UC Davis

UC Davis Previously Published Works

Title

Generation of ligand binding sites in T4 lysozyme by deficiency-creating substitutions.

Permalink

<https://escholarship.org/uc/item/2gd3m8px>

Journal

Journal of molecular biology, 277(2)

ISSN

0022-2836

Authors

Baldwin, E
Baase, WA
Zhang, XJ
et al.

Publication Date

1998-03-01

DOI

10.1006/jmbi.1997.1606

Peer reviewed

Generation of Ligand Binding Sites in T4 Lysozyme by Deficiency-creating Substitutions

Enoch Baldwin, Walter A. Baase, Xue-jun Zhang, Victoria Feher and Brian W. Matthews*

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

Several variants of T4 lysozyme have been identified that sequester small organic ligands in cavities or clefts. To evaluate potential binding sites for non-polar molecules, we screened a number of hydrophobic large-to-small mutants for stabilization in the presence of benzene. In addition to Leu99 → Ala, binding was indicated for at least five other mutants. Variants Met102 → Ala and Leu133 → Gly, and a crevice mutant, Phe104 → Ala, were further characterized using X-ray crystallography and thermal denaturation. As predicted from the shape of the cavity in the benzene complex, mutant Leu133 → Gly also bound *p*-xylene. We attempted to enlarge the cavity of the Met102 → Ala mutant into a deep crevice through an additional substitution, but the double mutant failed to bind ligands because an adjacent helix rearranged into a non-helical structure, apparently due to the loss of packing interactions. In general, the protein structure contracted slightly to reduce the volume of the void created by truncating substitutions and expanded upon binding the non-polar ligand, with shifts similar to those resulting from the mutations.

A polar molecule binding site was also created by truncating Arg95 to alanine. This creates a highly complementary buried polar environment that can be utilized as a specific “receptor” for a guanidinium ion. Our results suggest that creating a deficiency through truncating mutations of buried residues generates “binding potential” for ligands with characteristics similar to the deleted side-chain. Analysis of complex and apo crystal structures of binding and non-binding mutants suggests that ligand size and shape as well as protein flexibility and complementarity are all determinants of binding. Binding at non-polar sites is governed by hydrophobicity and steric interactions and is relatively permissive. Binding at a polar site is more restrictive and requires extensive complementarity between the ligand and the site.

© 1998 Academic Press Limited

Keywords: T4 lysozyme; benzene; cavities; guanidinium; structural complementarity

*Corresponding author

Present addresses: E. Baldwin, Molecular and Cellular Biology, 220 Briggs Hall, University of California, Davis, CA 95616, USA; X.-j. Zhang, Crystallography Program, Oklahoma Medical Research Foundation, 825 Northeast 13th Street, Oklahoma City, OK 73104, USA; V. Feher, Wadsworth Center, New York State Department of Health, Empire State Plaza, Albany, NY 12201-0509, USA.

Abbreviations used: rmsd, root-mean-square deviation; WT, wild-type; WT*, Cys-free pseudo WT; PEG, polyethylene glycol.

Introduction

One aspect of protein engineering is the design of proteins that bind a molecule of choice specifically (Eriksson *et al.*, 1992b; Fitzgerald *et al.*, 1994; McRee *et al.*, 1994; Perona *et al.*, 1994; Barrick *et al.*, 1997). Such tailored binding sites might have utility for detection of small organic molecules, and as starting points for developing enzyme-like catalysts. Antibodies and enzymes have highly complementary binding surfaces as a product of evolution. An evolution-like method of obtaining protein affinity molecules *de novo* involves combinatorial methods in conjunction with binding-selec-

tion protocols (Barbas *et al.*, 1992; Winter *et al.*, 1994; Ku & Schultz, 1995). Another strategy is to create a new site in a known structural scaffold using site-directed mutagenesis (Gonzalez *et al.*, 1996). Here, we investigate the possibilities and limitations of the latter approach.

Relatively few mutations in existing proteins have led to the non-covalent binding of small molecules in crystals or in solution (Eriksson *et al.*, 1992b; Fitzgerald *et al.*, 1994; McRee *et al.*, 1994; Perona *et al.*, 1994; Gonzalez *et al.*, 1996; Barrick *et al.*, 1997). In T4 lysozyme, truncating Leu99 to Ala creates a large internal cavity that binds benzene and a number of aromatic molecules with sub-millimolar affinities (Eriksson *et al.*, 1992a,b; Morton *et al.*, 1995; Morton & Matthews, 1995). Extensive structural and thermodynamic analysis indicated that binding resulted from the hydrophobic effect but that cavity flexibility, interactions between the ligand and the cavity, and changes in ligand entropy upon binding all modulate specificity. One implication of that work was that mutations that create large enough interior cavities should bind hydrophobic molecules of the appropriate size and shape. But what is the generality of this result? Can we create additional small molecule binding sites through other cavity-creating mutations? Alternatively we would like to create binding sites more like those in enzymes, antibodies and other receptors that typically bind ligands in solvent-accessible clefts. Can we use mutation to create a binding crevice? While desolvation can drive the binding of hydrophobic molecules, most biologically interesting ligands have polar moieties, which interact more favorably with water than a low dielectric environment. Furthermore, specificity in protein binding sites often is through directional polar interactions with the substrate. What kind of mutations can we use to create a polar molecule binding site and what are its ligands? Here, we address these questions by engineering both non-polar and polar binding sites in T4 lysozyme.

Results

Design of mutations

The T4 lysozyme mutant Leu99 → Ala contains a large internal cavity that binds benzene (Eriksson *et al.*, 1992b) and a number of other aromatic compounds (Morton *et al.*, 1995; Morton & Matthews, 1995). To determine if there is a simple relation between the size of a cavity and its ability to bind benzene we tested a large number of other mutants with large-to-small substitutions (Table 1). The survey includes all mutants in T4 lysozyme known to form interior cavities of volume in excess of 31 Å³ as well as a subset of the mutants with smaller cavities (Xu *et al.*, 1998). Mutants that form obvious surface crevices were also included.

We attempted to generate binding sites for ligands that were more polar in character. One

experiment was based on mutant M102A, which binds benzene (see below). By introducing the second mutation Met106 → Ala we attempted to connect this internal cavity to the surface, creating a deep hydrophobic crevice. We reasoned that such a variant might bind polar-substituted benzenes with the polar moiety at the surface and the phenyl group buried in the interior. In another experiment we mutated Arg95 to alanine. In the wild-type protein Arg95 is partially buried and makes a number of energetically important contacts with its highly complementary environment (see below). Because of these contacts we anticipated that the mutant Arg95 → Ala might bind a guanidinium moiety.

Screen for benzene binding

By screening a number of variants (Table 1), five new benzene binding mutants (Leu7 → Ala (L7A), M102A, F104A, L133G and the double mutant L121A/L133A) were identified. Binding resulted in both an increase in melting temperature (t_m), as well as the enthalpy of unfolding. The locations of those mutations for which ligand binding has been characterized structurally are shown in Figure 1. The thermodynamic parameters for the mutant lysozymes not described previously are given in Table 2.

For L99A, M102A, F104A, L133G/WT and L99A/F153A, where benzene binding was directly seen *via* crystallography, the average increase in t_m is +3.4 deg. C and in ΔH is +13 kcal/mol. L7A and L121A/L133A show comparable increases in both t_m and ΔH , suggesting that both variants would probably bind benzene at high occupancy in crystals. L118A, L121A and L133A have smaller increases in both t_m and ΔH (0.7 to 1.0 deg. C and 2 to 4 kcal/mol) in the presence of benzene but there was no evidence for binding in the crystals (data not shown). One other mutant, A73-[AAA], with three alanine residues inserted after Ala73, shows significantly increased t_m but reduced ΔH . This is the case at all benzene concentrations surveyed (data not shown). This mutant has a new declivity on its surface as a result of the insertion (Vetter *et al.*, 1996). It may bind benzene, although not necessarily at a unique site. For all other mutants tested for benzene binding, the average change in t_m was -0.15 deg. C and in ΔH was -3 kcal/mol. These values do not differ from zero by more than the estimated error of ± 0.2 deg. C and ± 6 kcal/mol, respectively.

Binding affinities

While changes of the melting temperature in the presence of ligand (Tables 1 and 2) are a sensitive assay for binding, thermodynamic parameters are difficult to extract (Shrake & Ross, 1992), especially for relatively insoluble, volatile compounds. Figure 2 shows for mutant L99A the dependence of the apparent equilibrium constant for thermal

Table 1. Thermal upshift screen for the binding of ligands to T4 lysozyme mutants

Mutant	t_m of ligand-free mutant (°C)	ΔH of ligand-free mutant at t_m (kcal/mol)	Ligand	Ligand concentration (mM)	Δt_m in presence of ligand (°C)	$\Delta\Delta H$ at t_m in presence of ligand (kcal/mol)	Binding detected to folded protein?
WT ^a	51.8	118	Benzene	10.4	−0.1	−5	No(No)
M6A ^b	46.0	95	Benzene	10	−0.4	−1	No
L7A ^b	43.6	90	Benzene	10	3.0	15	Yes
I17A ^c	43.3	87	Benzene	10	−0.4	−2	No
L33G ^c	37.5	50	Benzene	10	−0.1	−2	No
L39A ^d	48.4	103	Benzene	10	−0.5	−6	No
L46G ^e	37.0	50	Benzene	10	−0.8	−5	No
L46A ^f	43.2	90	Benzene	10	−0.4	−4	No
L66A ^b	38.3	69	Benzene	10	0.2	−2	No
F67A ^b	46.0	101	Benzene	10	−0.3	−3	No
A73-[AAA] ^g	30.0	32	Benzene	10	1.2	−1	Possibly
I78A ^b	47.0	105	Benzene	10	0.0	−3	No
L84A ^b	38.3	67	Benzene	10	−0.1	−6	No
L91A ^b	41.9	85	Benzene	10	−0.7	−4	No
R95A	36.2	nd	Benzene	10	−2.0	nd	No
R96A	46.9	93	Benzene	10	−0.9	−5	No
L99A ^{a,f}	36.2	82	Benzene	7.2	6.0	18	Yes(Yes)
I100A ^b	41.0	85	Benzene	10	−0.5	−4	No
M102A	41.0	78	Benzene	10	2.9	9	Yes(Yes)
V103A ^b	45.1	94	Benzene	10	−0.1	−5	No
F104A	42.0	82	Benzene	10	2.6	7	Yes(Yes)
F104A	42.0	82	Toluene	3	1.3	5	Possibly(No)
M106A	44.6	89	Benzene	10	−0.2	0	No
L118A ^f	39.6	76	Benzene	10	0.9	2	Possibly(No)
M120A ^h	51.1	112	Benzene	10	−0.3	−6	No
L121A ^f	42.5 ^g	81	Benzene	10	0.7	2	Possibly(No)
L133G/WT	37.9	50	Benzene	10	2.6	13	Yes(Yes)
L133G/WT	37.9	50	<i>p</i> -Xylene	satd (ca 0.5)	0.5	7	Yes(Yes)
L133A/WT ^f	43.0 ^g	92	Benzene	10	1.0	4	Possibly(No)
L133S/WT	39.2	64	Benzene	10	0.9	−1	Possibly
W138A	41.4	74	Benzene	10	0.4	−4	Unlikely
W138A	41.4	74	Indole	8	0.6	4	Possibly
F153A ^a	39.5	73	Benzene	7.2	−0.1	−2	No(No)
L46A/L121A ^b	31.8	51	Benzene	10	1.0	4	Possibly
L99A/F153A ^{a,g,i}	10.4	45	Benzene	10	3.0	18	Yes
M102A/M106A	36.5	65	Benzene	10	0.3	−8	No
L121A/L133A ^{b,j}	32.2	43	Benzene	10	5	17	Yes
WT ^{a,b,k}	65.3	133	Guanidinium hydrochloride	100	−0.4	4	No
R95A ^k	45.5	68	Guanidinium hydrochloride	20	1.5	15	Yes(Yes)
R95A ^k	45.5	68	1-Methyl- guanidinium hydrochloride	18	1.0	11	Yes
R96A ^l	57.6	104	Guanidinium hydrochloride	30	0.5	4	Possibly(No)

The buffer was as in Methods and Materials except as footnoted. All T4 lysozyme mutants were in the cysteine-free pseudo wildtype (C54T/C97A) background except as identified by/WT, which signifies the wild-type background. Thermal unfolding curves were modeled as two-state transitions except for R95A at pH 3 where t_m refers to the movement of the principal transition of a non-two-state system. For all others, t_m is the melting temperature and Δt_m the change in t_m in the presence of the putative ligand. ΔH is the enthalpy of unfolding of the mutant at its t_m . $\Delta\Delta H$ is the change in ΔH in the presence of putative ligand. The estimated error in Δt_m is $\pm 0.2^\circ\text{C}$ and ± 6 kcal/mol for $\Delta\Delta H$. The final column gives an overall assessment, based on Δt_m and $\Delta\Delta H$, as to whether or not the folded protein appears to bind the ligand in solution. In cases where binding was analyzed crystallographically the results are given in parentheses. As discussed in the text there can be differences in the approaches.

^a Eriksson *et al.* (1992b).

^b Xu *et al.* (1998).

^c Unpublished work of R. DuBose, W.A.B. and B.W.M.

^d Heinz *et al.* (1994).

^e Unpublished work of D. W. Heinz, W.A.B and B.W.M.

^f Eriksson *et al.* (1992a).

^g Vetter *et al.* (1996).

^h Blaber *et al.* (1995).

ⁱ Buffer was 0.025 M KCl, 0.015 M KH_2PO_4 , 0.003 M potassium citrate (pH 4.5).

^j Mutant L121A/L133A also binds hexane, diethyl disulfide, dipropyl disulfide and diisopropyl disulfide.

^k Measured in 0.15 M KCl, 0.010 M potassium phosphate (pH 6.7).

^l Measured in 0.2 M KCl, 0.010 M potassium phosphate (pH 6.7).

Figure 1. Stereo diagram of the backbone of T4 lysozyme showing the locations of bound ligands located by X-ray crystallography. The substitutions that were used to generate the binding sites are indicated with the α -carbon atom drawn filled.

unfolding, K_{app} , on the benzene concentration [Bz]. With the assumption of binding of one molecule of benzene to only the folded protein:

$$(1/K_{app}) = 1 + [Bz]/K_d$$

where K_d is the dissociation constant (Brandts, 1964; Eriksson *et al.*, 1992b). The value determined from Figure 2 ($K_d = 0.40$ mM) confirms that reported previously (Eriksson *et al.*, 1992b). Schellman (1976) has shown that the dependence of $(1/t_m)$ on the natural logarithm of ligand molarity is related to the difference between the number of ligands bound to the folded state and the number bound to the unfolded state. Such a plot (not shown) suggests that this difference is about 0.6 to 0.75. This indicates that there is some binding of benzene to unfolded L99A at pH 3.0.

This result also underscores one of the limitations of the thermal upshift assay, namely, a substantial increase in t_m and in ΔH is a firm indication of binding but an unchanged melting temperature does not exclude binding. Also, dissociation constants estimated as in Figure 2 are not strictly equivalent to those measured, for example, by NMR, which directly measures binding to the folded protein. The binding affinity of benzene to L99A measured here (0.40 mM) is, in fact, slightly less than that determined both by NMR (0.3 mM) and titration calorimetry (0.2 mM, 29°C; Morton *et al.*, 1995). The dissociation constants of benzene binding to M102A and F104A were estimated using NMR spectroscopy, at 10°C as 0.4 mM and 1.0 mM, respectively (Feher *et al.*, 1996). The change in t_m of L133G as a function of benzene

Table 2. Thermal stabilities for putative ligand-binding lysozymes

Mutant	t_m (°C)	ΔH (kcal/mol)	$\Delta\Delta G$ (kcal/mol)	$\Delta\Delta G_{tr}$ (kcal/mol)	$\Delta\Delta G_{cav}$ (kcal/mol)
A. Putative benzene-binding lysozymes and reference proteins at pH 3					
F104A	42.0	82	-3.1	-2.0	N/A
M102A	41.0	78	-3.3	-1.2	-1.5
M106A	44.6	89	-2.3	-1.2	N/A
M102A/M106A	36.5	66	-4.3	-2.4	N/A
WT*	51.7	114	-	-	-
L133G/WT	37.9	50	-5.6	-2.3	-2.7
WT	53.5	130	-	-	-
B. Guanidinium-binding lysozyme and WT* reference proteins at pH 5.4					
R95A	50.7	88	-4.9	-	-
WT*	65.3	133	-	-	-

The Table includes data only for variants not described previously (see references in Table 1). t_m is the melting temperature. ΔH is the enthalpy of unfolding at T_m and $\Delta\Delta G$ is the free energy of unfolding relative to the reference protein. $\Delta\Delta G_{tr}$ is the difference between the transfer free energies of the substituted amino acids, after Fauchère & Pliška (1983) and, for the cavity-creating mutants, $\Delta\Delta G_{cav}$ is the energy of cavity formation estimated as 24 cal M⁻¹ Å⁻³ (Eriksson *et al.*, 1992a).

Figure 2. Estimation of the affinity of binding of benzene to mutant L99A (see the text). The effect of benzene on the melting temperature of 1 μ M L99A was measured in 20 mM glycine-HCl, 1 mM EDTA at pH 3 (Carra *et al.*, 1996). The dissociation constant, K_d , is given by the relation $(1/K_{app}) = 1 + [Bz]/K_d$ and the straight line is the best least-squares fit to this relation using the data between 0 and 6 mM benzene.

concentration was similar to M102A and F104A, suggesting comparable affinity.

The thermal stability of R95A was evaluated under a variety of conditions. Low concentrations (0 to 40 mM) of guanidinium ion consistently increased t_m in all buffers at pH values above the pK_a of carboxylate, although by somewhat different amounts. The maximal effect, seen in 0.15 M KCl, 10 mM $K_{3/2}H_{3/2}PO_4$ (pH 6.7), was 2.7 deg. C at about 100 mM guanidinium chloride. Little change was detected in the t_m of the wild-type protein in the presence of 100 mM guanidinium hydrochloride (Table 1). In the above buffer, however, the thermal unfolding of R95A was not very reversible. Much better reversibility was observed in 0.1 M NaCl, 10 mM sodium acetate (pH 5.4)

and this buffer was therefore used to estimate $\Delta\Delta G$ in Table 2.

Crystal structures

The crystal structures were determined and refined to resolutions ranging from 1.8 Å to 2.1 Å with excellent stereochemistry (Table 3). Coordinates and structure factors have been deposited in the Brookhaven Data Bank for immediate release (I. D. codes 220L, 222L, 223L, 225L to 229L and 252L). In general, the conformations of most of the residues near the mutation sites were well-defined, with the exception of M102A/M106A, which had two conformations for residues 106 to 114 (see below). In the unliganded structures, the cavities had no observable excess electron density, while the crevices were occupied by bound solvent. As for the complex of L99A with benzene (L99A/BZ; Eriksson *et al.*, 1992b), the azimuthal orientation of the benzene moiety could not be ascertained in M102A/BZ or L133G/BZ. The electron density for the benzene moiety in F104A/BZ, however, was noticeably hexagonal (data not shown). Ligand thermal factors in the benzene and xylene complexes were higher than those of surrounding atoms or the deleted side-chains, whereas in R95A/GUD the bound guanidinium ion (GUD) had B -factors similar to those of the Arg95 guanidinium group.

Except for M102A/M106A, neither mutation nor ligand binding substantially perturbed the structure of the protein. Global positional and B -factor differences between the structures and wild-type, or between liganded and unliganded structures were generally small, 0.09 to 0.23 Å (main-chain rmsd) and -4 Å² to $+10$ Å². The positional and B -factor changes within 6 Å of the mutation sites were also modest (0.16 Å to 0.60 Å and -3 Å² to 7 Å²), excepting M102A/M106A. In most cases the structure surrounding the putative hydrophobic cavity or crevice shifted inward in the truncation mutant, and expanded outward in response to ligand binding (see below).

Table 3. Crystallographic data for unliganded and liganded mutant structures

Mutant	Name	Cell dimensions		Resolution (Å)	Completeness (%)	R_{merge} (%)	R_{iso} (%)	R (%)
		a, b (Å)	c (Å)					
Arg95 → Ala/guanidinium	R95A/GUD	60.6	97.5	∞ –1.8	92	3.3	16.9 ^{WT*}	16.0
Met102 → Ala	M102A	61.3	96.9	10–1.9	65	9.1	13.6 ^{WT*}	15.8
Phe104 → Ala	F104A	60.9	97.0	10–1.9	76	3.4	12.3 ^{WT*}	15.2
Leu133 → Gly	L133G/WT	61.2	96.0	10–1.8	56	N/A	19.4 ^{WT}	17.5
Met102/Met106 → Ala	M102A/M106A	60.8	97.2	∞ –2.1	85	4.9	19.7 ^{WT*}	16.3
Met102 → Ala/benzene	M102A/BZ	60.9	97.2	10–1.85	72	3.3	15.1 ^{M102A}	14.9
Phe104 → Ala/benzene	F104A/BZ	60.8	97.8	10–2.0	85	4.9	11.8 ^{F104A}	15.2
Leu133 → Gly/benzene	L133G/BZ	61.2	96.3	20–1.9	88	2.9	13.5 ^{L133G}	16.8
Leu133 → Gly/ <i>p</i> -xylene	L133G/XL	61.1	96.8	20–1.9	83	3.8	14.6 ^{L133G}	16.0

Mutants generated and named as in Table 1. All crystals were isomorphous with the wild-type, space group $P3_121$, cell dimensions $a = b = 60.9$ Å, $c = 96.8$ Å (Weaver & Matthews, 1987). $R_{merge} = \sum |I - \langle I \rangle| / \sum I$, where I are intensities of equivalent reflections. R_{iso} gives the average agreement between the observed structure factors and the reference data set, which is denoted by the superscript. R is the crystallographic residual for the refined structure. The deviations from ideal geometry were better than or equal to 0.015 Å for bond lengths and 2.1° for bond angles.

Table 4. Changes in cavity volume in mutant structures relative to wild-type

Structure	Volume of cavity in WT or WT* at site of mutation $V_c(\text{WT})$ ($\text{\AA}^3$)	Volume of cavity in mutant structure $V_c(\text{mut})$ ($\text{\AA}^3$)	Increase in cavity volume due to mutation $V_c(\text{mut}) - V_c(\text{WT})$ ($\text{\AA}^3$)	Expected cavity volume assuming no change in structure $V_{\text{mod}}(\text{mut})$ ($\text{\AA}^3$)	Reduction in cavity volume due to collapse of WT structure $V_{\text{mod}}(\text{mut}) - V_c(\text{mut})$ ($\text{\AA}^3$)	Binding of benzene
M6A ^a	0	58 ^b	58	71	13	No
L46A ^c	0	24	24	45	21	No
L99A ^c	34	149	116	130	-20	Yes(Yes)
I100A ^a	0	29	29	23	-6	No
M102A ^a	34	83	49	109	26	Yes(Yes)
V103A ^a	0	17	17	12	-5	No
L118A ^c	34	105	71	N/A ^c	N/A	Possibly(No)
L121A ^d	34 + 20 ^d	67 + 8 ^d	21 ^d	145	70	Possibly(No)
L133A ^c	41	142	101	169	27	Possibly(No)
L133G	41	176	136	194	18	Yes(Yes)
F153A ^c	34	101	67	132	31	No(No)

Wild-type lysozyme contains two voids in the carboxy-terminal domain, designated Cavity I and Cavity II (Eriksson *et al.*, 1992a; Xu *et al.*, 1998). In mutants L99A, M102A, L118A and F153A, Cavity I is incorporated within the cavity created by the mutant. In mutants L121A, L133A and L133G both Cavity I and Cavity II are incorporated. Cavity volumes, V_c , were calculated using the crystallographic coordinates of the wild-type or mutant structures (see the text). Cavity values, $V_{\text{mod}}(\text{mut})$, were obtained by truncating the side-chain from coordinates of the reference wild-type structure, and give the volume expected if there were no change in coordinates. Values of V_{mod} quoted here differ slightly from those given by Eriksson *et al.* (1992a) because of a change in the van der Waals radius of the C ^{β} atom (Xu *et al.*, 1998). The ability to bind benzene is based on upshift of t_m . Where an independent assessment is available from crystallographic experiments it is given in parentheses, as in Table 1.

^a Data from Xu *et al.* (1998).

^b Volume calculated in the absence of bound solvent.

^c Data from Eriksson *et al.* (1992a).

^d Data from Baldwin *et al.* (1996).

Met102 → Ala and

Met102 → Ala/benzene complex

If, as a model calculation, the side-chain of Met102 in the wild-type structure is truncated to alanine, the resultant cavity is 109 $\text{\AA}^3$ (Table 4). This includes 34 $\text{\AA}^3$ from a pre-existing cavity in the native protein (Table 4; Eriksson *et al.*, 1992a). The modeled cavity is benzene-shaped and includes a “dimple” created by the Leu133 C ^{$\delta 2$} methyl group at the ring center. In the M102A structure, however, inward shifts of the Met106, Phe114, Trp138 and Phe153 side-chains reduced the actual cavity size to 83 $\text{\AA}^3$ (Figure 3(a); Table 4). This volume is actually somewhat less than that of benzene (Table 5). Met106 changed torsion angles ($\chi_1 = 79^\circ$ (WT) to 41° (M102A), $\chi_3 = 113^\circ$ (WT) to -168° (M102A)). The Phe114 ring rotated 25° , the Trp138 ring tipped 7° and the side-chains of L121 and Phe153 translated 0.4 $\text{\AA}$ and rotated towards the cavity. Helices 93-104 and 143-154 rotated slightly towards each other and shifted inward, moving the 102 C ^{β} atom and the 149 C ^{$\gamma 1$} atoms 0.3 $\text{\AA}$. The side-chains of Asp10, Asn101, Gln105, Arg145, Arg148, Tyr161 and Sol208, which form an extensive buried polar network, shifted overall by 0.3 $\text{\AA}$.

Notwithstanding the reduction in volume of the cavity in mutant M102A, this variant does bind benzene (Figure 3(b)). In the complex (M102A/BZ) the residues lining the cavity shifted back towards their wild-type positions, particularly Phe114 and Phe153, expanding the cavity

by 39 $\text{\AA}^3$ (Figure 3(c); Table 5). The benzene site is an “aromatic box” formed by Phe114, Trp138 and Phe153, and is precisely benzene-shaped (Figure 3(d)). The ligand binds in the expected orientation, with the ring dimple resting on the Leu133 methyl group. The edge of the benzene moiety contacts the center of Phe114 ring with an interplanar angle of 52° , making a canonical ring-ring interaction (Burley & Petsko, 1985). The complex structure, overall, is very similar to WT, but is slightly expanded (13 $\text{\AA}^3$; Table 5) around the bound ligand due to outward rotations of helices 93-105, 108-113 and 116-123. The root-mean-square shift in main-chain atoms for residues 1 to 161 is 0.13 $\text{\AA}$, and 0.19 $\text{\AA}$ for all atoms with B -factors less than 50 $\text{\AA}^2$.

Leu133 → Gly and its complexes with benzene and *p*-xylene

In L133G, truncation of Leu133 to Gly fused the two wild-type cavities into a single elongated 176 $\text{\AA}^3$ void (Table 4), which included two narrower lobes on opposite sides, one near Met102 and the other near Leu133. Residues 106 to 121 shifted as a unit towards the core (0.25 $\text{\AA}$; Figure 4(a)). The Phe114 and Ile150 side-chains shifted towards the mutation site 0.4 $\text{\AA}$ and 0.3 $\text{\AA}$, respectively. Some side-chains lining the cavity had increased side-chain B -factors (Met102, +7 $\text{\AA}^2$; Ile150, +7 $\text{\AA}^2$; Phe153, +15 $\text{\AA}^2$).

In the benzene complex L133G/BZ, Leu99, Met102, Ser117, Leu121, Ile150 and Phe153 side-

Figure 3. (a) Superposition of the structure of mutant M102A (filled bonds) on WT* (open bonds). The asterisk (*) marks the site of the mutation. (b) Difference electron density map calculated with coefficients $F_0(\text{M102A/BZ}) - F_0(\text{M102A})$ and phased with the refined coordinates for M102A. Contours drawn at $\pm 4\sigma$ where σ is the root-mean-square density throughout the asymmetric unit. (c) Superposition of the benzene complex M102A/BZ (filled bonds) on M102A (open bonds), based on the main-chain atoms of residues 81 to 161. (d) Molecular surface surrounding the benzene binding site in the M102A/BZ complex.

Table 5. Changes in cavity volume on ligand binding

Structure	Volume of cavity in mutant structure $V_c(\text{mut})$ ($\text{\AA}^3$)	Expected cavity volume assuming no change in structure $V_{\text{mod}}(\text{mut})$ ($\text{\AA}^3$)	Unoccupied cavity volume in complex $V_c(\text{cplx})$ ($\text{\AA}^3$)	Calculated volume of cavity in complex with ligand removed $V_{\text{mod}}(\text{cplx})$ ($\text{\AA}^3$)	Expansion of cavity induced by ligand $V_{\text{mod}}(\text{cplx}) - V_c(\text{mut})$ ($\text{\AA}^3$)	Expansion in WT from removal of sidechain and replacement with ligand $V_{\text{mod}}(\text{cplx}) - V_{\text{mod}}(\text{mut})$ ($\text{\AA}^3$)	Difference between volume of cavity and volume of ligand $V_c(\text{mut}) - V_{\text{lig}}$ ($\text{\AA}^3$)
L99A/BZ complex	149	130	32	167	18	47	53
M102A/BZ complex	83	109	0	122	39	13	-13
L133G/BZ complex	176	194	8	199	23	5	80
L133G/XL complex	176	194	0	225	49	31	38

As in Table 4, cavity volumes, V_c , were calculated from the crystal structures, in this case of the complexes. $V_{\text{mod}}(\text{cplx})$ is the cavity volume in the crystal structure of the complex when the ligand atoms are deleted. V_{lig} is the van der Waals volume of the ligand, 96 $\text{\AA}^3$ for benzene and 138 $\text{\AA}^3$ for xylene, determined by molecular surface calculated with a 1.2 $\text{\AA}$ probe (Connolly, 1983). Volumes for the complex of L99A with benzene were calculated from the coordinates given by Eriksson *et al.* (1992b).

Figure 4. (a) Superposition of mutant L133G (filled bonds) on WT (open bonds) based on the main-chain atoms of residues 81 to 161. (b) Superposition of the structure of the benzene complex L133G/BZ (filled bonds) on the apo mutant L133G (open bonds).

chains and helix 108–113 shifted 0.3 to 0.4 Å away from the ligand, while Phe114 and Leu118 shifted 0.3 Å towards it (Figure 4(b)). The benzene moiety stacked against the Phe153 ring, making an interplanar angle of 22°. The Phe153 phenyl ring became better ordered (the overall *B*-factor decreased 10 Å²) while the mobility of Leu121 increased (*B*-factor increased 12 Å²). The cavity was not completely filled, with two empty regions remaining on either side of the bound benzene moiety (not shown).

In the *p*-xylene complex L133G/XL, the ligand ring retained its orientation, but translated 0.3 Å relative to the position occupied by the benzene moiety (Figure 5(a) to (c)); cf. Morton & Matthews, 1995). The regions that remained unoccupied in the benzene complex contained the *para*-methyl groups of the ligand. Although the molecular volume of *p*-xylene (138 Å³) is larger than that of benzene (96 Å³), the shifts and mobility changes were quite similar to L133G/BZ, except for the greater outward shifts of Met102 (0.5 Å) and Leu121 (0.4 Å) and a slight outward shift of helix 108–113 in the xylene complex (0.2 Å).

Phe104 → *Ala* crevice mutant and its complex with benzene

In F104A, removal of the Phe104 side-chain (92% buried) created a narrow, deep crevice (about 3 Å × 4 Å × 6 Å) roughly perpendicular to the plane of the 104 phenyl ring (Figure 6(a)). Adjacent residues rearranged slightly to both shorten and slightly expand the opening compared to WT.

Leu7, His31, Leu32 and Asp70 translated towards the crevice, while residues of the long helix 70–80 actually shifted away 0.3 Å (main-chain rms) moving the C^α atom of Ala74 by 0.4 Å. Other shifts propagated up to 15 Å from the Phe104 ring, into the region of residues 105 to 113 (0.7 Å). The Glu108 side-chain rotated towards the cleft, moving 1.9 Å, Thr109 and Gly110 translated 0.3 Å towards the site, and the C^ε atom of Met106 rotated 180° towards the active site. One new solvent occupied the cleft making a hydrogen bond with the carbonyl group of Ile29 (Figure 7(a) and 7(b)). Two additional water molecules (Sol265 and Sol267 in Figure 7), which are part of an extensive solvent network in the active site of the WT* protein, shifted towards the mouth of the crevice. Another solvent formed hydrogen bonds with the carbonyl oxygen atoms of Val103 and Ala104.

In the complex structure, F104A/BZ (Figure 6(b)), benzene bound in the crevice “edge-on” packing between Asp70 and Ala104 (Figure 6(c)). The C^β atom of Asp70 shifted outward by 0.4 Å accompanied by a translation of the long helix 60–80 (0.3 Å main-chain rmsd) that expanded the crevice on one side. These shifts were coupled to a rigid-body “hinge” motion of the N-terminal domain that opened the active site by about 1°. The benzene moiety is 89% buried and apparently rotationally restrained, since the electron density is distinctively hexagonal (not shown). The ligand ring plane is rotated about 70° with respect to the Phe104 phenyl group in WT*. Four solvent molecules in or near the mouth of the crevice (Sol265, Sol267, Sol299 and Sol401 in

Figure 5. (a) Difference electron density map for the complex of *p*-xylene with L133G, calculated with amplitudes $F_0(\text{L133G/XL}) - F_0(\text{L133G})$ and phases from L133G. Contoured at $\pm 3\sigma$. (b) Superposition of the *p*-xylene complex L133G/XL (filled bonds) on L133G (open bonds). (c) Molecular surface in the vicinity of *p*-xylene (filled bonds) bound in the L133G/XL complex. Surfaces drawn with 10 dots per $\text{\AA}^2$. The location of benzene bound in the L133G/BZ complex is also shown.

Figure 7(c)), were ejected, while other nearby water molecules maintained their positions, with Sol266 contacting the bound benzene moiety.

Arg95 → *Ala*/guanidinium complex

In the wild-type lysozyme Arg95 makes a variety of favorable interactions: (i) a salt-bridge with Asp92, which is estimated to contribute 2.5 kcal/mol of electrostatic stabilization (Anderson *et al.*, 1993) and orients Asp92 to make an N-cap interaction with helix 93-105 (Nicholson *et al.*, 1991); (ii) bidentate C-cap hydrogen bonds with helix 143-154; and (iii) a stacking interaction with Trp126. We were unable to grow crystals of R95A (see below). In the crystals of the complex, the guanidinium ion was 83% buried in a shallow crevice essentially replacing the corresponding moiety of the Arg95 side-chain (Figure 8(a) and 8(b)). The

pattern of hydrogen-bonding was preserved, with analogous ligand atoms participating in the salt-bridge with Asp92 O^{δ1} and a bidentate C-cap interaction with helix 143-154 *via* the carbonyl oxygen atoms of Phe153 and Arg154 (Figure 9(a) and 9(b)). The ligand rotated azimuthally by 25° relative to Arg95, moving the N1 atom 0.9 Å away from O154 and toward Asp92 O^{δ1}. The guanidinium plane is stacked on the indole ring of Trp126 but the planes of both groups have rotated 10° towards residue 157 while maintaining the same relative orientation as in wild-type. Residues 152 to 155 shifted (0.4 Å main-chain rmsd) towards the ligand moving the 153 carbonyl oxygen atom 0.7 Å. Residues 91 to 93 shifted 0.7 Å toward the ligand, moving the O^{δ1} atom of Asp92 closer to the guanidinium nitrogen atom by 0.4 Å. The ligand preserves contacts with surface solvent molecules Sol238 and Sol277, which mediate a contact with

Figure 6. (a) Superposition of mutant F104A (filled bonds) on WT* (open bonds) based on the main-chain atoms of residues 81 to 161. (b) Superposition of the benzene complex F104A/BZ (filled bonds) on the apo mutant structure F104A (open bonds). (c) Molecular surface in the vicinity of benzene bound in the F104A/BZ complex, with dots drawn at 5 per Å².

Thr21 from a symmetry-related molecule in the crystal (Figure 9(b)). This contact explains the difficulty in crystallizing R95A in the absence of guanidinium ions. An additional solvent makes hydrogen bonds with the carbonyl group of Thr155.

Met102 → Ala/Met106 → Ala

The model-building and refinement of this mutant was not straightforward (see Methods and Materials). The final model has two different conformations for residues 106 to 114 (Figure 10(a))

Figure 7. Changes in solvent structure near residue 104. Solvents are denoted by starred atoms and the broken lines connect solvent atom pairs that are within hydrogen bonding distance. (a) WT* lysozyme; (b) mutant F104A; (c) benzene complex F104A/BZ.

and 10(b)), one is close to wild-type (0.5 Å rmsd, weighted at 0.4; [Figure 10\(b\)](#)), while the predominant alternative ([Figure 10\(a\)](#)) is completely different (3.9 Å rmsd, weighted at 0.6). For the alternative conformer, helix 108-113 has completely unraveled into a non-helical S-shaped structure with no apparent secondary main-chain H bonds ([Figure 10\(b\)](#)). The conformation of residues 106 to 114 was most similar to that of residues 189 to 197 in flavocytochrome C sulfide dehydrogenase (PDB #1FCD, main-chain discrepancy 1.3 Å). Gly107 changed its backbone conformation from “disallowed” to left-handed helical, moving 3.9 Å. Two other glycine residues, Gly110 and Gly113, chan-

ged from the right-handed to the left-handed helical conformation. Glu108, Thr109, Val111 and Ala112 changed from α -helical to near β conformations. These alterations of the backbone essentially shifted the register of residues 108 and 109 such that Glu108 occupies the position of Gly107 in WT*, and Thr109 the position of Glu108. Gly113 and Ala112 moved 3.3 Å and 6.1 Å, respectively, away from the interior and towards the active site. Ala106 also shifted to make helical hydrogen bonds, extending the preceding helix 93-105 by one residue. Shifts propagated over 10 Å from the rearranged region and included translations of the side-chains of Phe104 (0.5 Å) and Thr115 (0.3 Å).

Figure 8. (a). Difference map for the complex of guanidinium with R95A calculated with amplitudes $F_0(\text{R95A/GUD}) - F_c(\text{WT-95})$ where $F_c(\text{WT-95})$ is the structure amplitude of wild-type lysozyme with the side-chain of Arg95 removed. The phases were from the same model. Contoured at $\pm 3\sigma$. (b) Superposition of the guanidinium complex R95A/GUD (filled bonds) on WT* (open bonds) based on the main-chain atoms of residues 81 to 161.

Residues 70 to 90 shifted 0.6 Å toward the back side of the molecule away from the active site. In spite of the disorder of 106 to 114, residues lining the mutation sites did not have greatly increased *B*-factors.

Both conformers formed a tubular channel into the center of the protein, lined by a number of hydrophobic residues that are buried in the wild-type. In the case of the alternative conformer, the channel curved back and exited to the surface in the vicinity of residues Leu84, Gly110 and Leu118. The channel contained a chain of new solvent molecules (Figure 10(a)), connecting to a network of solvents in the active site cleft.

Mutant M102A/M106A provides further support for the role of tertiary packing interactions in determining secondary structure. The substitutions M102A and M106A, in concert, weaken the interactions between α -helix 108-113 and the remainder of the protein. As a result, even though residues 108 to 113 retain the same sequence, they adopt a very different conformation in the "alternative" mode. This same region displayed yet another conformation in the mutant with three alanine residues inserted after Ala73 (Vetter *et al.*, 1996), suggesting that it may be especially sensitive to its local environment.

Stabilities of mutant proteins

Consistent with previous observations for buried residues (Kellis *et al.*, 1988; Shortle *et al.*, 1990; Eriksson *et al.*, 1992a; Buckle *et al.*, 1996; Xu *et al.*,

1998), all the substitutions were destabilizing (Tables 1 and 2). For the cavity-creating mutants, the stability changes can be adequately described by the change in solvent-transfer energy plus a cavity penalty (Eriksson *et al.*, 1992a; Xu *et al.*, 1998). For the crevice-creating substitution Phe104 → Ala, the stability change was comparable to truncation of a buried Phe (Xu *et al.*, 1998).

Truncation of the polar charged side-chain Arg95 → Ala (89% buried, polar atoms are 83% buried) leads to a large destabilization (−4.5 kcal/mol), indicating that this side-chain interacts more favorably with its folded environment than with solvent (see Discussion).

The double mutant M102A/M106A is destabilized (−4.3 kcal/mol), somewhat less than the sum of the individual destabilizations of M102A and M106A (−5.6 kcal/mol).

Discussion

Model binding sites created by mutations can provide a useful probe of ligand-protein interactions, in part because they are created free of selection for binding. The screen of putative ligand-binding mutants (Table 1) indicates that the primary requirement for the binding of a hydrophobic compound such as benzene is a non-polar cavity or crevice that is large enough and, through reasonable shifts, similar to the shape of the ligand. The mutants with the smallest cavities (M6A, L46A, I100A and V103A) do not appear to bind benzene; those with the largest cavities (L99A,

Figure 9. Conservation of the hydrogen bonding pattern of Arg95 in wild-type by guanidinium in R95A/GUD. The continuous lines denote the hydrogen bonds and the numbers indicate the distances in Å. The atom types are carbon (open), nitrogen (starred) and oxygen (filled). (a) Arg95 in WT* lysozyme; (b) guanidinium in the R95A/GUD complex.

L133G) do (Table 4). L133A, however, has a cavity volume of 142 Å^3 , considerably larger than the volume of benzene (96 Å^3) but appears not to bind the ligand. This indicates that some sites have limited flexibility and/or inappropriate shapes and cannot be sufficiently compensated by binding interactions to drive structural reorganization. In the case of L133A, the Ala methyl group apparently blocks binding, since truncation to Gly led to binding of benzene, and in the L133G/BZ complex, the benzene moiety overlapped the volume formerly occupied by the methyl group of Ala133. The distribution of unfilled space indicated that a para-disubstituted benzene (e.g. *p*-xylene) should be accommodated and, in L133G/XL, *p*-xylene was observed to bind in the expected orientation with only minor adjustments (Figure 5(b) and (c)).

Other sites appear to be more flexible. The cavity of M102A appears somewhat small to accommodate benzene (83 Å^3), but does, in fact, bind it. The site expanded *via* rotation of a relatively unconstrained surface residue, Phe114, and other small

adjustments, yielding a structure that was similar to that of the wild-type. Plasticity in the region of helix 108-113 was also observed in mutant L99A in order to accommodate a number of non-polar ligands (Morton & Matthews, 1995).

Similar to M102A, the crevice in F104A appeared too narrow to accommodate benzene but expanded in order to do so (Figure 6(a)).

Ligand orientation in the hydrophobic sites appears to be a function of the cavity shape rather than the original side-chain position. The benzene orientation in F104A/BZ is not the same as the phenyl ring of Phe104. This is for two reasons: (i) if a benzene ring were bound in this orientation it would collide with the 104 C β atom (the more appropriate mutant would be Phe104 $\rightarrow$ Gly); and (ii) because of existing packing defects that surround the Phe104 ring, the empty site is not a "negative image" of the side-chain but is in fact crevice-like.

The guanidinium ion in R95A/GUD occupies the position of the Arg95 head group in the wild-type (Figure 8(b)). This reflects the greater directionality of polar (hydrogen-bonding) relative to non-polar interactions. The binding of the guanidinium ion also shows that a specific polar binding site can be created by removing partially or fully buried polar side-chains. In this case, the overall binding affinity is low but not surprising considering the charge and polarity of the ligand, which result in an unfavorable transfer energy.

The double mutant M102A/M106A could not bind benzene because structural rearrangements distorted the M102A binding site. The change in conformation is likely the result of lost packing interactions. In the wild-type, Met102 and Met106 pack against Gly107, Gly110, Phe114 and Val111, the latter being the only hydrophobic "anchor" residue in helix 108-113 (Heinz *et al.*, 1993). In the wild-type crystal structure, this helix is somewhat irregular and has high main-chain *B*-factors (Weaver & Matthews, 1987) and low protection factors in NMR deuterium exchange experiments (Lu & Dahlquist, 1992). Presumably, the native helix conformation was destabilized enough by the removal of contacts with Met102 and Met106 to allow population of another conformer that is more favored. Loss of contacts with either residue alone cannot be responsible, since Met102 $\rightarrow$ Ala or Met106 $\rightarrow$ Ala substitutions alone do not significantly perturb this region (Xu *et al.*, 1998). Likewise, loss of the Val111 contacts is not entirely responsible, since helix 108-113 is retained in the mutant Val111 $\rightarrow$ Ala (Xu *et al.*, 1998). In this case, the slight preference for the alternative conformer is hard to reconcile, since more hydrophobic surface area is exposed. Stabilizing influences might include relief of conformational strain. The wild-type helical conformation may be favored only marginally due to the high glycine content of this region: Gly107, Gly110, Gly113 all adopted the left-handed configuration in M102A/M106A and

Figure 10. (a) Superposition of M102A/M106A (filled bonds) on WT* (open bonds) showing the "alternative" conformation with hydrogen bonds indicated by broken lines. (b) Superposition of residues 106 to 114 of the "alternative" conformation of M102A/M106A (filled atoms and bonds), together with the "wild-type-like" conformation of M102A/M106A (thin open bonds and starred atoms), on WT* (open thick bonds and atoms), based on atoms 81 to 161.

Gly107 shifts from a "disallowed" conformation (Laskowski *et al.*, 1993).

Solvent structural changes

As has been observed for most other non-polar cavities, the cavities of M102A and L133G lack any discernible electron density that might suggest occupancy by solvent or other small molecules. This is not surprising, because there is no unsatisfied hydrogen bond donor or acceptor within the cavity (Wolfenden & Radzicka, 1994; Matthews *et al.*, 1995).

In contrast, the crevice of F104A and the channel of M102A/M106A contain several ordered solvent molecules. In both cases, these water molecules form a path to the surface, hydrogen bonding to a network of solvents that is conserved in this crystal form. However, while the new solvent bound in F104A also makes a hydrogen bond with a free carbonyl group in the engineered crevice, the chain of solvents in M102A/M106A makes no direct protein hydrogen bond. They are, however, relatively well ordered ($\langle B \rangle = 30 \text{ \AA}^2$). Perhaps steric restrictions from van der Waals contacts with the wall of the channel restrict the mobility of these solvent molecules.

A general approach to ligand binding

This study suggests a generalized way to create ligand binding sites. Truncation of a buried residue will leave behind an environment that is complementary to the deleted side-chain. An appropriate ligand can then complement the engineered deficiency.

Generally, large-to-small hydrophobic substitutions will create low dielectric voids. These can

sequester hydrophobic molecules of complementary shape and size, with binding driven by favorable solvent transfer energies.

A polar residue, if buried, usually participates in an extensive polar network and is frequently saturated with hydrogen bonds and charge neutralized to compensate for the loss of solvation upon folding. Truncating such side-chains results in a natural site for the analogous ligand. Arg95 is an example in point. In the wild-type structure each nitrogen atom of the guanidinium side-chain donates at least one hydrogen bond to a complementary acceptor on the protein (Figure 9(a)). In such a case a guanidinium ion binds in place of the truncated side-chain. In contrast, only two of the three guanidino nitrogen atoms of Arg96 donate hydrogen bonds to the protein (see Figure 3 of Weaver *et al.*, 1989) and in this case the binding of guanidinium ion to R96A is weak, if at all (Table 1).

For non-polar ligands, the balance between solvent transfer energetics, steric interactions, van der Waals interactions and ligand and site flexibility determine affinity and specificity (Morton *et al.*, 1995). For polar ligands, charge-charge, charge-dipole and dipole-dipole interactions need to be considered. Different cavities and crevices, as well as ligands, have variable limits to flexibility that makes their binding preferences difficult to predict accurately. Indeed, binding of a non-polar compound generally causes the cavity or crevice to expand, while in the case of R95A the surrounding atoms actually moved towards the bound guanidinium ion. Some docking procedures attempt to account for site flexibility by relaxed distance tolerances or soft potentials (Kuntz *et al.*, 1994; Knegt *et al.*, 1997). A preferred model for cavities that bind hydrophobic compounds might be a truncated model of the wild-type protein. Such a model

will (i) usually have a larger cavity than the actual variant since most cavities are collapsed relative to wild-type (Table 4); and (ii) indicate the direction that the cavity might be likely to “flex”, since the wild-type side-chain is presumably accommodated with little strain (Baldwin *et al.*, 1996). The marked similarity of the M102A/BZ structure to WT* supports this idea.

Buried polar and non-polar truncating substitutions could, in principle, be combined to create extended sites for more complicated ligands with polar and non-polar regions. However, there are two primary limitations. First, such mutations are usually destabilizing (e.g. see Eriksson *et al.*, 1992a; Shoichet *et al.*, 1995; this work). Our experience has been that T4 lysozyme mutants destabilized to more than about 8 kcal/mol are difficult to work with. If destabilizations are additive, the practical limit is therefore likely to be three to five substitutions or thereabouts. A second limitation that is more difficult to anticipate is the stability of local secondary structure that may restrict the allowed substitutions, as observed in the M102A/M106A variant.

Methods and Materials

Mutant proteins

The substitutions were introduced into an M13 clone containing either the wild-type (WT) or Cys-free pseudo wildtype (WT*) T4 lysozyme gene (Matsumura & Matthews, 1989) using oligonucleotide-directed mutagenesis (Kunkel *et al.*, 1987). Mutant proteins were expressed and purified as described (Muchmore *et al.*, 1989; Poteete *et al.*, 1991).

Thermal data

Procedures for detection of ligand binding were as described (Morton *et al.*, 1995) except that the ligand was dissolved with gentle stirring in sealed optical cells for at least 60 minutes immediately prior to thermal denaturation. For the putative benzene-binding mutants (Table 1), measurements were made in 25 mM KCl, 17 mM KH_2PO_4 , 3 mM H_3PO_4 (pH 3.0). For the putative guanidinium ion binding mutants, pH 6.7 buffers were used (Table 1). Protein concentrations were 0.5 μM to 1.5 μM .

Protein stabilities (Table 2) were determined as described (Eriksson *et al.*, 1993), except that 0.1 M NaCl, 10 mM NaOAc (pH 5.4), was used for R95A and its reference protein, WT*. Reversibility was over 95% at pH 3.0 and about 80% at pH 5.4. For L133G/WT, $\Delta\Delta G$ was calculated relative to WT at 37.9°C using a ΔC_p of 1.8 kcal/mol deg. The data for this mutant could be fit with a two-state model but only poorly. $\Delta\Delta G$ values for the other benzene-binding mutants were calculated relative to WT* at 44°C using a ΔC_p of 1.8 kcal/mol deg. $\Delta\Delta G$ for R95A was calculated relative to WT* at 58°C using a ΔC_p of 2.5 kcal/mol deg.

Crystal growth, data collection and refinement

Crystals in space group $P3_221$ were grown at 4°C in hanging drops in ~2 M mixed sodium and potassium phosphate (pH ~7.0) as described (Eriksson *et al.*, 1993) or, for R95A/GUD, from 100 mM potassium phosphate,

100 mM sodium chloride, 20% (w/v) PEG 3350 (pH 7.1) in the presence of 250 mM guanidinium chloride. Crystals were mounted in glass capillaries and volatile hydrophobic ligands were introduced by vapor diffusion (Eriksson *et al.*, 1992b; Morton & Matthews, 1995), or by placing liquid ligand in contact with a plug of mother liquor containing the crystal one to two weeks. Trigonal crystals of Arg95 → Ala were not obtained in the absence of ligand under a variety of conditions. Crystals in space group $P2_1$ that diffracted to 4 to 5 Å were obtained but could not be further improved.

X-ray data were collected from a single crystal at room temperature using a Xuong-Hamlin area detector (Hamlin, 1985) or, in the case of M102A and L133G, from two crystals using film and processed as described (Schmid *et al.*, 1981). Data were no longer collected when $(I/\sigma(I))$ exceeded 2 (area detector) or 3 (film) for the highest-resolution shell.

Structures were determined by difference Fourier methods and refined with TNT (Tronrud *et al.*, 1987) starting with the WT or WT* model, essentially as described (Baldwin *et al.*, 1996). The ligand molecule positions could be ascertained from the $F_{\text{obs}}(\text{complex}) - F_{\text{obs}}(\text{apo})$ difference maps or $F_{\text{obs}}(\text{complex}) - F_{\text{calc}}(\text{apo})$ maps after rigid-body refinement.

Refinement of M102A/M106A

The M102A/M106A structure required special attention and was solved as follows. The $F_{\text{obs}}(\text{mut}) - F_{\text{obs}}(\text{WT}^*)$ map indicated that residues 106 to 114 had rearranged, but their positions were difficult to discern. These residues and all solvent molecules were deleted from the starting model and the atomic B -factors set to the Wilson B (21 Å²). Rigid-body refinement at 4 Å resolution was performed first with the whole molecule and then with five segments. The 16 segments that define the secondary structure elements were then refined using data to 3.3 Å resolution. Positional refinement of individual atoms was then performed using 3 Å data. At this point, residues 106 to 109 and 112 to 114 were modeled into difference density in conformations that were completely different from wild-type. Several trial structures were built. After further refinement at 3 Å, residues 110–111 were added. After model building and positional refinement at 3 and 2.5 Å resolution, followed by positional and restrained thermal factor refinement at 2.1 Å, the different trial structures essentially agreed. Solvent atoms, except those in the rearranged region, were placed followed by more refinement. At this point positive difference density in an omit $F_{\text{obs}} - F_{\text{calc}}$ map was well fit by Thr109, Val111 and Ala112 in the wild-type conformation. The wild-type-like conformation for residues 106–114 was initially traced from a map in which occupancies of the alternative conformation were set to 0.5 and the B -factors for all atoms were reset to 21 Å². Further positional and thermal factor refinement was performed in which the B -factors for just residues 106 to 114 in both conformations were restrained to the TNT BCORRELS library (Tronrud, 1996). Model building was aided using two different $F_{\text{obs}} - F_{\text{calc}}$ difference maps in which residues in each conformation were omitted. The final such maps are shown in Figure 11(a) and (b). While there is only weak positive difference density ($< 2\sigma$) for Gly110 in either conformation and for Val111 in the alternative conformation, the geometry of the adjacent well-defined residues was sufficient to constrain their main-chain conformations. The side-chain conformation

Figure 11. (a) Omit map showing the electron density and structure for the "alternative" conformation of residues 106 to 114. The map was calculated with amplitudes $F_o(M102A/M106A) - F_c(\text{omit})$ where F_c are the structure amplitudes for the model of the M102A/M106A structure with residues 106 to 114 in the alternative conformation deleted (see the text). Contours drawn at $\pm 2.5\sigma$. (b) Omit map, following Figure 10(a), in which residues 106 to 114 in the "wild-type-like" conformation for the M102A/M106A structure were deleted from the structure factor and phase calculation. In this case the density and the model correspond to the wild-type-like conformation. Contours drawn at $\pm 3\sigma$.

of Val111 was not well determined. The relative occupancies of the mutant and wild-type conformations were estimated to be 0.6 and 0.4 by both refinement of group occupancies and *B*-factors for each conformation, and by simultaneous refinement of individual positions, thermal factors and group occupancies.

The model retained 101 of the wild-type solvent atoms and added 36 new ones, 11 of which contact the rearranged residues. In addition, the two C-terminal residues, 163 and 164, were visible and were included in the final model. The observed conformation matches features observed in a difference map, using observed data, of wild-type *versus* a truncated wild-type in which residues 163 and 164 were deleted (P. Pjura & B.W.M., unpublished data).

Calculations

Atomic position and thermal factor shifts and solvent-accessible surface area (Lee & Richards, 1971), were calculated using EDPDB (Zhang & Matthews, 1995). For calculating rmsd in atom positions, the structures were first superimposed using the main-chain atoms of residues 81 to 161. This accounts for shifts of the entire molecule and/or hinge-bending motion. WT coordinates are Brookhaven Data Bank reference 3LZM; those of WT* are 1L63. Cavity volumes were calculated using the program MS (Connolly, 1983), with a probe radius of 1.2 Å. Atomic van der Waals radii (Eriksson *et al.*, 1992a) are available from the authors upon request.

Acknowledgments

We thank Larry Weaver, Mike Blaber and Dale Tronrud for helpful advice, and Joel Lindstrom, Joan Wozniak and Sheila Snow for excellent technical sup-

port. We thank Mike Blaber for data on the binding of benzene to crystals of L118A lysozyme. This work was supported in part by NIH grant GM21967 to B.W.M.

References

- Anderson, D. E., Lu, J., McIntosh, L. & Dahlquist, F. W. (1993). The folding, stability and dynamics of T4 lysozyme: a perspective using nuclear magnetic resonance. In *NMR of Proteins* (Clower, M. & Gronenborn, A., eds), pp. 258–304, Macmillan Press Ltd, Boca Raton, FL.
- Baldwin, E., Xu, J., Hajiseyedjavadi, O., Baase, W. A. & Matthews, B. W. (1996). Thermodynamic and structural compensation in "size-switch" core repacking variants of T4 lysozyme. *J. Mol. Biol.* **259**, 542–559.
- Barbas, C. F., Bain, J. D., Hoekstra, D. M. & Lerner, R. A. (1992). Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc. Natl Acad. Sci. USA*, **89**, 4457–4461.
- Barrick, D., Ho, N. T., Simplaceanu, V., Dahlquist, F. W. & Ho, C. (1997). A test for the role of the proximal histidines in the Perutz model of cooperativity in haemoglobin. *Nature Struct. Biol.* **4**, 78–83.
- Blaber, M., Baase, W. A., Gassner, N. & Matthews, B. W. (1995). Alanine scanning mutagenesis of the α -helix 115–123 of phage T4 lysozyme: effects on structure, stability and the binding of solvent. *J. Mol. Biol.* **246**, 317–330.
- Brandts, J. F. (1964). The thermodynamics of protein denaturation. II. A model of reversible denaturation and interpretations regarding the stability of chymotrypsinogen. *J. Am. Chem. Soc.* **86**, 4302–4314.
- Buckle, A. M., Cramer, P. & Fersht, A. R. (1996). Structural and energetic responses to cavity-creating mutations in hydrophobic cores: Observation of a buried water molecule and the hydrophilic nature

- of such hydrophobic cavities. *Biochemistry*, **35**, 4298–4305.
- Burley, S. K. & Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science*, **229**, 23–28.
- Carra, J. H., Murphy, E. C. & Privalov, P. L. (1996). Thermodynamic effects of mutations on the denaturation of T4 lysozyme. *Biophys. J.* **71**, 1994–2001.
- Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709–713.
- Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992a). The response of a protein structure to cavity-creating mutations and its relationship to the hydrophobic effect. *Science*, **255**, 178–183.
- Eriksson, A. E., Baase, W. A., Wozniak, J. A. & Matthews, B. W. (1992b). A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature*, **355**, 371–373.
- Eriksson, A. E., Baase, W. A. & Matthews, B. W. (1993). 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.
- Fauchère, J.-L. & Pliška, V. (1983). Hydrophobic parameters π of amino acid sidechains form the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **18**, 369–375.
- Feher, V. A., Baldwin, E. P. & Dahlquist, F. W. (1996). Access of ligands to cavities within the core of a protein is rapid. *Nature Struct. Biol.* **3**, 516–521.
- Fitzgerald, M. M., Churchill, M. J., McRee, D. E. & Goodin, D. B. (1994). Small molecule binding to an artificially created cavity at the active site of cytochrome *c* peroxidase. *Biochemistry*, **33**, 3807–3818.
- Gonzalez, L., Jr., Plecs, J. J. & Alber, T. (1996). An engineered allosteric switch in leucine-zipper oligomerization. *Nature Struct. Biol.* **3**, 510–515.
- Hamlin, R. (1985). Multiwire area X-ray diffractometers. *Methods Enzymol.* **114**, 416–452.
- Heinz, D. W., Baase, W. A., Dahlquist, F. W. & Matthews, B. W. (1993). How amino-acid insertions are allowed in an α -helix of T4 lysozyme. *Nature*, **361**, 561–564.
- Heinz, D. W., Baase, W. A., Zhang, X.-J., Blaber, M., Dahlquist, F. W. & Matthews, B. W. (1994). Accommodation of amino acid insertions in an α -helix of T4 lysozyme. Structural and thermodynamic analysis. *J. Mol. Biol.* **236**, 869–886.
- Kellis, J. T., Nyberg, K., Sali, D. & Fersht, A. R. (1988). Contribution of hydrophobic interactions to protein stability. *Nature*, **333**, 784–786.
- Knegtel, R. M., Kuntz, I. D. & Oshiro, C. M. (1997). Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **266**, 424–440.
- Ku, J. & Schultz, P. G. (1995). Alternate protein frameworks for molecular recognition. *Proc. Natl Acad. Sci. USA*, **92**, 6552–6556.
- Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367–382.
- Kuntz, I. D., Meng, E. C. & Shoichet, B. K. (1994). Structure-based molecular design. *Acc. Chem. Res.* **27**, 117–123.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
- Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
- Lu, J. & Dahlquist, F. W. (1992). Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry*, **31**, 4749–4756.
- Matsumura, M. & Matthews, B. W. (1989). Control of enzyme activity by an engineered disulfide bond. *Science*, **243**, 792–794.
- Matthews, B. W., Morton, A. G. & Dahlquist, F. W. (1995). Use of NMR to detect water within non-polar cavities. *Science*, **270**, 1847–1849.
- McRee, D. E., Jensen, G. M., Fitzgerald, M. M., Siegel, H. A. & Goodin, D. B. (1994). Construction of a bisquo heme enzyme and binding by exogenous ligands. *Proc. Natl Acad. Sci. USA*, **91**, 12847–12851.
- Morton, A. & Matthews, B. W. (1995). Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry*, **34**, 8576–8588.
- Morton, A., Baase, W. A. & Matthews, B. W. (1995). Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry*, **34**, 8564–8575.
- Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. (1989). Expression and ^{15}N labelling of proteins for proton and nitrogen-15 NMR. *Methods Enzymol.* **177**, 44–73.
- Nicholson, H., Anderson, D. E., Dao-pin, S. & Matthews, B. W. (1991). Analysis of the interaction between charged sidechains and the α -helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry*, **30**, 9816–9828.
- Perona, J., Hedstrom, L., Wagner, R. L., Rutter, W. J., Craik, C. S. & Fletterick, R. (1994). Exogenous acetate reconstitutes the enzymatic activity of trypsin Asp189Ser. *Biochemistry*, **33**, 3252–3259.
- Poteete, A. R., Dao-pin, S., Nicholson, H. & Matthews, B. W. (1991). Second-site revertants of an inactive T4 lysozyme mutant restore activity structuring the active site cleft. *Biochemistry*, **30**, 1425–1432.
- Schellman, J. A. (1976). The effect of binding on the melting temperature of biopolymers. *Biopolymers*, **15**, 999–1000.
- Schmid, M. F., Weaver, L. H., Holmes, M. A., Grütter, M. G., Ohlendorf, D. H., Reynolds, R. A., Remington, S. J. & Matthews, B. W. (1981). An oscillation data collection system for high-resolution protein crystallography. *Acta Crystallog. Sect. A*, **37**, 701–710.
- Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. (1995). A relationship between protein stability and protein function. *Proc. Natl Acad. Sci. USA*, **92**, 452–456.
- Shortle, D., Stites, W. E. & Meeker, A. K. (1990). Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry*, **29**, 8033–8041.
- Shrake, A. & Ross, P. D. (1992). Origins and consequences of ligand-induced multiphasic thermal protein denaturation. *Biopolymers*, **32**, 925–940.
- Tronrud, D. E. (1996). Knowledge-based B-factor restraints for the refinement of proteins. *J. Appl. Crystallog.* **29**, 100–104.
- Tronrud, D. E., Ten, Eyck L. F. & Matthews, B. W. (1987). An efficient general-purpose least-squares

- refinement program for macromolecular structures. *Acta Crystallog. Sect. A*, **43**, 489–503.
- Vetter, I. R., Baase, W. A., Heinz, D. W., Xiong, J.-P., Snow, S. & Matthews, B. W. (1996). Protein structural plasticity exemplified by insertion and deletion mutants in T4 lysozyme. *Protein Sci.* **5**, 2399–2415.
- Weaver, L. H. & Matthews, B. W. (1987). Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* **193**, 189–199.
- Weaver, L. H., Gray, T. M., Grütter, M. G., Anderson, D. E., Wozniak, J. A., Dahlquist, F. W. & Matthews, B. W. (1989). High-resolution structure of the temperature-sensitive mutant of phage lysozyme, Arg96 → His. *Biochemistry*, **28**, 3793–3797.
- Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**, 433–455.
- Wolfenden, R. & Radzicka, A. (1994). On the probability of finding a water molecule in a nonpolar cavity. *Science*, **265**, 936–937.
- Xu, J., Baase, W. A., Baldwin, E. & Matthews, B. W. (1998). The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci.* **7**, 158–177.
- Zhang, X.-J. & Matthews, B. W. (1995). EDPDB: A multi-functional tool for protein structure analysis. *J. Appl. Crystallog.* **28**, 624–630.

Edited by P. E. Wright

(Received 10 November 1997; received in revised form 24 December 1997; accepted 24 December 1997)