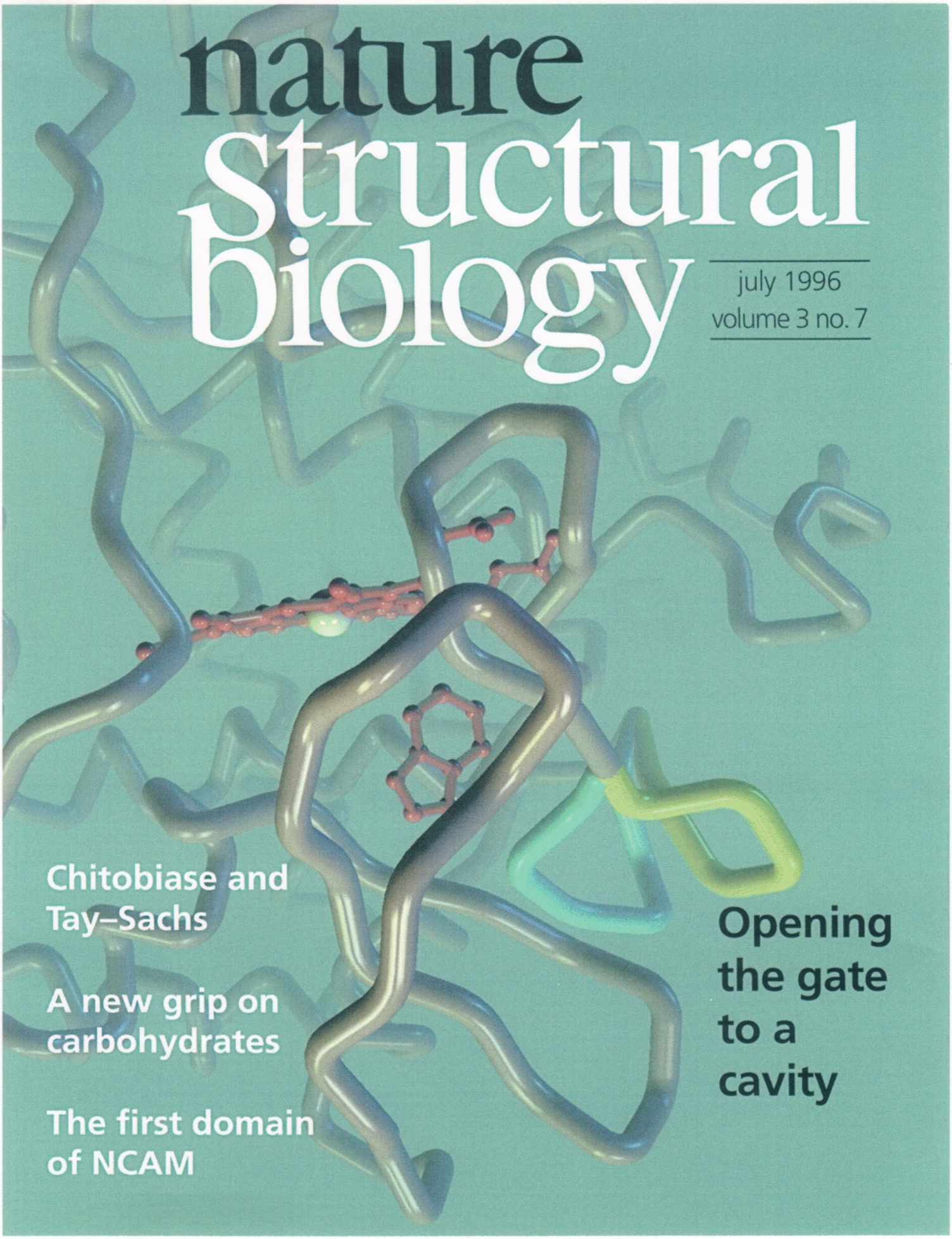
nature structural biology

july 1996
volume 3 no. 7



Chitobiase and Tay–Sachs

A new grip on carbohydrates

The first domain of NCAM

Opening the gate to a cavity

A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity

Melissa M. Fitzgerald, Rabi A. Musah, Duncan E. McRee and David B. Goodin

Conformational changes that gate the access of substrates or ligands to an active site are important features of enzyme function. In this report, we describe an unusual example of a structural rearrangement near a buried artificial cavity in cytochrome *c* peroxidase that occurs on binding protonated benzimidazole. A hinged main-chain rotation at two residues (Pro 190 and Asn 195) results in a surface loop rearrangement that opens a large solvent-accessible channel for the entry of ligands to an otherwise inaccessible binding site. The trapping of this alternate conformational state provides a unique view of the extent to which protein dynamics can allow small molecule penetration into buried protein cavities.

Department of Molecular Biology, MB8, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

Correspondence should be addressed to D.B.G.

Fluctuations in protein surface loop structures play an important role in molecular recognition. For many enzymes, including dihydrofolate reductase[1], triosephosphate isomerase[2], cytochrome P450[3], myoglobin[4] and HIV protease[5,6], conformational gating can provide a mechanism to recruit substrates to the active site, to exclude solvent, to sequester reactive intermediates, or to increase substrate specificity by 'induced fit'[7,8]. With a few exceptions [9–14], the dynamics and extent of such loop motion is not well characterized. Buried, solvent inaccessible cavities have been introduced into proteins such as T4 lysozyme by mutagenesis, and in some cases small molecules such as benzene will bind these cavities[15–17]. Again, while significant dynamics of the protein structure is implied in the process of ligand binding, the nature of these fluctuations is unclear.

A buried cavity was recently introduced into the active site of the haem enzyme cytochrome *c* peroxidase[18] (CCP). Trp 191, which is reversibly oxidized to a cation free radical in the native enzyme[19–24], was replaced by glycine (W191G). The resulting cavity was found to specifically bind a number of small cationic molecules such as protonated imidazoles by simply soaking crystals of the protein in buffers containing the target compound[18,25–26]. This result is surprising because the crystal structures of W191G or W191G containing bound imidazole derivatives (PDB entries 1cmq and 1cmp respectively[18]) showed that the cavity is completely buried from solvent access, providing no obvious pathway for ligands to enter. Thus, the facile binding of ligands and solvent molecules within this cavity implies that significant dynamic changes are occurring in the protein structure to provide one or more pathways for ligand entry even in the crystalline state. In this study, we structurally characterize a ligand-bound state of the enzyme, which reveals that an alternate conformation of a surface loop provides solvent access to the cavity. This appears to be a trapped 'gate open' state through which ligands may enter and exit. The kinetic properties of ligand binding to the W191G mutant provide information about the energetics of this motion.

Ligand-induced opening of a channel

While smaller ligands such as protonated 1,2-dimethylimidazole (DMI^+) bind to the cavity of W191G without significant changes in the protein structure[18], protonated benzimidazole ($BzIm^+$) binding induces a rearrangement of a surface loop which, in the ligand-free structure, blocks solvent access to the cavity. Crystal structures were determined at pH 4.5 for both DMI^+ (pK_a=7.9) and $BzIm^+$(pK_a=5.5), as previous studies[18,26] have shown that the cavity preferentially binds imidazole as its protonated cation. The $F_{BzIm^+}-F_{unsoaked}$ difference Fourier map gave clear difference features consistent with the displacement of solvent by $BzIm^+$. In addition, negative difference density was observed along the main chain of the PGGAAN sequence (a surface loop composed of residues 191–195), while new density was observed nearby. A model for a new 'open' conformation of the main chain from residues 190–195 was built and refined against the W191G/$BzIm^+$ data (Table 1). The 'omit map' (Fig. 1) exhibits clear density for the benzimidazole, which binds in a position identical to the Trp 191 side chain of the WT enzyme, and displaces the bound potassium and waters found in the ligand-free W191G structure[25,26]. Asp 235 hydrogen bonds with one of the imidazole N atoms of $BzIm^+$, analogous to the interaction with Trp 191 in the native enzyme.

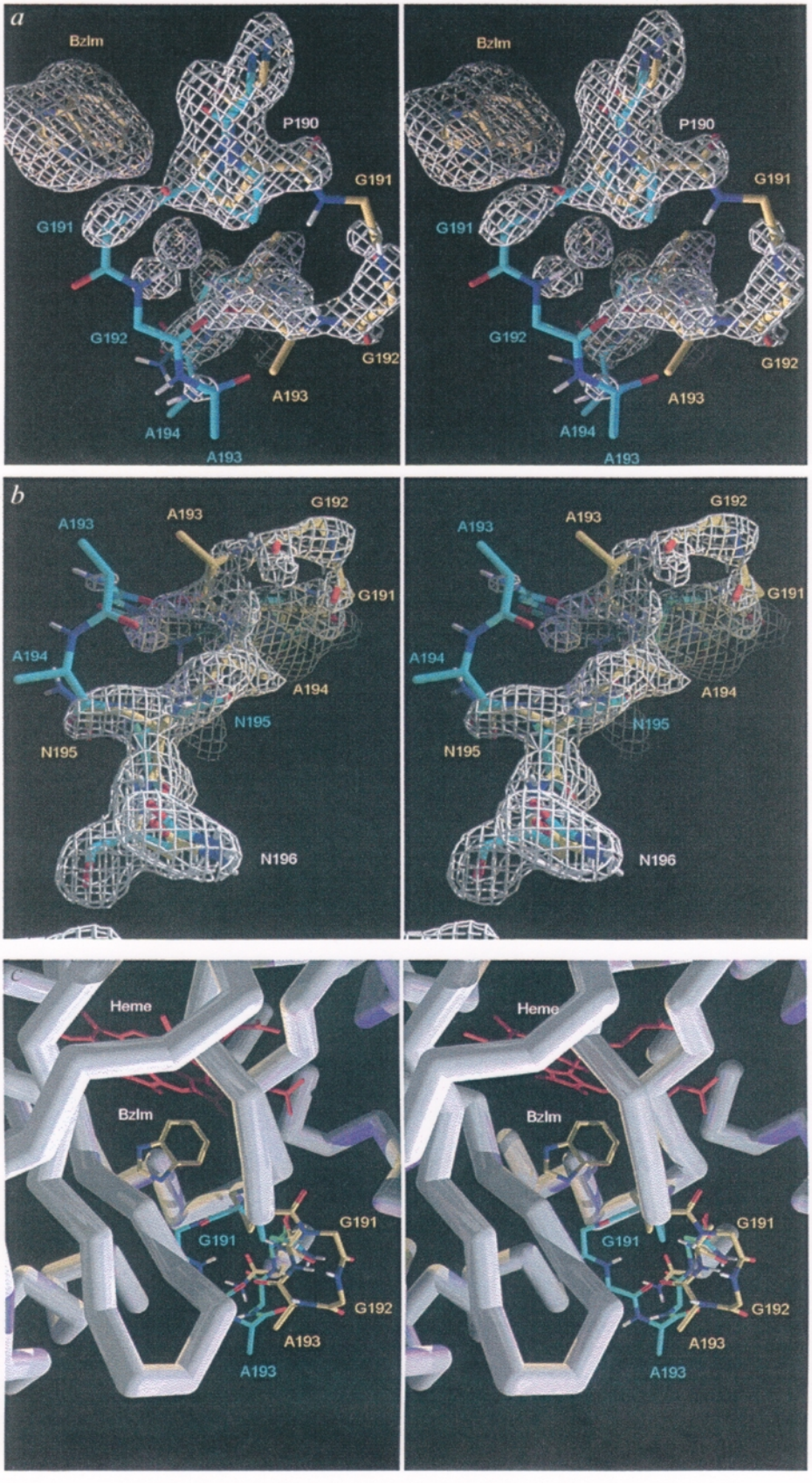


Fig. 1 Omit map (F_o–F_c) electron density for W191G with Bzlm$^+$ bound in the artificial cavity of CCP. The map is contoured at 3σ to show the main chain occupancy of an alternate conformation of the loop containing residues 190–195, which was induced by the binding of Bzlm$^+$. The refined structures at 1.9 Å resolution are superimposed on this map for the ligand bound, 'gate open' conformation (yellow), and the unbound, 'gate-closed' conformation (blue). *a*, The conformational change involves a *cis-trans* isomerization of Pro 191 at one hinge, and *b*, rotation about the ψ (from 16.0° (closed) to 174.5° (open)) and ϕ (from –95.2° (closed) to –131.1° (open)) torsion angles of Asn 195 at the second hinge. The side chain of Asn 195 in one conformation mimics the main chain of that residue in the alternate conformation. c, The relationship of the loop to the protein surface, haem, and bound Bzlm$^+$ is shown from approximately the same viewpoint as in (*a*). The omit map was calculated by omitting the Bzlm$^+$ ligand and residues 189–196 from the model followed by 150 cycles of refinement before calculation of the F_o–F_c map.

Table 1 Data collection and refinement statistics for W191G/BzIm[1]

Data Collection:		
	Unit cell	$P2_12_12_1$ (a= 108.0 Å, b = 77.3 Å, c = 51.8 Å)
	Total reflections	68,690
	Unique reflections	30,535
	Resolution (Å)	1.9
	$I/\sigma_{I(ave)}$	15.4
	$I/\sigma_{I(last\ shell)}$	0.95
	Completeness (%)	89
	R_{sym}[2]	0.052
Refinement:		
	R_{cryst} (%)[3]	16.4
	Resolution (Å)	5.0–1.8
	No. of reflections	21,900
	r.m.s. bond (Å)[4]	0.013
	r.m.s. angle (°)	2.7
	No. of solvent molecules	99

[1] I/σ_I (last shell) represents that value for the 10% of the data of highest resolution.
[2] $R_{sym} = \Sigma |I_h - \bar{I}_h| / \Sigma I_h$.
[3] $R_{cryst} = \Sigma (|F_{obs} - F_{calc}|) / \Sigma (F_{obs})$.
[4] R.m.s. bond and r.m.s. angle represent the root-mean-squared deviation between the observed and ideal values.

The ligand-induced loop conformational change in the structure of W191G/BzIm$^+$ involves a double-hinged rotation about Pro 190 and Asn 195. Electron density is observed (Fig. 1) for the main-chain atoms of residues 190–195 in the new open conformation, with the exception of Cα of residue 191. A *cis-trans* isomerization is observed at Pro 190, allowing the main chain to re-pack away from the bound benzimidazole. The altered loop conformation rejoins the native structure at Asn 195, which undergoes a rotation of 40° and 158° about the φ and ψ torsion angles, respectively, compared with those of ligand-free W191G. This results in a rough interchange in the positions of the Asn 195 side-chain and main-chain atoms. The ability of the Asn side chain to mimic the main chain has been noted as a factor responsible for the preponderance of Asn residues at the beginning of helices[27]. A significant increase in *B*-values for residues 189–194 is observed on binding BzIm$^+$ (Fig. 2) which indicates that either this segment is undergoing significant thermal motion in the bound state, or that the open conformation is not fully populated. We note that similar increases in *B*-values for residues 189–195 were reported on binding of NO to the haem of CCP, but without a significant change in the observed protein conformation[28]. It was suggested that these residues became disordered on complex formation with NO, and a role was proposed for structural adjustment of these residues in facilitating electron transfer to Trp 191 in the electron transfer complex with cytochrome *c*.

Variations in the chemical identity of the ligand appear to determine the extent of loop movement on binding. Structures of several bound cationic compounds which are similar to BzIm$^+$ were characterized by crystallography. These included the cationic forms of indoline (In$^+$, K_d=0.26 mM), aniline (An$^+$, K_d=1.31 mM) and imidazo[1,2-a]pyridine (IP$^+$, K_d=0.26 mM) each at pH 4.5. In addition, the structure with DMI$^+$ was re-determined at pH 4.5. For both In$^+$ and IP$^+$, no clear electron density for residues 190–195 was observed in either the open or closed state, indicating significant structural disorder in this loop. However, for W191G/An$^+$ and W191G/DMI$^+$, well defined electron density was observed only for the closed conformation. That the three larger ligands (BzIm$^+$, In$^+$ and IP$^+$) appear to result in partial loop opening, while the smaller ligands (Im$^+$, DMI$^+$ and An$^+$) bind with the loop closed indicates that van der Waals contacts with the larger ligands prevent the closing of the loop. None of the ligands make direct hydrogen bonds with the loop sequence. In addition, if BzIm$^+$ were to bind with the closed loop conformation, it would be denied the potential for one of the ring nitrogens to hydrogen bond to solvent molecules in the channel. This may provide an explanation for the more highly populated open conformation with this ligand.

The hinged loop movement in W191G/ BzIm$^+$ introduces, in the open state, a large channel that suggests a pathway for ligand entry. The direction of the loop movement with respect to the bound imidazole and the protein surface (Fig. 3) suggests that access to ligand binding may appear transiently during motional dynamics of the W191G structure. A solvent-accessible surface of these two states is shown in Fig. 4*a*,*b*. The cavity dimensions increase from ~180 Å^3 in the closed, ligand-free form to ~450 Å^3 in the open, liganded form. The deep channel that is observed in the open conformation and the access that it provides to the haem centre is very suggestive of the active-site channel seen in the structure of prostaglandin H_2 synthase-1[29].

Kinetics and energetics of loop opening

The association and dissociation rate constants for the binding of two imidazole compounds to W191G indicate that the difference in binding affinity is deter-

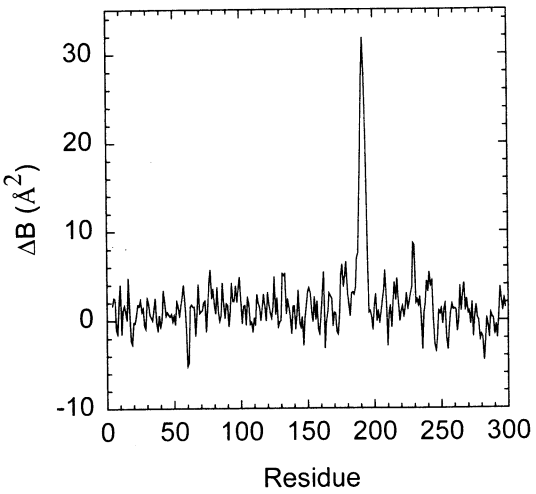


Fig. 2 Plot of the signed difference between the averaged crystallographic temperature factors (*B*-values) of main-chain atoms for W191G with and without benzimidazolium bound. The largest difference increase occurs for residues 189–194.

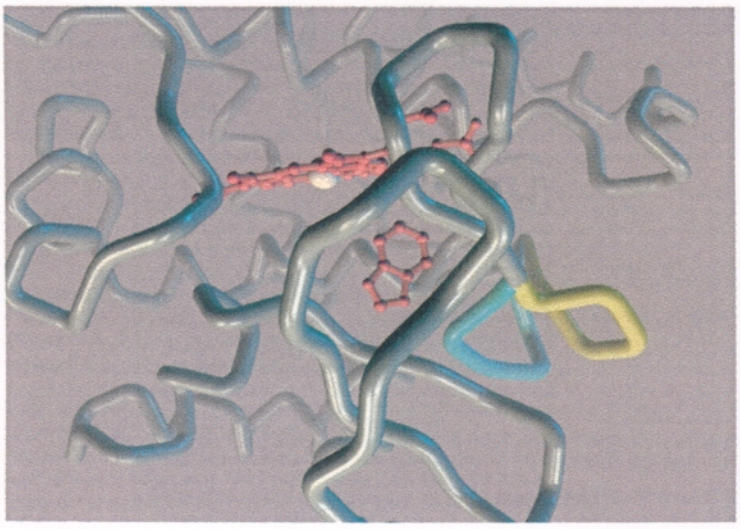


Fig. 3 Tube representation of the Cα main chain of W191G with BzIm+ bound at the position of Trp 191. The unbound closed conformation of the loop from residues 190–195 is shown in blue and the ligand-bound open conformation is in yellow.

mined by the off rates. Stopped-flow absorption measurements were carried out to determine the association rates for DMI^+ and $BzIm^+$. At 20 °C, k_{on} was observed to be 2.9×10^4 $M^{-1}sec^{-1}$ for DMI^+ and 6.5×10^4 $M^{-1}sec^{-1}$ for $BzIm^+$. Under these conditions, the dissociation constant K_d was 0.05 mM for DMI^+ and 0.84 mM for $BzIm^+$. The values for k_{off}, determined from k_{on} and K_d were 1.4 sec^{-1} for DMI^+ and 55 sec^{-1} for $BzIm^+$. Thus, the difference in affinity of these two ligands for the cavity is due primarily to the variation in the off rates, in which the larger $BzIm^+$ ligand dissociates ~50-fold faster than DMI^+. This is fully consistent with the structural results because the loop remains open after $BzIm^+$ binding but is observed to be closed when DMI^+ is bound.

While standard analysis of the activation parameters for ligand binding to W191G would indicate that the transition state for binding is similar for both compounds, and is achieved by a transition through a high entropy state of the protein, the large scale motions that accompany binding suggest that this apparent simplicity may belie more complex behaviour. The temperature dependence of k_{on} for binding DMI^+ or $BzIm^+$ to the W191G cavity appears to follow an Arrhenius relationship (Fig. 5) over a narrow temperature range. Standard transition state analysis gives activation free-energies of 10.9 and 10.4 kcal mol^{-1} for DMI^+ and $BzIm^+$ respectively, and provides activation enthalpies (27.1 and 27.0 kcal mol^{-1}) and entropies (54.3 and 55.7 entropy units, cal deg^{-1} $mole^{-1}$) for binding that are quite similar for the two ligands despite their differences in size and hydrophobicity. Indeed, the activation enthalpy of 27 kcal mol^{-1}, which may well reflect the isomerization of Pro 190, is nearly identical to that measured for a proline isomerization in the refolding of apoplastocyanin[30]. In addition, the similar thermodynamic parameters observed for the on rate of both ligands are to be expected if the transition state for binding is defined by the opening of the 'gate'. However, studies of ligand-gated protein rearrangements in myoglobin[11–14] have revealed non-Arrhenius temperature dependencies and non-exponential kinetics when examined over large scales. This behaviour—resulting from transitions through complex conformational energy surfaces with multiple minima and statistical averaging over a multitude of possible

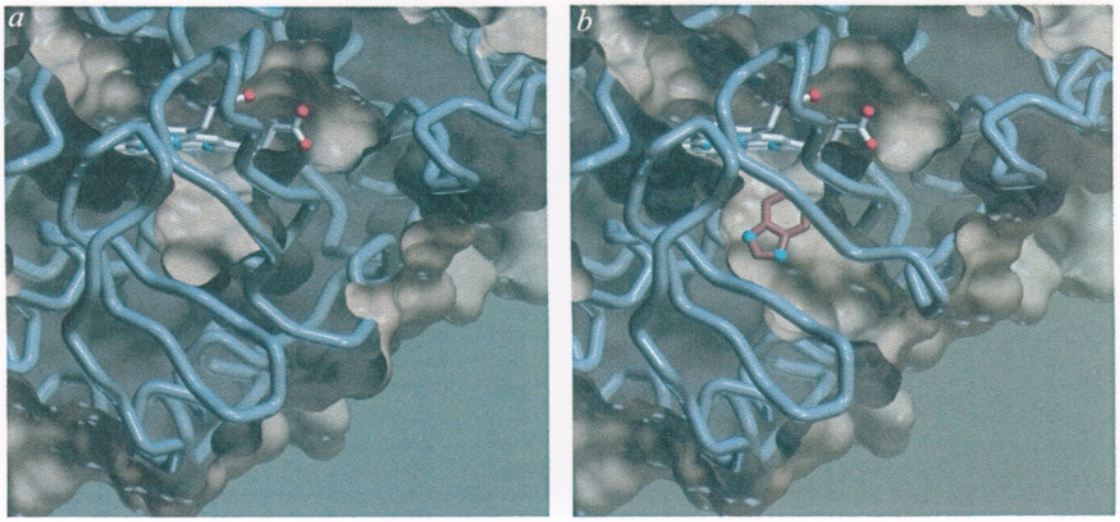


Fig. 4 Solvent-accessible surface calculated for the artificial cavity of *a*, W191G or *b*, W191G with BzIm+ bound in place of Trp 191. The cavity dimensions increase from ~180 $Å^3$ in the ligand-free form to ~450 $Å^3$ in the liganded form due to a conformational change in the loop forming the bottom and side walls of the W191G artificial cavity. The viewpoint of this figure is approximately the same as in Fig. 3. To show both the cavity and the loop, the clipping plane of this figure cuts the cavity only. The altered conformation of the loop creates a large solvent-accessible pathway to the otherwise buried cavity. The solvent accessible surfaces were calculated using a 1.4 Å probe sphere[33].

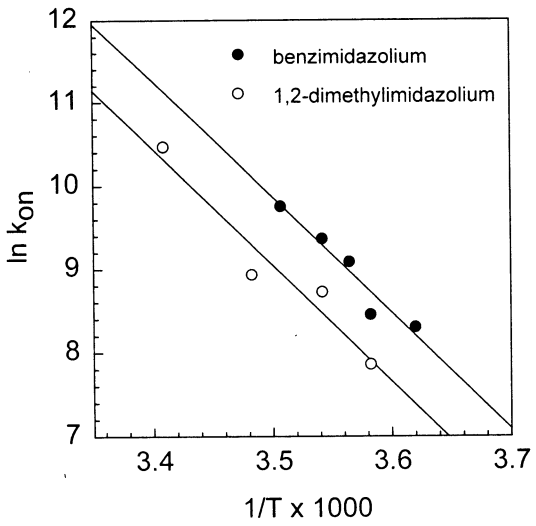


Fig. 5 Arrhenius plot of the association constant, k_{on}, for the binding of $BzIm^+$ (○) or DMI^+ (●) to W191G. The linear least squares fits of the data are given by $\ln(k_{on})=(-1.39\pm0.22\times10^4)\ (1/T)+(58.5\pm7.8)$ for $BzIm^+$, and $\ln(k_{on})=(-1.39\pm0.25\times10^4)\ (1/T)+(57.8\pm8.8)$ for DMI^+.

pathways—is expected for large scale protein motions involving many bond torsions. Evidence that such complexity may exist in the opening of the W191G loop is provided in the pre-exponential terms of the Ahrrenius plots which are much too high (10^{24}–10^{25}) to be interpreted in terms of a meaningful frequency factor. Thus, strict interpretation of the pre-exponential component of the Ahrrenius plot in terms of transition state entropy may not be valid, inviting a detailed examination of the conformational dynamics of this transition.

Implications

Removal of Trp 191, which is oxidized by the haem of the native enzyme, not only creates a ligand-binding cavity in CCP but also appears to introduce sufficient flexibility in residues 190–195 to enable the opening of a pathway for ligands to reach the buried cavity. The kinetic data indicate that the conformational dynamics of the loop movement are rate limiting in the binding of ligands to the cavity. However, the off rate in the open conformation is ~50 sec^{-1}, suggesting that a high rate of substrate dissociation and turnover may be possible through further protein modifications to permanently open the channel. This suggests that cavities large enough to recognize biologically relevant molecules such as steroids and nucleic acids could be introduced into proteins by this method of 'cavity complementation'. Such proteins could serve as highly specific designed catalysts and receptors which may have use as therapeutic agents and biosensors.

Methods

The W191G mutant of cytochrome c peroxidase was constructed by site-directed mutagenesis, overexpressed in *Escherichia coli.* BL21(DE3), and purified as previously described[18]. Single crystals for X-ray diffraction were grown by vapour diffusion from 25% 2-methyl-2,4-pentanediol (MPD). Solutions for crystal soaks were prepared by dissolving the probe compound in 40 % MPD, adjusting the pH with dilute H_3PO_4, and diluting with distilled H_2O to a final probe molecule concentration of 50 mM in 24% in MPD. Crystals were soaked for at least one hour before mounting. Cu K radiation from a Rigaku X-ray generator was used to produce X-ray diffraction data, which was collected at 15 °C with a Siemens area detector. Analysis and model building were performed as described previously[18] using difference Fourier methods as implemented in XtalView[31], and least squares refinement of coordinates as implemented in XPLOR 3.1[32]. Crystals of W191G obtained for this study were in a different crystal form than that of PDB entry 1cmq, but were the same as a second determination of the W191G structure reported in the same study[18]. The cavity volumes were calculated by creating a fine 0.1 Å grid that covered the region of the cavity. The grid points lying outside the cavity solvent-accessible surface were removed, and the volume was determined by the number of enclosed grid points. For the open form of the loop, the neck of the channel was manually blocked by adding a water molecule during the calculation. Coordinates have been deposited in the Brookhaven Protein Data Bank, accession code 1RYC.

Rate constants for the binding of benzimidazole or 1,2-dimethylimidazole to W191G were determined from stopped flow absorption measurements using a Hi Tech Scientific SHU stopped-flow single-wavelength spectrophotometer interfaced to a Hewlett-Packard 9000 computer and equipped with a variable temperature bath. It has been shown that imidazole binding produces a 1–2 % intensification of the haem Soret absorption band[18], providing a spectroscopic probe for monitoring the binding event. Kinetic traces were obtained over a range of ligand concentrations (0.5–8 mM benzimidazole or 0.01–0.5 mM 1,2-dimethylimidazole in the reaction cell) and temperatures (3–20 °C). Solutions were buffered with 0.1 M ammonium acetate buffer, pH 4.5. 1 ml of 10 μM W191G was placed in one syringe, and 1 ml of the ligand solution was placed in the other syringe. Following a five minute temperature equilibration, 100 μl of each reactant was mixed in the stopped-flow apparatus, and the absorption at 400 nm in a 1 cm cell was monitored as a function of time using a 100 μs photomultiplier time constant and 1.5 mm slit-width. k_{obs} for each ligand concentration was obtained by fitting an exponential to the averaged data from 4–8 scans. The fit was performed over the initial 200 or 300 ms for benzimidazole, and over the initial 1000 or 2000 ms for 1,2-dimethylimidazole. At each temperature, the second order rate constant, k_{on}, for the binding of the respective ligand (L) to W191G was determined from the linear least-squares fit of the data to the equation $k_{obs}=k_{off}+k_{on}$ [L]·k_{off} at 20 °C was determined from the expression $k_{off}=k_{on}K_d$, where k_{on} at 20 °C was calculated from an Arrhenius plot for k_{on}, and K_d was measured by optical absorbance difference titrations of W191G[18] with benzimidazole or 1,2-dimethylimidazole in 0.1 M ammonium acetate buffer, pH 4.5, at 20 °C.

Received 1 March; accepted 23 May 1996.

Acknowledgements
We gratefully acknowledge G.M. Jensen, S.K. Wilcox, H. Frauenfelder, I.A. Wilson, P.E. Wright, B. Matthews and A.E. McDermott for helpful discussions and critical reading of the manuscript. This research was supported by National Institutes of Health grants to D.B.G., to D.E.M., and to M.M.F.

1. Li, L.Y., Falzone, C.J., Wright, P.E. & Benkovic, S.J. Functional role of a mobile loop of *Escherichia coli* dihydrofolate reductase in transition-state stabilization. *Biochemistry* **31**, 7826–7833 (1992).
2. Sampson, N.S. & Knowles, J.R. Segmental movement - definition of the structural requirements for loop closure in catalysis by triosephosphate isomerase. *Biochemistry* **31**, 8482–8487 (1992).
3. Li, H. & Poulos, T.L. Modeling protein-substrate interactions in the heme domain of cytochrome P450BM-3. *Acta Cryst.* D**51**, 21–32 (1995).
4. Nobbs, C.L. *Heme and Hemoproteins* **1**, 143–147 (Academic Press, New York, 1966).
5. Jaskolski, M. *et al.* Structure at 2.5 Å resolution of chemically synthesized human immunodeficiency virus type I protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry* **30**, 1600–1609 (1991).
6. Rao, J.K.M., Erickson, J.W. & Wlodawer, A. Structural and evolutionary relationship between retroviral and eucaryotic aspartic proteinases. *Biochemistry* **30**, 4663–4671 (1991).
7. Koshland, D.E. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA* **44**, 98–104 (1958).
8. Herschlag, D. The role of induced fit and conformational changes of enzymes in specificity and catalysis. *Bioorg. Chem.* **16**, 62–96 (1988).
9. Williams, J.C. & McDermott, A.E. Dynamics of the flexible loop of triosephosphate isomerase: the loop motion is not ligand gated *Biochemistry* **34**, 8309–9319 (1995).
10. Falzone, C.J., Wright, P.E. & Benkovic, S.J. Dynamics of a flexible loop in dihydrofolate reductase from *Escherichia coli* and its implication for catalysis *Biochemistry.* **33**, 439–442 (1994).
11. Young, R.D. *et al.* Time- and temperature dependence of large-scale conformational transitions in myoglobin. *Chemical Physics* **158**, 315–327 (1991).
12. Frauenfelder, H. & Wolynes, P.G. Rate theories and puzzles of hemeprotein kinetics. *Science* **229**, 337–345 (1985).
13. Frauenfelder, H., Sligar, S.G. & Wolynes, P.G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
14. Beece, D. *et al.* Solvent viscosity and protein dynamics. *Biochemistry* **19**, 5147–5157 (1980).
15. Eriksson, A.E. *et al.* Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178–183 (1992).
16. Eriksson, A.E., Baase, W.A., Wozniak, J.A. & Matthews, B.W. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **355**, 371–373 (1992).
17. Eriksson, A.E., Baase, W.A. & Matthews, B.W. Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4-lysozyme have different structural and thermodynamic consequences. *J. Mol. Biol.* **229**, 747–769 (1993).
18. Fitzgerald, M.M., Churchill, M.J., McRee, D.E. & Goodin, D.B. Small molecule binding to an artificially created cavity at the active site of cytochrome c peroxidase. *Biochemistry* **33**, 3807–3818 (1994).
19. Erman, J.E., Vitello, L.B., Mauro, J.M. & Kraut, J. Detection of an oxyferryl porphyrin π-cation-radical intermediate in the reaction between hydrogen peroxide and a mutant yeast cytochrome c peroxidase. Evidence for tryptophan-191 involvement in the radical site of compound I. *Biochemistry* **28**, 7992–5 (1989).
20. Scholes, C.P. *et al.* Recent ENDOR and pulsed electron paramagnetic resonance studies of cytochrome c peroxidase-compound I and its site-directed mutants *Isr. J. Chem* **29**, 85–92 (1989).
21. Sivaraja, M., Goodin, D.B., Smith, M. & Hoffman, B.M. Identification by ENDOR of Trp191 as the free-radical site in cytochrome c peroxidase compound ES. *Science* **245**, 738–740 (1989).
22. Houseman, A.L.P., Doan, P.E., Goodin, D.B. & Hoffman, B.M. Comprehensive explanation of the anomalous EPR spectra of wild-type and mutant cytochrome-c peroxidase compound-ES. *Biochemistry* **32**, 4430–4443 (1993).
23. Huyett, J.E. *et al.* Compound ES of cytochrome c peroxidase contains a Trp π-cation radical: characterization by CW and pulsed Q-band ENDOR spectroscopy. *J. Am. Chem. Soc.* **117**, 9033–9041 (1995).
24. Jensen, G.M., Goodin, D.B. & Bunte, S.W. Density functional and MP2 calculations of spin densities of oxidized 3-methyl indole: models for tryptophan radicals. *J. Phys. Chem.* **100**, 954–959 (1996).
25. Miller, M.A., Han, G.W. & Kraut, J. A Cation binding motif stabilizes the compound I radical of cytochrome c peroxidase. *Proc. Natl. Acad. Sci. USA* **91**, 11118–11122 (1994).
26. Fitzgerald, M.M., Trester, M.L., Jensen, G.M., McRee, D.E. & Goodin, D.B. The role of aspartate-235 in the binding of cations to an artificial cavity at the radical site of cytochrome c peroxidase. *Protein Science* **4**, 1844–1850 (1995).
27. Richardson, J.S. & Richardson, D.C. Amino acid preferences for specific locations at the ends of α helices. *Science* **240**, 1648–1652 (1988).
28. Edwards, S.L., Kraut, J. & Poulos, T.L. Crystal structure of nitric oxide inhibited cytochrome c peroxidase. *Biochemistry* **27**, 8074–8175 (1988).
29. Picot, D., Loll, P.J. & Garavito, R.M. The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1. *Nature* **367**, 243–249 (1994).
30. Koide, S., Dyson, H.J. & Wright, P.E. Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. *Biochemistry* **32**, 12299–12310 (1993).
31. McRee, D.E. A Visual protein crystallographic software system for X11/Xview. *J. Mol. Graphics* **10**, 44–46 (1992).
32. Brünger, A.T. & Karplus, M. Crystallographic R-factor refinement by molecular dynamics. *Science* 235, 458–460 (1987).
33. Connolly, M.L. Solvent accessible surfaces of proteins and nucleic acids. *Science* **221**, 709–713 (1983).

Not just your average structures

Gregory A. Petsko

Rosensteil Basic Medical Research Center, Brandeis University, Waltham, Massachusetts 02254, USA

Structures determined by X-ray crystallography and NMR spectroscopy are averages, whereas many important biological functions require significant departure from the 'average' conformation.

One of my favorite short stories concerns a day on which the Law of Averages mysteriously has been repealed. Instead of having a variety of breakfasts at a range of times, on that day every person in the New York metropolitan area has toast at exactly 7:23 a.m. The resulting power drain blacks out the entire region. A little later, all of the commuters who take the Holland Tunnel decide to enter it at precisely the same instant; the resulting traffic jam extends for fifty miles and stops all traffic for seven hours. The rest of the day is more of the same.

Aside from being amusing, this story makes a serious point. Our society depends upon randomized behaviours that average out to a rough regularity. No one works exactly nine to five, people do not shower at exactly the same moment every day, and so forth. We have designed our civilization around averages, but we understand that reality consists of fluctuations about those averages.

The beautiful protein structures that grace the pages of this journal are averages too. Every protein crystal structure, for example, is a double average over all asymmetric units in the crystal lattice and the time required to collect the diffraction data. NMR-derived structures are also time and ensemble averages. But so powerful are the static images of structures we present, that it is necessary to remind ourselves that some aspects of protein function depend on what we don't see: the excursions from the average. In this and the June issues of *Nature Structural Biology*, three papers illustrate of the importance of such excursions in dramatic fashion: by showing that large organic molecules can penetrate deep into proteins and bind to cavities that are inaccessible in the average structure.

This observation is not, in itself, new. Perutz and Kendrew recognized in the 1960s that oxygen could not reach the haem iron in myoglobin and haemoglobin unless the protein was flexible. Ringe and Ortiz de Montellano showed in the early 1980s that aryl groups like phenylhydrazine could also diffuse through myoglobin to that site[1]. It has also been known for decades that xenon atoms can bind in buried pockets in myoglobin and other proteins[2] (Fig. 1). One of the xenon-binding cavities in myoglobin can also bind mercury triiodide (D. Ringe, personal communication); this huge molecule can actually penetrate the protein in the crystalline state, ruling out global, or even local, protein unfolding as a required mechanism for penetration. The ligand must 'worm' its way through the structure with the help of countless small displacements of atoms and groups of atoms about their equilibrium conformation. And camphor, an even larger molecule, is able to penetrate through tens of Ångstroms of 'solid' protein to reach the completely buried active site in *Pseudomonas putida* cytochrome P450 (ref. 3).

But none of this work carries the subliminal power of the lovely ribbon diagrams crystallographers publish that show THE structure. Our NMR colleagues are somewhat more scrupulous about this: they show families of similar structures that fit their distance restraints. Yet in the end, even they produce atomic coordinates, and people use these coordinates, and the moment they do, the message that this is THE structure is there, lurking on the edges of consciousness like the rhinoceros in the Ionesco play.

Our insistence on equating the average structure with the only structure has consequences for everything we do. Crystallographic *R*-factors are higher than they should be for protein structures; as Ed Lattman points out in a wonderful essay[4], this is at least in part due to our insistence on imposing 'ideal' geometry on our models. The average of an ensemble of structures may have some significantly non-ideal bond-lengths, angles and non-bonded contacts. This point was made years ago[5], and was promptly ignored in the search for 'correct' structures. Other examples could be given. So it is not just useful, it is imperative, to remind ourselves that protein flexibility gives these molecules capabilities that are not obvious from simple examination of the average structure. That is what the three articles in this and the last issue do.

Dalquist and associates[6] use NMR to examine the kinetics of exchange between free and bound ligands (indole and benzene) to engineered hydrophobic cavities in T4 lysozyme. Although the static crystal structures of ligand–T4 complexes suggest that backbone fluctuations on the order of Ångstroms are required to allow these ligands to reach their buried binding sites, the NMR data indicate that the association constants are on the order of 10^6–10^7 per mole per second. In other words, fluctuations of that magnitude must occur on a microsecond time-scale. These numbers are completely consistent with the known association rates of oxygen, carbon monoxide, and other ligands with the buried haem site in myoglobin, but it is important to see that larger ligands can pentrate proteins at similar rapid rates. These authors further argue that 'polarity barriers', rather than steric limitations, may govern the general inabili-

Fig. 1 Filling cavities. Four xenon atoms occupy preformed cavities of packing voids within myoglobin[2]. The Cα backbone is shown in green and the haem in purple. The cavities are highlighted by a blue Conolly dot surface using a 1.4 Å probe sphere.

ty of enzyme substrates to diffuse freely through the proteins and the slow exchange rate of water with some buried amide groups. It has long been known that xenon will not bind to polar cavities in a protein, even if they are large enough to accommodate a xenon atom and even if the partial pressure of xenon is very high. This concept also explains why some protein cavities are truly 'empty' (that is, they do not contain even disordered solvent molecules) while others are not. It is worth noting that water can penetrate proteins rapidly, too. Denisov, *et al.*[7], in the June issue, describe dispersion mesurements on everyone's favorite small protein, bovine pancreatic trypsin inhibitor, that show the residence time of a buried water molecule to be 170±20 μs.

Goodin and associates use X-ray crystallography to show that a hinged main-chain rotation of a six-residue loop opens a large surface channel for the entry of protonated benzimidazole to an artificial cavity in cytochrome *c* peroxidase[8]. They also use stop-flow measurements to show that the rate constants for ligand binding are on the millisecond time-scale. The activation free energies for this process (~10 kcal mol^{-1}) are at least twice as large as those estimated for ligand binding to T4 lysozyme, but still small enough to allow rapid binding. Such motions can be important for catalysis: for example, the hinged movement of the active-site loop in triosephosphate isomersase, which renders the active site inaccessible to solvent and which causes some atoms to move by more than 10 Å, must occur on a sub-microsecond time-scale to account for the observed rate of catalysis[9]. The work described in this paper demonstrates that such motions are indeed compatible with the half-times of many significant biological processes.

Goodin and associates speculate that cavities large enough to recognize steroids or other drugs can be introduced into proteins as a vehicle for the creation of specific catalysts and receptors. Alber and co-workers, in the June issue, have taken the idea even further[10]. They engineered a GCN4 leucine zipper mutant that switches from a two-to-three-stranded coiled-coil on binding of hydrophobic ligands to the engineered intramolecular cavity. They confirm by X-ray crystallography that benzene indeed binds to the core of the trimer, as it was designed to. In addition to showing that core packing can be the determining factor for the oligomerization state of a protein, the work of Alber *et al.* has profound implications for structure-guided drug design. It has already been shown that protein flexibility is important for protein function. Some of the Sterling–Winthrop antiviral compounds are thought to work by changing the conformational entropy of the virus capsid[11]. Some of the inactivating mutants of ribulose bis-phosphate carboxylase oxygenase can only be explained by assuming that they reduce the flexibility of the enzyme. There are numerous other examples. Alber and co-workers show that hydrophobic packing interactions at even a single site can select among possible oligomers. Extending this observation, it should be possible to design ligands that disrupt protein oligomerization or protein–protein interactions. Many of the most interesting targets in biology depend on such interactions for function, but there are no good small-molecule ligands that interrupt them. Studies of ligand binding to protein cavities offer hope that such drugs can be designed or found by screening.

Protein structures are not the static creatures that their average pictures subtly suggest them to be. They are, in the wonderful phrase of Gregorio Weber, "kicking, screaming, stochastic molecules". Individual atoms are in constant, picosecond fluctuation. On a similar time-scale, collective motions of side-chains and non-bonded atomic clusters occur. Larger-scale movements of loops and secondary structure elements are only slightly slower, on a nanosecond time-scale, and some of these motions can be Ångstroms in magnitude. We all know that ligand-induced conformational changes in proteins can be gigantic; what we often forget is that the random movements that are subject of the work in these papers provide the 'lubrication' that allow such conformational changes to occur.

We can never remind ourselves too often that our structures are averages and that, seductive as they are, they are only a part of the story. Many attempts have been made to mimic protein behaviour with small, rigid molecules. These attempts usually fail. I think that one reason they fail is that proteins are very special molecules: they have regular three-dimensional structures but they are also highly flexible. Legend has it that Galileo, forced to recant his assertion that the Earth was not the immobile center of the Cosmos, muttered as he was led away, "Still, it moves." Still, so do proteins.

1. Ringe, D., Petsko, G.A., Kerr, D.E. & Ortiz de Montellano, P.R. *Biochemistry* **23**, 2–4 (1984).
2. Tilton, R.F., Kuntz, I.D. & Petsko, G.A. *Biochemistry* **12**, 2849–2857 (1985).
3. Poulos, T.L. *Pharm. Res.* **5**, 67–75 (1988).
4. Lattman, E.E. *Proteins Struct. Funct. Genet.* **25**, 1 (1996).
5. Ringe, D. & Petsko, G.A. *Progr. Biophys. Mol. Biol.* **45**, 197–235 (1985).
6. Feher, V.A., Baldwin, E.P. & Dahlquist, F.W. *Nature Struct. Biol.* **3**, 516–521 (1996).
7. Denisov, V.P., Peters, J., Hörlein, H.D & Halle, B. *Nature Struct. Biol.* **3**, 505–509 (1996).
8. Fitzgerald, M.M., Musah, R.A., McRee, D.E. & Goodin, D.B. *Nature Struct. Biol.* **3**, 626–631 (1996).
9. Alber, T. *et al.* Cold Spring Harbor Symp. Quant. Biol. LII, 603–613 (1988).
10. Gonzalez, Jr., L. Plecs, J.J. & Alber, T. *Nature Struct. Biol.* **3**, 510–515 (1996).
11. Phelps, D.K. & Post, C.B. *J. Mol. Biol.* **254**, 544–551 (1995).