Use of sequence duplication to engineer a ligand-triggered, long-distance molecular switch in T4 lysozyme

Mohammad S. Yousef*, Walter A. Baase, 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 24, 2004

We have designed a molecular switch in a T4 lysozyme construct that controls a large-scale translation of a duplicated helix. As shown by crystal structures of the construct with the switch on and off, the conformational change is triggered by the binding of a ligand (guanidinium ion) to a site that in the wild-type protein was occupied by the guanidino head group of an Arg. In the design template, a duplicated helix is flanked by two loop regions of different stabilities. In the ''on'' state, the N-terminal loop is weakly structured, whereas the C-terminal loop has a well defined conformation that is stabilized by means of nonbonded interactions with the Arg head group. The truncation of the Arg to Ala destabilizes this loop and switches the protein to the ''off'' state, in which the duplicated helix is translocated ≈20 Å. Guanidinium binding restores the key interactions, restabilizes the C-terminal loop, and restores the ''on'' state. Thus, the presence of an external ligand, which is unrelated to the catalytic activity of the enzyme, triggers the inserted helix to translate 20 Å away from the binding site. The results illustrate a proposed mechanism for protein evolution in which sequence duplication followed by point mutation can lead to the establishment of new function.

The ability to create and manipulate ligand-induced conformational changes is one of the major challenges in protein engineering and biotechnology. This ability demands a detailed understanding of the interplay between binding, structure, dynamics and energetics (1, 2). Several steps have been made toward developing possible ''nanoallostery'' modules, many cases of which have used a protein template already known to change conformation upon ligand binding. The approaches involve either mutating residues so that one state is preferentially stable over others (3–7), manipulating the binding specificity so that unnatural ligands can bind (8–10), or fusing the template to another protein that can sense the signal (11–13). Such approaches are subject to the structural and functional limitations imposed on the template during evolution. In an experiment notable for the use of different templates (14), allosteric switching was observed when two proteins (ubiquitin and barnase), neither of which undergoes conformational changes in its native form, were fused in such a way that the folding of one protein unfolded the other. However, the lack of a regulatory site that bound ligand(s) limited the ability to switch between the two states.

In the present report, a nanostructural module was added to T4 lysozyme. Part of the module, a duplicated secondary structure element, switches conformation upon binding a ligand, in this case guanidinium ion. The ligand is unrelated to the function of the protein, but its binding induces a large-scale conformational change.

The reference protein on which the design is based is designated L20 and has been described in ref. 15. In this protein, residues 40–50, corresponding to the B helix of T4 lysozyme, are duplicated in tandem. In the crystal structure, the inserted helical segment extends the ''parent'' helix at its N terminus and leaves the C terminus intact. The conformation at the C terminus appears to be stabilized by a loop that includes Arg-63-Asn-64-Thr-65-Asn-66. In the related mutant L20-polyglycine, this C-terminal loop was weakened by mutating residues in this loop to glycines (16). As a consequence, the residues within helix B translocated ≈20 Å toward its C terminus. In this case, the conformation at the N terminus of the helix is the same as in WT.

Materials and Methods

Cloning, Protein Purification, and Crystallization. Starting with the gene for the mutant L20 (15), the construct L20/R63A was created by QuikChange site-directed mutagenesis protocol (Stratagene) by using the internal primers 5′-GAA TTA GAT AAA GCT ATT GGG GCT AAT ACT AAT GGT GTA ATT ACA and 5′-TGT AAT TAC ACC ATT AGT ATT AGC CCC AAT AGC TTT ATC TAA TTC. The Arg to Ala mutation was confirmed by sequencing, and the product of the PCR (the mutated vector) was transformed into Xl1-Blue genetic strain and subsequently into the RR1 expression system (17) in *Escherichia coli*. Cells were grown at 37°C to high density. After induction, the protein was predominantly obtained in inclusion bodies by increasing the temperature to 42°C. The protein was then purified by using the standard protocol for T4 lysozyme (18, 19) and dialyzed against 100 mM sodium phosphate (pH 6.5)/500 mM NaCl/0.02% NaN_3. For the liganded protein, crystals in space group $P3_221$ were grown at 4°C by hanging-drop vapor diffusion, equilibrating against ≈1.8 M mixed potassium and sodium phosphate (pH 6.5) and in the presence of 0.2 M guanidinium chloride. For the drop, a 15 mg/ml protein solution was mixed 1:1 with the reservoir precipitant solution. The crystals grew to 0.5 mm × 0.5 mm × 0.3 mm within 2–4 days. For the unliganded form, the protein was further dialyzed against 50 mM Tris (pH 7.5) and 100 mM NaCl and concentrated to 20 mg/ml. Crystals were grown at 16°C with 30% (wt/vol) polyethylene glycol 3400, 100 mM Hepes buffer (pH 7.5), and 200 mM ammonium acetate. The crystals grew to thin plates of ≈0.2 mm × 0.2 mm × 0.1 mm within 15–21 days.

X-Ray Data Collection and Model Refinement. For the liganded crystals, cryocrystallographic data were collected on a Rigaku (Tokyo) R-AXIS-II image plate detector. The crystals were flash-cooled to 100 K in a nitrogen stream with 20% glycerol added to the crystallization mother liquor as a cryoprotectant. For the unliganded form, diffraction data were collected at Advanced Light Source (Beamline 8.3.1, λ = 1.0 Å) at 100 K, and 20% glycerol was added as a cryoprotectant. Data were integrated and scaled by using the HKL suite of programs DENZO, XDISPLAYF, and SCALEPACK (20). The structures of both the liganded and the unliganded proteins were solved by molecular

Data deposition: Atomic coordinates and structure factor files have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 1T8A and 1T97 for the liganded and unliganded forms, respectively).

*On leave from: Biophysics Department, Faculty of Science, Cairo University, Giza, Egypt.

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

Table 1. X-ray data collection and refinement statistics for L20/R63A

	Liganded form	Unliganded form
Data processing		
Space group	$P3_221$	$P2_1$
Unit cell dimensions		
a, b, and *c,* Å	$a = b = 60.4$, $c = 96.2$	$a = 53.9$, $b = 54.2$, $c = 58.0$
γ (liganded) and β (unliganded), °	$\gamma = 120$	$\beta = 103.9$
No. of monomers*	1	2
Resolution range (outer shell), Å	30–2.0 (2.07–2.0)	30–2.7 (2.8–2.7)
No. of observations	149,998	118,498
Unique reflections	13,956	11,359
Completeness (outer shell), %	99 (99)	99 (99)
R_{sym} (outer shell), %	4.6 (21.4)	6.7 (15.0)
$I/\sigma(I)$ (outer shell)	33 (5)	16 (7)
Structural refinement		
Resolution range, Å	30–2.0	30–2.7
R_{work}/R_{free}, %	21.8/25.1	23.6/28.9
rms deviation for bond length†, Å/angles, °	0.005/1.0	0.007/1.3
Distribution of (ϕ,ψ) for favored region, % (disallowed)	93 (0)	87 (0)
Mean *B* value		
Protein, Å^2	31	38
Guanidinium, Å^2	36.3	—
Water, Å^2 (*n*)	35 (131)	36 (82)
Amino acids with alternate conformations	6	0

—, not determined.
*The number of monomers represents those per asymmetric unit.
†The rms deviation for bond lengths and bond angles represents the rms deviations between the observed and ideal values.

replacement (21) and refined by CNS (22). CNS was also used for the calculations of thermal factor and accessible surface area profiles. A summary of the data processing and refinement statistics is given in Table 1. The refined structure of the liganded form includes the guanidinium ion, a chloride ion, and a molecule of the oxidized form of β-mercaptoethanol, present in the crystallization solution. The unliganded form was crystallized with two molecules per asymmetric unit related by a noncrystallographic twofold axis. For monomer A, there was no visible electron density map for residues 54–61; therefore, they were not modeled. Monomer B is complete and was used for the analysis shown in the figures.

Thermal Analysis. Thermal stabilities were determined in 0.1 M NaCl/10 mM NaOAc, pH 5.4 as described in ref. 23. Thermal denaturation experiments to compare the effect of 0.2 M KCl with that of 0.2 M guanidinium chloride were done in 1.9 M $Na_{0.55}K_{0.9}H_{1.55}PO_4$ (pH 6.5) buffer. Unfolding was irreversible in this buffer. Transition temperature increments were determined by direct overlay of unfolding curves in KCl versus guanidine hydrochloride for identical concentrations of protein, here 0.015 mg/ml.

Results and Discussion

The motivation behind the current design was to modify the C-terminal loop so that its stability would depend on the binding of a ligand, which, in turn, would trigger a switch between the two conformations.

Inspection of the C-terminal loop (Fig. 1*a*) suggests that it is stabilized primarily by multiple interactions with the guanidino head group of Arg-63. We therefore substituted Ala at this site to obtain the mutant L20/R63A. We reasoned that guanidinium

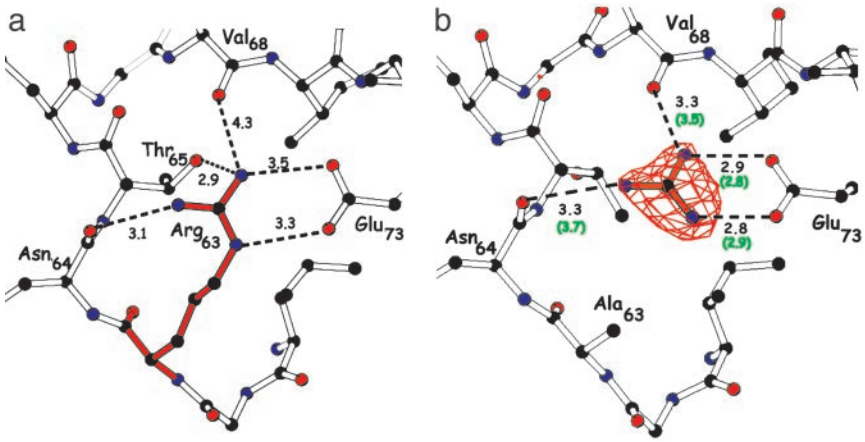

Fig. 1. Details of the interactions that stabilize the loop at the C terminus of the duplicated helix. (*a*) L20 (the design template). (*b*) L20/R63A in the presence of guanidinium. Distances (black) are shown in Å; in green are the corresponding distances in the WT structure. The superimposed $F_o - F_c$ difference map contoured at 3.3 σ (red) defines the position of the ligand.

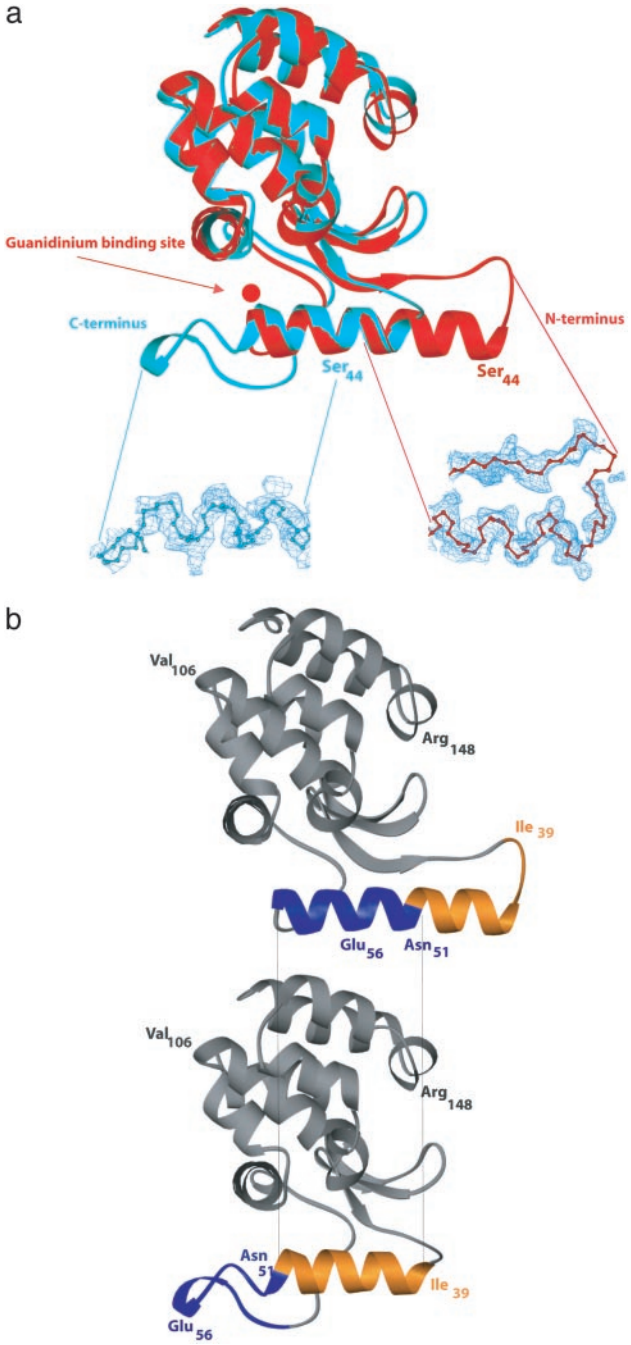

Fig. 2. (*a*) Superposition of liganded (red) on the unliganded (cyan) forms of L20/R63A. As representative examples, the alternative positions of Ser-44 are labeled. On the lower left and right are simulated-annealing omit maps (contoured at 1.1 σ) with backbone representations of the helix extended in both directions. (*b*) Detailed sketch showing the structures of the liganded (*Upper*) and the unliganded (*Lower*) forms. The ''inserted'' residues (Asn-40-Ile-50) are colored orange, and the ''parent'' residues (Asn-51-Ile-61, renumbered because of the 11-residue insert) are colored blue. The vertical bars connecting the two structures show the location of helix B in WT. In the presence of the guanidinium ion (*Upper*), the inserted helix (in orange) extends at its N terminus. In the absence of the ion (*Lower*), the inserted sequence occupies the position of helix B and the parent sequence extends the helix at its C terminus.

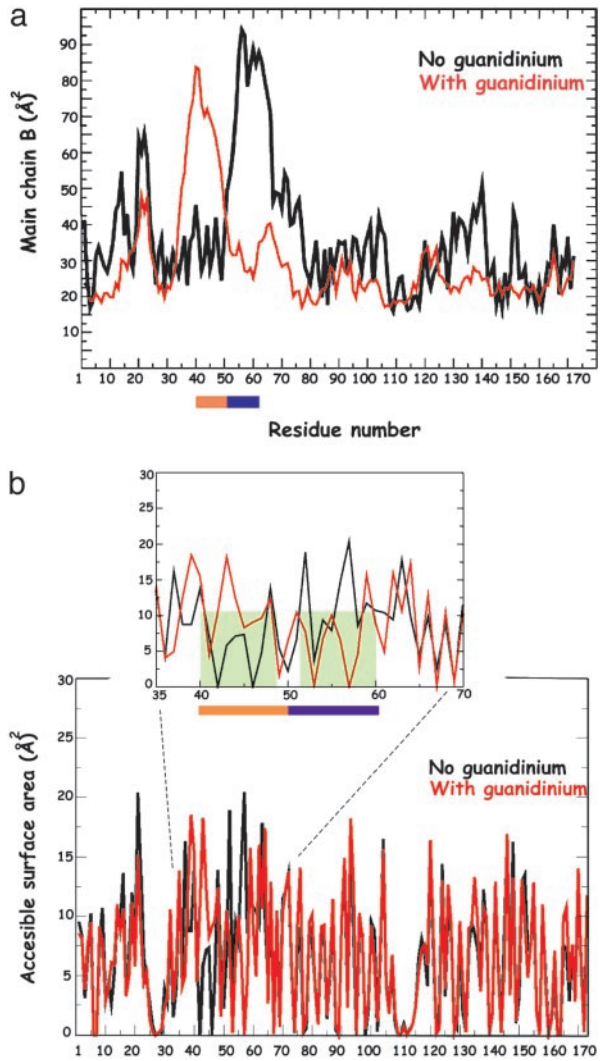

Fig. 3. (*a*) Thermal factor profiles for the liganded (red) and the unliganded form (black). Some of the differences shown are presumably due to the different resolutions and different crystal packing. The most dramatic differences, however, are in the vicinity of the duplicated helix. The orange and blue bars indicate the duplicated sequence. (*b*) Comparison of the residue-accessible surface-area profiles of the liganded (red) and unliganded (black) structures. The orange and blue bars indicate the duplicated sequence. For parts of the protein away from the region of duplication, the two profiles are essentially identical. Within the region of duplication there are major differences highlighted in *Inset* (see text for discussion).

ion might act as a surrogate for the head group of Arg-63 (24). The stability of L20/R63A is reduced by 6.1°C [1.8 kcal/mol (1 cal = 4.184 J)] relative to L20, confirming that Arg-63 does contribute to the stability of the protein. Also, in a high-salt buffer similar to that used for crystallization (see *Materials and Methods*), the melting temperature of L20/R63A is increased by 1.7°C in the presence of 0.2 M guanidinium hydrochloride, confirming the binding of the ion. This stabilizing effect is observed notwithstanding that guanidinium at high concentrations is routinely used as a protein denaturant.

The structural results confirm that the design was successful. When L20/R63A was crystallized in the presence of 0.2 M guanidinium, the structure clearly showed the presence of the bound ion (Fig. 1*b*). The site of binding for the guanidinium ion and its interactions with the surrounding protein closely match those of the Arg head group. Using such a truncation mutation (Arg to Ala) satisfies the steric and physicochemical complementarily between the ligand and the binding site. Moreover, in the guanidinium-bound structure, the duplicated α-helix was

extended at its N terminus as in the parent mutant L20. In contrast, when L20/R63A was crystallized in the absence of guanidinium, the duplicated α-helix extended in the opposite direction (i.e., as in L20-polyglycine). In both structures, more than half of the inserted sequence adopts a helical conformation (about two turns) before looping out to connect to the rest of the protein (Fig. 2). The presence or absence of the ligand determines the choice between the two conformations of the helical repeat. The distance from the binding site to the most distal part of the altered structure is $\approx$25 Å. The switch is triggered by the balance between the stabilizing forces at the ends of the duplicated helix (25). Ligand binding modulates these competing interactions and controls the conformation of the molecule.

As the protein switches between the on and off states, there is a major change in the B-factor profiles (Fig. 3*a*). Residues 51–65 are ordered when the guanidinium is bound. This ordering correlates with the stabilization of the C-terminal loop. Residues 35–45 on the N-terminal loop are ordered only in the unliganded form when the helix extends toward its N terminus. In the liganded form (Fig. 3 *a* and *b*, red trace), residues 40–50 extend helix B at its N terminus and are less well ordered than the remainder of the molecule. In contrast, in the nonliganded form (Fig. 3 *a* and *b*, black trace), residues 40–50 are within helix B, whereas residues 51–61 extend the helix at its C terminus. In this situation, residues 51–61 are relatively mobile. As a complementary result, the residues that are located within helix B in the respective structures have very similar solvent accessibility profiles (Fig. 3*b*). In contrast, residues within the loop region are, on average, much more accessible to solvent. The black profile between residues 40 and 50 is very similar to the red profile between residues 51 and 61 (Fig. 3*b Inset*). These regions correspond, respectively, to helix B in the guanidinium-free and the guanidinium-bound structures. Ala-42 and Leu-46 are completely buried in the former structure, whereas Ala-53 and Leu-57 are buried in the latter. In contrast, the residues that are outside helix B are, on average, much more solvent-exposed in both the ligand-bound and ligand-free states. For example, Leu-57, which is fully buried in the guanidinium-bound state, is very solvent-exposed in the absence of the ligand. The slight dip in Ala-63 in the profile for the guanidinium-bound structure is due to the binding of the ligand.

Because the binding of the ligand causes such large conformational changes in the loops at each end of the B helix, this construct, or others like it, might represent a template for nanobiotechnology. For example, it might be possible to incorporate into one or another of the flanking loop regions a macromolecule whose function could be modulated by these conformational changes. The ligand-linked motion might also be translated into a detectable macroscale signal (e.g., a change in fluorescence) suitable for the design of biosensors.

It might also be noted that gene duplication and sequence repetition, possibly followed by mutation, have been suggested as possible mechanisms for protein evolution (26–28). The present design illustrates how these steps might occur. Moreover, the results also suggest that regulatory molecules that mimic functional groups within the set of common amino acids (e.g., formate, guanidinium, imidazole, etc.) may represent a likely set of "primitive" effector molecules during the evolution of allosteric processes. Modulation of weakly structured nonconserved surface loops could be a plausible strategy to introduce new functions.

In summary, a molecular switch has been designed in T4 lysozyme. Modulation of surface loops that flank duplicated segments of secondary structures might be a general strategy for the design of nanobioswitches.

We thank Dr. Martin Sagermann for helpful discussions, Dr. Michael Blaber for constructive comments on the manuscript, and Damon Hamel for collecting the x-ray data for the unliganded form at Advanced Light Source (Beamline 8.3.1).

1. Weber, G. (1992) in *Protein Interactions* (Chapman & Hall, London).
2. Mizoue, L. S. & Chazin, W. J. (2002) *Curr. Opin. Struct. Biol.* **12,** 459–463.
3. Abadou, A. & Desjarlais, J. R. (2001) *Protein Sci.* **10,** 301–312.
4. Abadou, A., Shenvi, R. A. & Desjarlais, J. R. (2001) *Biochemistry* **40,** 12719–12726.
5. Luo, B. H., Springer, T. A. & Takagi, J. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 2403–2408.
6. Nelson, M. R., Thulin, E., Fagan, P. A., Foresen, S. & Chazin, W. J. (2002) *Protein Sci.* **11,** 198–205.
7. Joyce, M. G., Girvan, H. M., Munro, A. W. & Leys, D. A. (2004) *J. Biol. Chem.* **279,** 23287–23293.
8. Benson, D. E., Haddy, A. E. & Hellinga, H. W. (2002) *Biochemistry* **41,** 3262–3269.
9. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. (2003) *Nature* **423,** 185–190.
10. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 11255–11260.
11. Doi, N. & Yanagawa, H. (1999) *FEBS Lett.* **453,** 305–307.
12. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11241–11246.
13. Guntas, G. & Ostermeier, M. (2004) *J. Mol. Biol.* **336,** 263–273.
14. Radley, T. L., Markowska, A. I., Bettinger, B. T., Ha, J. H. & Loh, S. N. (2003) *J. Mol. Biol.* **332,** 529–536.
15. Sagermann, M., Baase, W. A. & Matthews, B. W. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 6078–6083.
16. Sagermann, M., Gay, L. & Matthews, B. W. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 9191–9195.
17. Poteete, A. R., Dao-pin, S., Nicholson, H. & Matthews, B. W. (1991) *Biochemistry* **30,** 1425–1432.
18. Vetter, I. R., Baase, W. A., Heinz, D. W., Xiong, J.-P., Snow, S. & Matthews, B. W. (1996) *Protein Sci.* **5,** 2399–2415.
19. 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.
20. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
21. Rossmann, M. G. (1972) in *Molecular Replacement Method* (Gordon & Breach, New York).
22. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M. & Pannu, N. S. (1998) *Acta Crystallogr. D* **54,** 905–921.
23. Eriksson, A. E., Basse. W. A. & Matthews, B. W. (1993) *J. Mol. Biol.* **229,** 747–769.
24. Baldwin, E., Baase, W. A., Zhang, X.-J., Feher, V. & Matthews, B. W. (1998) *J. Mol. Biol.* **277,** 467–485.
25. Blaber, M. (2004) *Trends Biotechnol.* **22,** 1–2.
26. Groves, M. R. & Barford, D. (1999) *Curr. Opin. Struct. Biol.* **9,** 383–389.
27. Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. (2001) *Struct. Biol.* **134,** 117–131.
28. Hartmann, M., Schneider, T. R., Pfeil, A., Heinrich, G., Lipscomb, W. N. & Braus, G. H. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 862–867.