) in *Techniques in Protein Chemistry VIII*, ed. Marshak, D. (Academic, New York), pp. 851–863.
23. Navaza, J. (1994) *Acta Crystallogr. A* **50**, 157–163.
24. Tronrud, D. E. (1996) *J. Appl. Crystallogr.* **29,** 100–104.
25. Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987) *Acta Crystallogr. A* **43**, 489–503.
26. DuBose, R. F., Baase, W. A., Zhang, X.-J., Xu, J. & Matthews, B. W. (1999) in *Perspectives in Structural Biology: A Volume in Honor of G. N. Ramachandran*, eds. Vijayan, M., Yathindra, N. & Kolaskar, A. S. (Indian Academy of Sciences and the University Press, Hyderbad, India), pp. 139–152.
27. Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524.
28. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.