Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structures of the B1 domain of protein L from *Peptostreptococcus magnus* with a tyrosine to tryptophan substitution

Jason W. O'Neill,[a,b] David E. Kim,[c] David Baker[c] and Kam Y. J. Zhang[a,b]*

[a]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA, [b]Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195, USA, and [c]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

Correspondence e-mail: kzhang@fhcrc.org

Received 14 September 2000
Accepted 3 January 2001

PDB References: protein L B1 domain, with zinc, 1hz5; protein L B1 domain, without zinc, 1hz6.

The three-dimensional structure of a tryptophan-containing variant of the IgG-binding B1 domain of protein L has been solved in two crystal forms to 1.7 and 1.8 Å resolution. In one of the crystal forms, the entire N-terminal histidine-tag region was immobilized through the coordination of zinc ions and its structural conformation along with the zinc coordination scheme were determined. However, the ordering of the histidine tag by zinc does not affect the overall structure of the rest of the protein. Structural comparisons of the tryptophan-containing variant with an NMR-derived wild-type structure, which contains a tyrosine at position 47, reveals a common fold, although the overall backbone root-mean-square difference is 1.5 Å. The Y47W substitution only caused local rearrangement of several side chains, the most prominent of which is the rotation of the Tyr34 side chain, resulting in a 6 Å displacement of its hydroxyl group. A small methyl-sized cavity bounded by β-strands 1, 2 and 4 and the α-helix was found in the structures of the Y47W-substituted protein L B1 domain. This cavity may be created as the result of subsequent side-chain rearrangements caused by the Y47W substitution. These high-resolution structures of the tryptophan-containing variant provide a reference frame for the analysis of thermodynamic and kinetic data derived from a series of mutational studies of the protein L B1 domain.

1. Introduction

Small proteins are useful experimental systems that simplify the study of the biophysical consequences of amino-acid substitutions. The 64-residue immunoglobulin G (IgG) binding B1 domain of protein L in *Peptostreptococcus magnus* (Kastern *et al.*, 1990) is a good model system for studying the thermodynamics and kinetics of protein folding. The structure of the wild-type protein L B1 domain (referred to as WT) was first determined by NMR (Wikström *et al.*, 1994), which revealed a $\beta\beta\alpha\beta\beta$ fold. WT lacks disulfide bridges, metal cofactors and prolines, which can complicate folding kinetics.

The thermodynamics and kinetics of folding and unfolding of the B1 domain have been extensively studied by various methods such as changes in fluorescence (Gu *et al.*, 1995, 1997; Scalley *et al.*, 1997; Yi & Baker, 1996). To follow these changes, a tyrosine at position 47 was substituted with a tryptophan (Y47W), which has a 5.5-fold greater fluorescence signal than tyrosine (Cantor *et al.*, 1980). In this paper, the B1 domain containing the Y47W substitution will be referred to as the pseudo-wild-type (WT*) and all other mutant B1 domains discussed here contain the Y47W substitution. The Y47W mutation is located in the protein core and has little effect on

stability ($\Delta G = -19.2 \pm 0.8$ kJ mol^{-1} for WT* and $\Delta G = -21.6 \pm 0.8$ kJ mol^{-1} for WT; Scalley *et al.*, 1997).

A series of 70 point mutations distributed throughout WT* were analyzed for their thermodynamics and kinetics of folding and unfolding (Kim *et al.*, 2000). It was found that in the folding transition state β-strands 1 and 2 and the first β-hairpin turn are largely structured, while β-strands 3 and 4, the second β-hairpin turn and the α-helix are largely disrupted (Kim *et al.*, 2000). Interpretation of the mutational effects in that paper was based upon the NMR-derived WT structure, which contains a tyrosine at position 47.

The accurate interpretation of the thermodynamics and kinetics of the mutants derived from the tryptophan-containing variant requires the atomic resolution structure of this pseudo-wild-type protein L B1 domain. We have therefore solved orthorhombic (form 1) and trigonal (form 2) forms of WT* to 1.7 and 1.8 Å resolution, respectively. The form 2 crystal utilizes zinc ions to facilitate crystal formation through coordination of the N-terminal histidine tag (His tag). Comparison of the NMR-derived WT structures with the two crystal-derived WT* structures revealed significant root-mean-square differences (r.m.s.d.s). A methyl-sized cavity, which may be the result of the Y47W substitution, was found in the WT* structures. Interestingly, the cavity is not located at W47 but adjacent to those residues that readjusted to compensate for the mutation. The cavity in WT* provides a route to explain why several mutations have greater than expected destabilizing effects.

2. Materials and methods

2.1. Purification, crystallization and data collection

The initial cloning, purification and crystallization of the Y47W-substituted B1 domain of protein L (WT*) in the orthorhombic form (form 1) has been described previously (Johnsen *et al.*, 2000). The sequence numbering is consistent with Kim *et al.* (2000), where residues 2–64 in this paper are equivalent to 92–155 in Wikström *et al.* (1993, 1994). The WT* sequence used in our study lacks the first 16 amino acids that were disordered in the NMR structure (Wikström *et al.*, 1994). The protein contains a leader His tag (numbered −7 MHHHHHHA 0), making the total length of the protein 72 residues (Fig. 1). Purified WT* was crystallized in 30% PEG 8000 and 0.2 *M* ammonium sulfate at 277 K (form 1). The crystallization buffer with additional 20% PEG 400 was used as cryoprotectant. The form 2 crystals were grown in 150 m*M* ZnOAc, 1% PEG 8000 and 50 m*M* MES pH 6.0 at 277 K. The crystallization buffer plus 25% glycerol and 5% PEG 400 was used as cryoprotectant. All diffraction data were collected on an R-AXIS IV image plate at 105 K using Cu $K\alpha$ radiation ($\lambda = 1.5418$ Å) generated by a Rigaku rotating-anode generator operating at 50 kW. The data-collection statistics for both form 1 and form 2 crystals are summarized in Table 1.

2.2. Structure determination, model refinement and validation

The structures of form 1 and 2 crystals were solved by molecular replacement using the program *EPMR* (Kissinger *et al.*, 1999). The initial search model for the form 1 crystal was the 71% identical protein L C3 domain structure, whose 2.3 Å coordinates were kindly provided by Dr Tommy Wan and Dr Brian Sutton (private communication). Non-identical side chains were truncated to alanine (18 of 64 residues). Reflections between 3.5 and 15 Å were used for molecular replacement. Three WT* molecules related by improper non-crystallographic symmetry were found in the asymmetric unit. Attempts to use the NMR-derived WT structures as molecular-replacement models failed.

The automatic model-building features of *wARP* (version 5.0; Perrakis *et al.*, 1997) were utilized to build the initial form 1 WT* molecule. *wARP* creates an initial map based on the molecular-replacement solution and uses this map to build its own model. After ten building cycles and 100 refinement cycles, *wARP* autobuilt three contiguous chains totaling 184 residues and 685 free waters, many of which belonged to atoms of side chains and ordered waters. The final *wARP* refinement statistics were $R = 19.2\%$ and $R_{\rm free} = 27.7\%$. The final *wARP*-generated $2F_o - F_c$ map is shown in Fig. 2(*a*). This map has clear density for most side chains and waters. The programs *O* (Jones *et al.*, 1991) and *XFIT* (McRee, 1992) were used for manipulation of the molecular-replacement solutions and model building.

A free *R* set (Brünger, 1992) containing 5% of the data was generated using the *CCP*4 program *FREERFLAG*. The program package *CNS* (Brunger *et al.*, 1998) was used for structural refinement, which included iterative cycles of positional, simulated-annealing and individual *B*-factor refinement as well as automatic water picking. A simulated-annealing composite omit $2F_o - F_c$ map was generated after each cycle of refinement to verify atomic positions and to aid model rebuilding. The final structure from the form 1 crystal contains residues H−1 to G64 for molecule

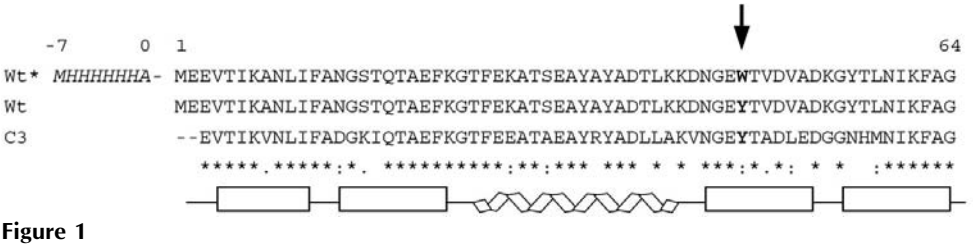

Figure 1
A sequence alignment of the protein L B1 domain containing the Y47W substitution (WT*), the B1 domain containing Y47 (WT) and the C3 domain (C3). The secondary-structure elements of the WT* molecule are shown below the alignment. Identical residues between WT* and the C3 domain are denoted by '*'. Conservative amino-acids differences are symbolized as ':' or '.'. The N-terminal 6×His tag found only on the WT* protein is in italics and is numbered M−7 to A0. The sequence numbering for WT* is consistent with Kim *et al.* (2000).

Table 1
Structural refinement statistics of the pseudo-wild-type (Y47W) protein L B1 domain crystals.

	Form 1	Form 2
Space group	$P2_12_12_1$	$P3_221$
Unit-cell parameters (Å)	a = 51.62, b = 54.01, c = 95.00	$a = b$ = 66.44, c = 109.18
Resolution (Å)	1.70	1.80
Completeness (%)	96.3	99.8
Redundancy	5×	4×
$I/\sigma(I)$†	17 (7)	19 (4)
R_{merge} (%)	7.5	6.6
Unique reflections	28825	26449
Reflections in cross-validation set	1471	1341
R (R_{free}‡) (%)	19.3 (21.8)	18.8 (19.3)
Luzzati coordinate error (Å)	0.19	0.18
Solvent content (%)	53	66
Visible residues in each monomer	66, 62, 62	72, 72
No. of water molecules	302	237
Number of Zn^{2+} ions	0	10
R.m.s.d. bonds (Å)	0.0056	0.0060
R.m.s.d. angles (°)	1.22	1.30
Average B factors, protein atoms (Å^2)	21.3	18.3
Average B factors, waters (Å^2)	32.9	32.0
Average B factors, Zn^{2+} (Å^2)	—	36.7
Ramachandran statistics for each monomer§		
Residues in most favored regions	59, 56, 55	64, 65
Residues in additional allowed regions	1, 1, 2	3, 2
Glycine residues	5, 5, 5	5, 5

† The value in parentheses is $I/\sigma(I)$ in the highest resolution shell. ‡ Crystallographic R_{free} values (Brünger, 1992) were calculated with 5% of the total structure factors and were not included in the structure refinement. § The structures of each monomer in the asymmetric unit were assessed using *PROCHECK* (Laskowski *et al.*, 1993).

Table 2
The main-chain r.m.s.d. between the form 1 and 2 monomers of WT*, the NMR-derived WT structures and the C3 domain.

The first and second values represent the r.m.s.d. for all the residues (4–62) and the core residues (4–12, 16–23, 26–40, 45–53 and 56–62), respectively. The three molecules in the asymmetric unit of the form 1 crystals are labeled WT* *A*, WT* *B* and WT* *C*. The two non-crystallographic symmetry-related molecules in the form 2 crystals are labeled WT*Z *A* and WT*Z *B*. Non-crystallographic symmetry was not used for the structure refinement of WT* and WT*Z. The larger differences between the WT* monomers in the asymmetric unit occurred in turn regions. Other minor shifts in the backbone occurred around surface-exposed residues. Core residues had the smallest overall differences, with the exception of a partially buried V51. Interestingly, there were smaller r.m.s.d. differences between the C3 domain and WT* than there were between WT and WT*.

R.m.s.d. (Å)	WT* *B*	WT* *C*	WT*Z *A*	WT*Z *B*	NMR	C3†
WT* *A*	0.55/0.50	0.69/0.67	0.38/0.33	0.49/0.42	1.51/1.20	1.10/0.98
WT* *B*	—	0.56/0.53	0.53/0.52	0.48/0.48	1.46/1.22	1.07/0.90
WT* *C*	—	—	0.69/0.66	0.68/0.64	1.50/1.27	1.22/1.07
WT*Z *A*	—	—	—	0.29/0.28	1.47/1.21	1.05/0.94
WT*Z *B*	—	—	—	—	1.49/1.22	1.05/0.91
NMR‡	—	—	—	—	*1.25/0.80*§	1.75/1.45

† Model *A* of the two C3 monomers was used for comparison (r.m.s.d. between two monomers in the asymmetric unit is 0.064 Å owing to the use of NCS restraints). ‡ Solution #1 from the NMR structures of WT was used to make the r.m.s.d. valuations. § Mean r.m.s.d. between NMR solutions 1, 3, 6, 9, 13 and 17. The standard deviation of measurement is 0.11 Å.

A, residues E2–G64 for molecule *B* and residues E2–G64 for molecule *C*, as well as 302 waters. The final R factor for the form 1 crystal is 19.3% and the final R_{free} is 21.8% (Table 1). The form 1 WT* structure was also independently determined from the initial molecular-replacement solution using a combination of model building by *O* into simulated-annealing composite omit maps and model refinement by *CNS* without the aid of *wARP*. The same final structure was obtained.

The refined form 1 structure was used as a molecular-replacement search model to determine the form 2 structure by *EPMR* with data between 3.5 and 15 Å resolution. Refinement was carried out using *CNS* with a maximum-likelihood target function and torsion-angle dynamics. The His tags were ordered in the form 2 crystals owing to coordination with Zn^{2+} ions from the precipitant, thereby allowing all 72 residues to be built for each monomer in the asymmetric unit (Fig. 4*a*). Simulated-annealing composite omit maps were utilized to ensure unbiased model building as well as accurate identification and placement of zinc ions (Fig. 2*b*). The final R factor was 18.8% and R_{free} was 19.3% (Table 1). Non-crystallographic symmetry restraints were not used for the refinement of either form 1 or form 2 structures. Consequently, the comparison of molecules in the asymmetric unit could serve as an internal cross-validation of their structures as well as a means of revealing regions of conformational flexibility in the molecule.

The stereochemical properties of WT* structures in form 1 and form 2 crystals were examined by *PROCHECK* (Laskowski *et al.*, 1993). The quality of the main-chain and side-chain parameters were judged as being mostly 'better' and sometimes 'inside' the normal range of comparable structures at the same resolution. The Ramachandran plot showed that the monomers in both crystal forms had at least 98% of the residues in the most favored region (Table 1). Two structure-validation methods that utilize structural properties not imposed during the refinement were also used to assess the structure quality. The environmental preference for each residue was assessed by the program *VERIFY-3D* (Lüthy *et al.*, 1992). The distribution of non-bonded atoms in the neighborhood of each atom was analyzed by *ERRAT* (Colovos & Yeates, 1993). No violations were detected by either method.

2.3. Cavity and volume calculations

The program *VOIDOO* (Kleywegt & Jones, 1994) was used to calculate cavity volumes for WT* and WT. At least four cavity calculations were carried out for each molecule with randomly generated orientations in order to minimize measurement artifacts. A rolling probe with the radius of 1.2 Å was used for the cavity-volume calculation; for reference, *VOIDOO* would use a 1.4 Å radius to mimic a water molecule. The use of a 1.2 Å radius attempts to standardize our calculated volume results with those reported in a comprehensive analysis of cavity formation in T4 lysozyme structures (Xu *et al.*, 1998).

3. Results and discussion

3.1. The protein L B1 WT* structures

The WT* structure has a very similar fold to the WT structure as determined by Wikström *et al.* (1994) and consists of a central α-helix packed against a four-stranded β-sheet (Fig. 3). The topology of β-strands in WT* are mixed, where the central β-strands (1 and 4) are parallel and β-strands 1 and 2 as well as 3 and 4 are antiparallel. The β-sheet also has a typical right-handed twist. The loop and turn regions are well defined in the WT* structures and showed little variation among the monomers in the two crystal forms.

Each of the three molecules in the asymmetric unit of the form 1 crystals is in a heterogenic environment, making different crystal contacts with neighboring molecules and slightly affecting both backbone and side-chain conformations. The resulting main-chain and side-chain variations are reflected in main-chain r.m.s.d.s of 0.56 Å between *A* and *B*, 0.48 Å between *B* and *C*, and 0.69 Å between *A* and *C* for residues 4–62 (Table 2). The two monomers in the asymmetric unit of the form 2 crystals are in very similar environments, unlike those in the form 1 crystals. In fact, the main-chain r.m.s.d. between the form 2 monomers is only 0.29 Å. The *A* molecules are the most similar in forms 1 and 2, with an r.m.s.d. of 0.38 Å. The most significant side-chain variations occur at surfaces and crystal contact sites.

For the form 1 crystal there were 1484 protein atoms identified with an average *B*-factor value of 22.2 Å^2, with the *C* molecule having the lowest overall *B* factor, 20.2 Å^2. These *B* values correlate with the amount of surface area buried for each molecule: molecule *A* has 726 Å^2 buried, molecule *B* has 1146 Å^2 buried and molecule *C* has 1230 Å^2 buried. The highest *B*-factor values correspond to the C-termini of the α-helix in all three molecules. Overall, there were 302 waters identified in the form 1 structure; 23 of these solvent molecules have *B* factors of 20 Å^2 or below and are important in making crystal contacts by forming at least two or three hydrogen bonds. There were no buried waters observed. The form 2 crystal had an average *B* factor of 20.8 Å^2, including 1131 protein atoms, ten zinc ions and 214 waters (Table 1).

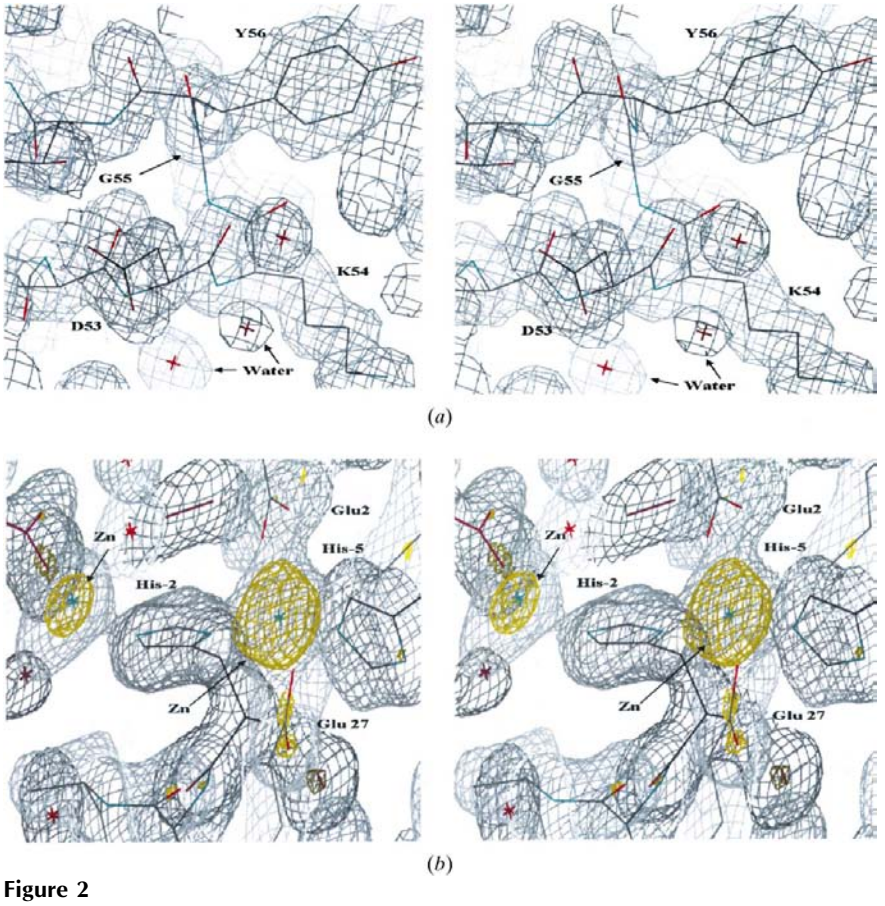

Figure 2
(*a*) A *wARP*-generated map in the WT* second β-turn region. Based on an initial molecular-replacement model of the C3 domain and diffraction data from the form 1 crystal, a stereoview of a *wARP*-generated $2F_o - F_c$ map shows clear electron density for side chains and waters. The majority of waters picked from this map have the spherical character exhibited here. The wire-frame model is taken from the refined form 1 molecule *A* structure, showing the second β-turn formed by residues D53, K54 and G55, which have three consecutive positive φ angles. The map is contoured at 1σ. (*b*) A simulated-annealing composite omit map around a zinc-coordinated His-tag region. From the form 2 data, the electron density for Zn^{2+} ions coordinated to the His tag show up clearly in a simulated-annealing composite omit $2F_o - F_c$ map contoured at 1σ (gray) and 5σ (yellow). Zn^{2+} ions are represented as dark green crosses, N atoms are in cyan and O atoms and waters are in red. Shown here is H−2 coordinating with two Zn^{2+} ions (Zn3 and Zn8). Symmetry-related residues are shown in purple. Maps were imaged using the *XtalView* program (McRee, 1992).

3.2. Zinc coordination of the histidine tag in the form 2 structure

The His tags (−7 MHHHHHHA 0) of the three monomers in form 1 crystals are mostly disordered; only H−1 is visible in molecule *A*. However, the His tags were ordered through coordination with zinc ions in the form 2 crystals (Fig. 4*a*). Furthermore, the zinc is acting as a crystal contact mediator that orders symmetry-related His-tag and non-His-tag residues (Fig. 4*b*). Zinc ions were identified in the form 2 crystals based on the height of the electron-density peaks and coordination geometry with neighboring atoms. Moreover, zinc is the only divalent cation in the crystallization buffer. The presence of zinc was also confirmed by the fluorescence wavelength scan around the zinc anomalous absorption edge (data not shown). The primary coordination to zinc is T4 tetrahedral, although T6 octahedral also occurs (Alberts *et al.*, 1998). The average distance between zinc and its coordinating atoms in the form 2 WT* structure is 2.2 Å for proteins and 2.9 Å for waters, which is in agreement with Alberts *et al.* (1998). A dual zinc coordination occurs with H2, where both amides of the histidine imidi-

zole ring coordinate separate zinc ions (Fig. 4*b*). This is the first example reported in the PDB of a completely ordered 6×His tag coordinated to zinc being visualized by X-ray crystallography.

3.3. The second β-hairpin turn

The second β-hairpin turn has a distorted type I′ configuration. It is composed of four amino acids, three of which, D53, K54 and G55, have positive φ angles in all five monomers across the two crystal forms (*e.g.* φ = 50.2, 60.2, 98.1° and ψ = 47.1, 26.3, −4.5° for residues 53, 54 and 55, respectively, in molecule *A* of the form 1 crystal). A simulated-annealing composite omit map calculated using *CNS* showed that the electron density in this β-hairpin turn region is very clear, and the assignments of both the main-chain and side-chain conformations are unambiguous. The main chain of this turn is nearly perpendicular to the plane of the β-sheet formed by strands 3 and 4. It is stabilized in this unusual position by Y56 packing against F26, which is located at the N-terminus of the α-helix, as well as by a hydrogen bond between the carbonyl O atom of Y56 and the amide N atom of K7 in the first β-strand. This highly unusual turn places residues D53, K54 and G55 in a region of φ/ψ space that is usually restricted for a left-handed α-helix on a Ramachandran plot. Since this is a less favorable region energetically, it is thought that this region of the protein is strained and therefore destabilizing. In support of this, a K54G mutation increased the overall stability ($\Delta\Delta G$ = +2 kJ mol^{-1}; unpublished data). This increased stability may come from decreased strain in the second β-turn.

In contrast, increasing the overall strain in this turn by replacing the preferred glycine at position 55 with an alanine decreased stability by −8.8 kJ mol^{-1} and dramatically increased the unfolding rate 16-fold over WT* (Kim *et al.*, 2000). In the WT* structure, G55 has a positive φ angle (98.1°) placing it in the 'generously allowed region' of the Ramachandran plot. By modeling a G55A mutation onto the WT* structure and performing energy minimization with *CNS*, steric clashes were predicted owing to the C^{β} methyl group of A55 being only 2.5 Å away from the carbonyl O atom of A55. The increased strain in the second β-hairpin turn caused by steric hindrance could explain the significant loss in stability for the G55A mutant. Generally, non-bonded C^{β} atoms are at

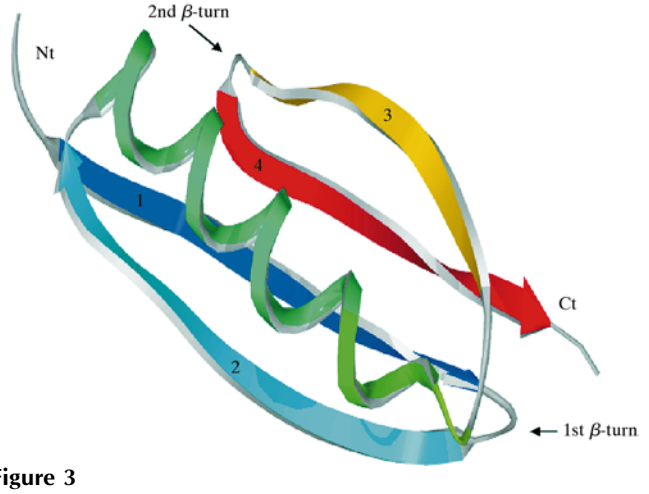

Figure 3
A color-coded ribbon diagram representing the WT* (Y47W) protein-L B1 domain. The β-strand 1 (residues 4–12) is in blue, β-strand 2 (residues 15–24) is in cyan, α-helix (residues 26–43) is in green, β-strand 3 (residues 46–52) is in orange and β-strand 4 (residues 57–63) is in red. The first β-turn (residues 13–14) and second β-turn (residues 53–56) are marked by arrows and labeled. Ribbon diagrams were created using the program *SwissPDBViewer* (Guex & Peitsch, 1997) and rendered using the program *POV-Ray* (http://www.povray.org).

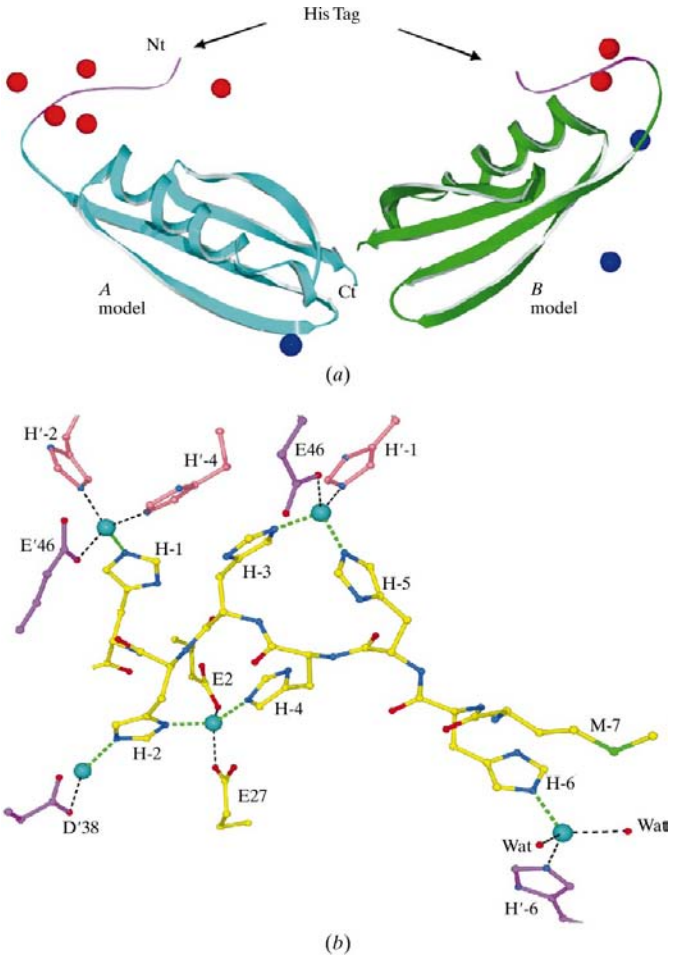

Figure 4
The zinc coordination scheme and the conformation of the N-terminal 6×His tag in molecule *A* of the form 2 crystal. (*a*) The asymmetric unit of the form 2 crystal contains two monomers related to each other approximately by a 120° rotation. The zinc ions are shown as either red or blue spheres, representing zinc ions coordinated either to the N-terminal His tag or to other side chains in the structure. (*b*) The yellow residues represent molecule *A*; purple and peach residues are symmetry mates and are numbered in italics. Label examples are H−2, His residue '−2'; D′38, Asp38 from molecule *B* in the neighboring asymmetric unit. The Zn^{2+} ions are represented as cyan balls. All Zn^{2+} ions coordinated to the His tag have greater than 5σ electron-density peaks in $2F_o - F_c$ maps. H2 coordinates with two Zn^{2+} ions (Zn3 and Zn8). The interatomic zinc to side-chain distance ranges from 2.00 to 2.36 Å. Coordination is primarily tetrahedral or octahedral. Bonds between the zinc and histidines in the His tag are represented by green dotted lines. Bonds between other amino acids and zincs are shown as black dotted lines. Blue spheres represent nitrogen and red spheres are oxygen unless denoted as water (Wat). The zinc between D′38 and H−2 is likely to coordinate two more waters, though the electron density for the placement of these waters was ambiguous.

least 2.8 Å from carbonyl O atoms, although 90% of the time the distance is 3.4 Å or greater (Singh & Thornton, 1990). Furthermore, the strain observed in the G55A model suggests that the increased unfolding rate of the G55A protein results from the strain acting like a spring, which is relieved upon the opening of the second β-hairpin turn during unfolding (Kim *et al.*, 2000).

3.4. Effect of Y47W substitution in WT* and comparisons with the NMR structure

Protein L has five homologous B domains and three C domains that share 68–90% residue identity. Using equivalent numbering, Fig. 1 shows a sequence alignment for WT*, WT and the 71% identical C3 domain used for the molecular-replacement solution. The main-chain r.m.s.d.s for either all residues or core residues within the regular secondary-structural elements between each of the independent WT* molecules, WT and C3 domains are given in Table 2. The three monomers in the asymmetric unit are very similar to each other, with an all-residue main-chain r.m.s.d. of 0.55–0.69 Å. However, a representative sampling of the 98% identical NMR WT-molecule structures (solutions 1, 3, 6, 9, 13 and 17) showed a main-chain r.m.s.d. of 1.5 Å when compared with WT* (Fig. 5*a*). This is greater than the difference between the 71% homologous C3 crystal structure and WT*, which has a main-chain r.m.s.d. of 1.1 Å (Table 3). As seen in Fig. 5(*a*), the largest differences occur in loop and turn regions, especially at the first β-hairpin turn and the second helix turn. When only the core residues are compared, the main-chain r.m.s.d. between the WT* and WT structures is 1.27 Å, which is still significantly larger than the main-chain r.m.s.d. of 0.98 between the WT* and C3 structures. The large differences between the WT* and WT structures reflect a greater overall coordinate uncertainty in the WT NMR structures rather than the effect of the Y47W mutation. The Ramachandran plots of a representative sample of NMR-solution structures showed a mean percentage of residues in the most favored region of only 68 ± 5%. Furthermore, the average r.m.s.d. of main-chain atoms between the representative NMR solutions is 1.25 Å for all residues and 0.80 Å for the core residues. This large internal variation in the WT NMR structures makes them less reliable for the interpretation of the small effects that amino-acid substitution may have upon interactions such as hydrogen bonding or hydrophobic interactions.

The Y47W substitution caused a significant number of side-chain rearrangements when comparing the WT* and WT structures. The most significant side-chain rearrangement was found with Y34. Owing to the substitution of tyrosine by a bulkier tryptophan, the side chain of Y34 in WT* rotated around χ_1 by about 72° and was pushed away by about 5.3 ± 0.6 Å judging from its hydroxyl group (Fig. 5*b*). Also, Y36 moved by about 2.7 ± 0.6 Å with respect to its hydroxyl positions in the WT* and WT structures. Both the Y47 of the WT structure and W47 of the WT* structure have similar χ_1 values.

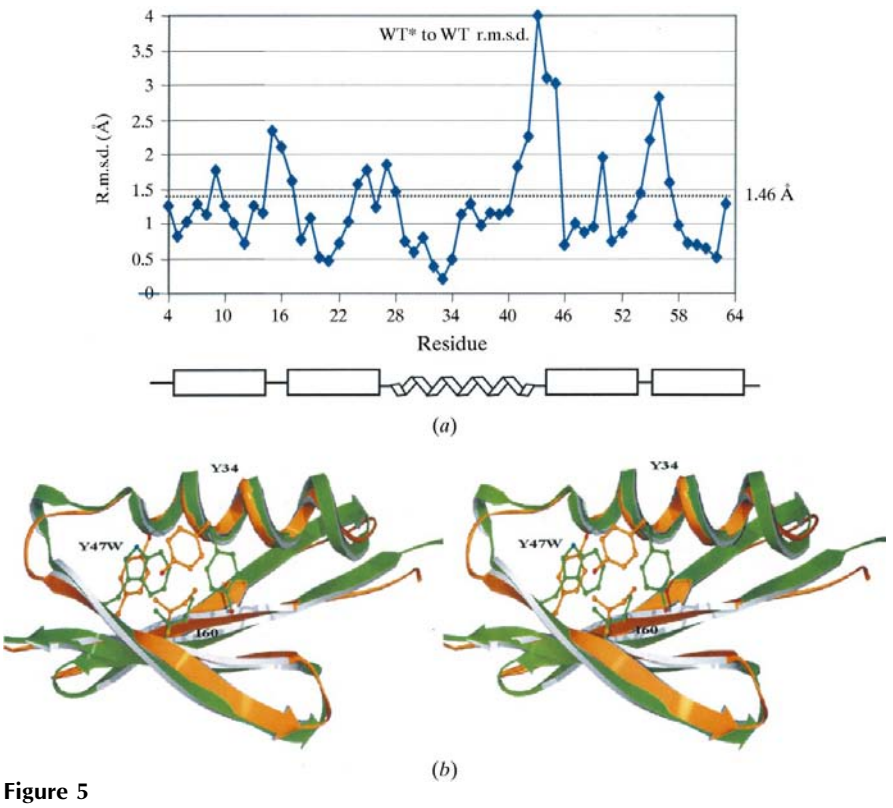

Figure 5
(*a*) Main-chain r.m.s.d. between molecule *A* in WT* form 1 and the WT NMR solution #1 model. The average r.m.s.d. is 1.46 Å, which is denoted by a dotted line through the chart and labeled to the right side of the chart. The r.m.s.d. was calculated using the *CCP*4 program *LSQKAB* (Kabsch, 1976). (*b*) Stereoview of ribbon model overlays of the WT* form 1 molecule *A* (green) and the WT NMR-derived solution #1 model (orange). The proteins exhibit the same overall fold, though the main-chain r.m.s.d. is 1.46 Å. Major differences occur in the loop regions as well as the N-terminus of β-strand 4. The side chains for residues Y47/W47, Y34 and I60 are shown. The substitution of Y47 in WT by W47 in WT* caused Y34 in WT* to swing 6 Å away relative to the Y34 position in WT as measured by the side-chain hydroxyl group.

3.5. Interior hydrophobic cavity

The repositioning of main and side chains arising from the Y47W substitution and primarily the movement of Y36 may be responsible for our observation of a small hydrophobic cavity in WT*. This cavity in WT* was detected with the program *VOIDOO* (Kleywegt & Jones, 1994). It is surrounded by the following residues: A8, L10, A20, F22, A33, Y36, A37 and I60 (Table 3*a*). A similar cavity of about 20 Å^3 was detected in two of the five random molecular orientations from only one (solution 6) of the six randomly chosen representative WT structures (solutions 1, 3, 6, 9, 13 and 17). It is more likely that there is no cavity in the NMR-derived WT structures based on our

Table 3
Cavity-surrounding residues.

(*a*) Residues and atoms surrounding the cavity. Listed in this table are the eight residues and the 23 atoms lining the cavity.

Residue	Contributing atoms
Ala8	CB
Leu10	CG, CD1, CD2
Ala20	CA, CB
Phe22	CD1, CE1, CE2, CZ
Ala33	C, O, CA, CB
Tyr36	CA, CB, CG, CD2, CE2
Ala37	N, CA, CB
Ile60	CD1

(*b*) Cavity sizes for each molecule in the form 1 and 2 asymmetric unit. The program *VOIDOO* (see §2) was used to calculated volume for the given molecule as well as for the NMR solution-structure models. Each round of cavity-volume estimation was carried out using a randomly rotating molecule. Each of WT* monomers in form 1 and 2 had different cavity sizes owing to slight rotations of side-chain and main-chain positions.

Molecule	Cavity (Å^3)
WT**A*	20.3 ± 0.9
WT* *B*	30.8 ± 0.4
WT* *C*	17.1 ± 0.5
WT*Z *A*	27.6 ± 1.8
WT*Z *B*	33.0 ± 0.5

(*c*) Thermodynamic effects of mutations on residues surrounding the cavity. The cavity-surrounding residues have greater loss of stability upon mutation than those same residue types located at other locations in the structure, including those involved with core interactions. The data on stabilities is taken from Kim *et al.* (2000).

Residue	Mutation	$\Delta\Delta G$ (kJ mol^{-1})
Ala8	Gly	−10.9
Leu10	Ala	−11.7
Ala20	Val, Gly	+5.4, −8.8
Phe22	Ala, Leu	−23.9, −14.2
Ala33	Gly	−15.9
Tyr36	Ala	−10.0
Ala37	Gly	−14.2
Ile60	Ala, Val	−19.2, −6.7

(*d*) Frequencies of recovered cavity-surrounding mutations with larger side chains than those in WT*. A series of random sequences were inserted into the first β-strand, α-helix or fourth β-strand. Listed here are the number of recovered residues that had a larger side chain over the total number of recovered mutants from those cavity-lining residues (data on frequency of replaced residues was taken from Kim *et al.*, 2000). The majority of the tested replacements with larger side-chain substitutions also had increased folding rates. This data suggests that larger side chains would satisfy more van der Waals contacts by filling the cavity.

Residue	Replaced residue frequency
Ala8	Val or Thr, 14/23
Leu10	Tyr or Phe, 16/24
Ala33	Leu, Val or Ile, 21/23
Ile60	Phe or Tyr, 20/25

sampling statistics. However, it is possible that this cavity was present in the WT structure and could not be detected unequivocally owing to the large coordinate uncertainty in the NMR-derived WT structures. To clarify this issue a high-resolution crystal structure of the WT molecule would need to be determined.

The slight variations in the placement of side chains surrounding the cavity are represented in Fig. 6, where three sets of residues corresponding to the three molecules in the form 1 asymmetric unit are superimposed. Also, *VOIDOO* calculated cavities of varying sizes in the five monomers in the asymmetric units of form 1 and 2 crystals. In form 1, the largest cavity was found in the *B* molecule at 31 Å^3, while the *A* molecule had a 20 Å^3 cavity and the *C* molecule had the smallest cavity at 17 Å^3. In form 2, the cavity sizes were 27 and 33 Å^3 (Table 3*b*) in molecules *A* and *B*, respectively. These values correlate with the amount of buried surface area found at the interfaces of the molecules in the asymmetric units, suggesting the slight changes in side-chain orientation and cavity size are a consequence of some dynamic expansion and collapse in the structure.

The discovery of the cavity in WT* also helps to explain why the cavity-surrounding mutations listed in Table 3(*c*) have greater $\Delta\Delta G$ values (Kim *et al.*, 2000) than other mutations in the same class (*e.g.* Ala→Gly) that do not surround cavities: it is likely that they increase the size of the cavity. The exception is Y34A ($\Delta\Delta G$ = −12.5 kJ mol^{-1}), which may be more destabilizing than Y36A since it makes important hydrophobic contacts to W47. The cavity in WT* may also explain the relatively small effect of the Y47W mutation on the thermostability. One would expect that the substitution of a tyrosine by a tryptophan in the hydrophobic core should significantly increase the thermostability. However, the increased hydrophobic contacts with W47 could be cancelled out owing to the creation of this cavity.

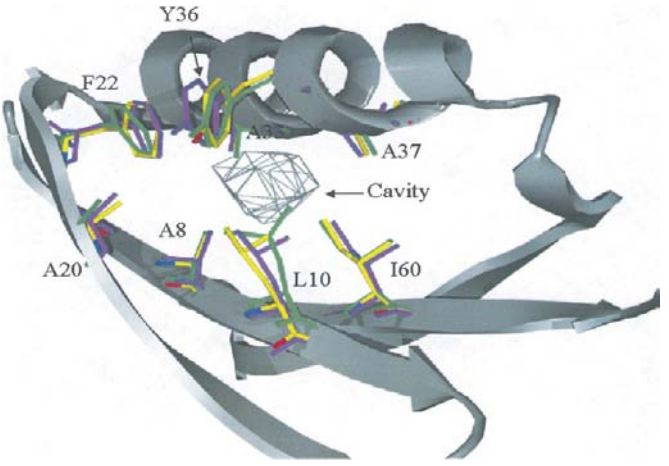

Figure 6
An overlay of the cavity and the cavity-surrounding side chains. The hydrophobic cavity, shown in gray gridwork, is located between β-strands 1, 2, 4 and the α-helix and is centered at (7.25, 8.50, 6.25 Å) in (*x*, *y*, *z*) coordinates in the form 1 *B* molecule. Side chains from the form 1 molecule *A* (magenta), *B* (yellow) and *C* (green) are superimposed, showing the dynamic range of side-chain organization. The residues that surround this cavity include: A8, L10, A20, F22, A33, Y36, A37, I60. A rolling probe of 1.2 Å was used to calculate volumes using the program *VOIDOO*. The cavity sizes were calculated for at least four randomly rotated molecules and then averaged. Shown here is the *B* molecule with cavity volume of 31 Å^3. The other molecules have cavity sizes that range from 17 to 32 Å^3. Parameters used in *VOIDOO* were: grid = 0.25, shrink factor = 0.90, plot grid = 0.2, growth factor = 1.1, minimum size of secondary grid = 10, convergence criterion (% and Å) = 0.1.

The cavities in model proteins such as T4 lysozyme (Eriksson, Baase, Wozniak *et al.*, 1992; Eriksson, Baase, Zhang *et al.*, 1992; Xu *et al.*, 1998), barnase (Buckle *et al.*, 1993, 1996) and chymotrypsin inhibitor 2 (CI2; Jackson *et al.*, 1993) have primarily resulted from large-to-small side-chain substitutions. In barnase, cavities of the size found in WT* can be created through Ile→Val mutations (Buckle *et al.*, 1993). In T4 lysozyme, a series of non-polar side chains were substituted with alanine, creating cavities that ranged in size from 10 to 89 Å^3. The consequence of these cavity-creating mutations is a reduction in stability owing to loss of favorable van der Waals contacts and hydrophobic interactions. Xu *et al.* (1998) estimate the energetic penalty paid for creating a cavity in T4 lysozyme to be 92 J mol^{-1} Å^{-3}. Based on this value, we could expect that the cavity volume (25 Å^3 on average) found in WT* destabilizes the protein by approximately -2.3 kJ mol^{-1} (92 J mol^{-1} Å^{-3} $\times$ 25 Å^3). Thus, if the cavity could be filled without inducing any additional strain, we could predict that the $\Delta G \simeq -21.6$ kJ mol^{-1} (WT* has $\Delta G = -19.2$ kJ mol^{-1}).

The C^{β} of A20 contributes to the wall of the cavity. Experimental studies reveal that the A20V mutation results in a 5.4 kJ mol^{-1} increases in stability (Kim *et al.*, 2000). This suggests that A20V stabilizes the protein to a greater extent than just filling the cavity. In another study, several segments of WT* (including β-strand 1, β-turn 1, α-helix, β-turn 2, β-strand 4) were mutated simultaneously and then selected for IgG binding using phage-display methods (Kim *et al.*, 2000). It is interesting to note that for the cavity-surrounding residues there is a preference for larger non-polar side chains (Table 3*d*). Even though stabilities tend to be lower, these mutations increased the overall folding rates. Taken together, these results suggest that the bulkier non-polar side chains satisfy more van der Waals contacts in or around this cavity. Indeed, energy-minimized modeling of L10F (from Table 3*d*) onto WT* suggests that the addition of three C atoms would fill this cavity, with only minor steric clashes. This approach has been attempted with T4 lysozyme, where two naturally occurring cavities of 39 and 23 Å^3 exist (Karpusas *et al.*, 1989). Here, an L133F or an A129V substitution was introduced to fill the larger or smaller cavities, but the resulting steric clashes decreased the stability by 1.2 and 2.9 kJ mol^{-1}, respectively. However, a S117F substitution filled the larger cavity and stabilized (4.6 kJ mol^{-1}) the T4 lysozyme. For L10F substitution to stabilize WT*, the side chains surrounding the phenylalanine would only need to adjust slightly to remove potential steric clashes.

The WT* structure has been useful in refining our understanding of how mutations cause their kinetic and thermodynamic effects. We anticipate that the high-resolution structure studies of WT* and selected mutants will contribute substantially to our understanding of protein L stability and folding.

We are grateful to Dr Tommy Wan and Dr Brian Sutton for the protein L C3 domain coordinates. We would also like to thank Dr B. Stoddard, Dr R. Strong, Mr D. Goetz, Dr B. Kuhlman, Dr T. Nguyen and Dr B. Shen for useful comments, assistance and advice. This work was supported in part by grants from NIH NRSA T32 GM07270 and from the Fred Hutchinson Cancer Research Foundation.

References

Alberts, I. L., Nadassy, K. & Wodak, S. J. (1998). *Protein Sci.* **7**, 1700–1716.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Buckle, A. M., Cramer, P. & Fersht, A. R. (1996). *Biochemistry*, **35**, 4298–4305.
Buckle, A. M., Henrick, K. & Fersht, A. R. (1993). *J. Mol. Biol.* **234**, 847–860.
Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry*, Part II. New York: Freeman & Co.
Colovos, C. & Yeates, T. O. (1993). *Protein Sci.* **2**, 1511–1519.
Eriksson, A. E., Baase, W. A., Wozniak, J. A. & Matthews, B. W. (1992). *Nature (London)*, **355**, 371–373.
Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992). *Science*, **255**, 178–183.
Gu, H., Kim, D. & Baker, D. (1997). *J. Mol. Biol.* **274**, 588–596.
Gu, H., Yi, Q., Bray, S. T., Riddle, D. S., Shiau, A. K. & Baker, D. (1995). *Protein Sci.* **4**, 1108–1117.
Guex, N. & Peitsch, M. C. (1997). *Electrophoresis*, **18**, 2714–2723.
Jackson, S. E., Moracci, M., el Masry, N., Johnson, C. M. & Fersht, A. R. (1993). *Biochemistry*, **32**, 11259–11269.
Johnsen, K., O'Neill, J. W., Kim, D. E., Baker, D. & Zhang, K. Y. J. (2000). *Acta Cryst.* D**56**, 506–508.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.
Karpusas, M., Baase, W. A., Matsumura, M. & Matthews, B. W. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 8237–8241.
Kastern, W., Holst, E., Nielsen, E., Sjobring, U. & Bjorck, L. (1990). *Infect. Immun.* **58**, 1217–1222.
Kim, D. E., Fisher, C. & Baker, D. (2000). *J. Mol. Biol.* **298**, 971–984.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kleywegt, G. J. & Jones, T. A. (1994). *Acta Cryst.* D**50**, 178–185.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lüthy, R., Bowie, J. U. & Eisenberg, D. (1992). *Nature (London)*, **356**, 83–85.
McRee, D. E. (1992). *J. Mol. Graph.* **10**, 44–46.
Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.
Scalley, M. L., Yi, Q., Gu, H., McCormack, A., Yates, J. R. III & Baker, D. (1997). *Biochemistry*, **36**, 3373–3382.
Singh, J. & Thornton, J. M. (1990). *J. Mol. Biol.* **211**, 595–615.
Wikström, M., Drakenberg, T., Forsén, S., Sjöbring, U. & Björck, L. (1994). *Biochemistry*, **33**, 14011–14017.
Wikström, M., Sjöbring, U., Kastern, W., Björck, L., Drakenberg, T. & Forsén, S. (1993). *Biochemistry*, **32**, 3381–3386.
Xu, J., Baase, W. A., Baldwin, E. & Matthews, B. W. (1998). *Protein Sci.* **7**, 158–177.
Yi, Q. & Baker, D. (1996). *Protein Sci.* **5**, 1060–1066.