Protein Engineering, Design & Selection vol. 22 no. 2 pp. 93–101, 2009
Published online December 10, 2008 doi:10.1093/protein/gzn072

An antibody loop replacement design feasibility study and a loop-swapped dimer structure

Louis A. Clark[1,2], P. Ann Boriack-Sjodin, Eric Day, John Eldredge, Christopher Fitch, Matt Jarpe, Stephan Miller, You Li, Ken Simon and Herman W.T. van Vlijmen

Biogen Idec Inc., Cambridge, MA 02142, USA

[1]Present address: Codexis Inc., 200 Penobscot drive, Redwood City, CA 94063, USA.

[2]To whom correspondence should be addressed. E-mail: louie@alumni.northwestern.edu

A design approach was taken to investigate the feasibility of replacing single complementarity determining region (CDR) antibody loops. This approach may complement simpler mutation-based strategies for rational antibody design by expanding conformation space. Enormous crystal structure diversity is available, making CDR loops logical targets for structure-based design. A detailed analysis for the L1 loop shows that each loop length takes a distinct conformation, thereby allowing control on a length scale beyond that accessible to simple mutations. The L1 loop in the anti-VLA1 antibody was replaced with the L2 loop residues longer in an attempt to add an additional hydrogen bond and fill space on the antibody–antigen interface. The designs expressed well, but failed to improve affinity. In an effort to learn more, one design was crystallized and data were collected at 1.9 Å resolution. The designed L1 loop takes the qualitatively desired conformation; confirming that loop replacement by design is feasible. The crystal structure also shows that the outermost loop (residues Leu51–Ser68) is domain swapped with another monomer. Tryptophan fluorescence measurements were used to monitor unfolding as a function of temperature and indicate that the loop involved in domain swapping does not unfold below 60°C. The domain-swapping is not directly responsible for the affinity loss, but is likely a side-effect of the structural instability which may contribute to affinity loss. A second round of design was successful in eliminating the dimerization through mutation of a residue (Leu51Ser) at the joint of the domain-swapped loop.

Keywords: antibody loop design/antibody structure/binding affinity/protein design/protein stability

Introduction

Predictable control of loop conformation is essential to the expansion of accessible structure diversity for protein design. The combination of high throughput screening and rational design is becoming recognized as a highly effective way to run protein engineering projects, especially when solution-phase selection systems are not available (Fox *et al*., 2007). Recent improvements in energy functions and optimization procedures have led to many examples of successful computational protein design (Lippow and Tidor, 2007). These techniques can scan sequence space and focus evolution efforts. High throughput screening methods coupled with tailored DNA libraries can then be used to screen thousands of variants. Without large in silico conformation space exploration, the expected synergy between design and high throughput screening is not well realized. For example, small variations, such as design of specific mutations at a single residue position, are often not worth rationally designing because it is easier and more reliable to make and screen a small library. Computational design techniques need to be developed more in the area of sequence length and backbone shape variations.

There are still large unknowns in the protein design field. Most reported design calculations are done with a fixed backbone and have been successful in repacking small domains with non-native residues (Dahiyat and Mayo, 1997). Small variations in the backbone have allowed successes such as design of a novel fold (Kuhlman *et al*., 2003), peptide affinity increases (Sood and Baker, 2006) and conformational switches (Ambroggio and Kuhlman, 2006). Protein design success is dependent on the protein under consideration and it is often not clear why certain designs fail. For example, some proteins can be stabilized by repacking cores (Kuhlman and Baker, 2000; Korkegian *et al*., 2005) and some apparently cannot (Kuhlman and Baker, 2000; Mooers *et al*., 2003). To advance our understanding, it could be argued that it is equally important to study and learn from challenging design failures as to report successes.

Antibodies constitute the major growth class of biotherapeutics (Leader *et al*., 2008). Even small advances in antibody design capability could have significant impact on the success rate and development time of therapeutic antibodies. Computational design approaches for improvement of antibody affinity have focused on the hypervariable loops since these are the residues that are directly involved in antibody–antigen binding. Predictable control of loop conformation would be a very useful computational technology because it would significantly expand the accessible structural diversity for protein design.

Antibodies are a logical place to begin trying loop replacement strategies because there are many examples of loop length variations and the protein scaffolds are structurally conserved. They are one of the most studied and crystallized protein classes. It is the natural loop length and conformation variation in antibodies that make them so effective in their physiological role. There is clear precedent for believing that antibody loops can be cut and pasted into similar frameworks. Humanization is an established technique where a full set of CDR loops from a non-human antibody are grafted onto a highly similar human scaffold (Morea *et al*., 2000). Loop lengths are often effectively varied during solution-phase library-based antibody evolution (see Fellouse *et al*., 2007 for an example of cooperative length variation in the L3 and H3 loops). In a related class of β-sandwich proteins, the fibronectin type III domain, a recent success in designing a new 10-residue

loop conformation and validating it via crystallography is particularly encouraging (Hu *et al*., 2007).

Loop length variations have been attempted for antibodies and other immunoglobulin scaffolds. Lamminmaki *et al*. (Lamminmaki *et al*., 1999) made random peptide insertions in the CDR H2 and were successful in improving affinity by 12x using libraries and an *in vitro* selection system. Insertions and deletions in CDRs were advocated by Lantto and Ohlin (Lantto and Ohlin, 2002). Such length changing mutations are known to occur during the somatic hypermutation process (Wildt *et al*., 1999). Random loop insertions were also tried to investigate tolerance and designability for a type III fibronectin domain (Batori *et al*., 2002). It has been suggested that certain combinations of CDR loop lengths will result in antibodies biased to recognized different classes of antigens (Collis *et al*., 2003; Almagro, 2004). Recently, this hypothesis has been successfully tested (Almagro *et al*., 2006; Persson *et al*., 2006).

A number of interesting loop length variation libraries in combination with selection systems have been successful in enzyme redesign. Macbeath *et al*. (Macbeath *et al*., 1998) inserted a six-residue random peptide into the middle of a helix which contributes to the active site of a chorismate mutase and allows the enzyme to dimerize. Variants were found which retain activity and convert the protein to monomeric form, suggesting that the loop variation can be used to influence tertiary structure and activity simultaneously. Bocola *et al*. (Bocola *et al*., 2005) used structure and sequence information for homologs of monooxygenases to design a loop removal in the active site which improved the substrate acceptance of the enzyme. Insertion of loops into enzymes is tolerated to varying degrees, depending on location (Mathonet *et al*., 2006). Park *et al*. substantially modified four loops around the active site of a metallohydrolase to convert it into a β-lactamase (Park *et al*., 2006). Their use of sequence and structure information from the target enzyme class again suggests that a combination of rational design and DNA libraries with a selection system can be very powerful.

To explore the feasibility of rational loop replacement in antibodies, we attempted to execute a loop change which could have a useful application. The anti-VLA1 molecule has the potential to inhibit the entry of activated T-cells and monocytes to sites of inflammation by blocking binding of $\alpha_1\beta_1$-integrins (Ben-horin and Bank, 2004). We have previously been successful at using computationally designed point mutations to increase affinity of the anti-VLA1 complex by about an order of magnitude (Clark *et al*., 2006). We tried mutations at most of the interface residues and eventually began to suspect that only other strategies would yield further improvements. Visualization of the VLA-1 antigen complex crystal structure (Karpusas *et al*., 2003) with its antibody suggested that the antibody was making poor contact with the antigen in the vicinity of the L1 CDR loop. An attempt is described in this work to form additional interactions with the antigen by lengthening this loop.

Materials and methods

Design methodology

While the primary goal of the work was to test the feasibility of designed antibody CDR loop length changes, we also hoped to improve the affinity of the antibody. Figure 1 shows the structural rationale behind this secondary goal. The intention was to introduce additional contacts between the CDR L1 loop and the VLA1 antigen. A model of the intended interface structure is shown in Fig. 1 and compared to the wildtype complex crystal structure (Karpusas *et al*., 2003). Addition of two residues in the L1 loop was predicted to introduce a bulge in the loop at position Ser32 to make a new hydrogen bond with Glu115 on the antigen side of the interface.

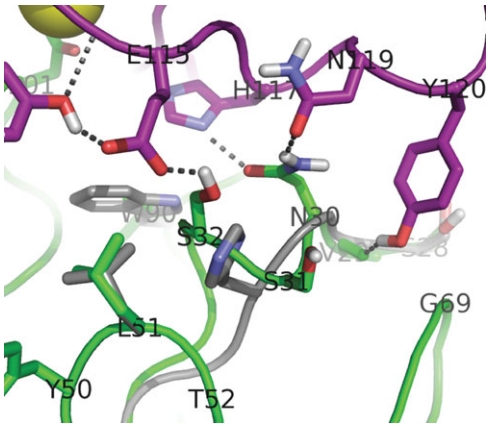

Fig. 1. One intended purpose of the CDR L1 loop lengthening was to add additional contacts on the antibody–antigen interface. If successful, the design would place L:Ser32 within hydrogen bonding distance of Glu115 on the antigen. The antigen is shown above in purple and the intended antibody conformation in green. For reference, the wildtype (1MHP) antibody structure is shown in gray. PyMOL was used to make all protein structure figures (DeLano, 2002).

Visualization of each loop length group was done to rule out loop lengths with obvious steric clashes. Groups of loops with a single length were fit to the wildtype crystal structure (Karpusas *et al*., 2003) (1MHP) at Ser24, Ala25, Met41 and Phe42 backbone positions. Within a given loop length, all loops had the same approximate conformation. This makes the loop length choice less complicated and increases the chance that the loop will take the desired conformation when inserted into the antibody framework and expressed. For each loop length, the fitted structures varied by 0.6 to 2.3 Å RMS, with the most variation observed for loop lengths of 12, 13 and 15. Representative examples of each loop length from 11 to 17 are shown in Fig. 2 overlayed onto the 1MHP framework. There are some notable exceptions to the conservation of conformation. Of the length 12 L1 loops, that from 1FIG takes a slightly different shape. Loops from 1CD0 and 1PEW of length 13 and the loop from 1BFV of length 16 differ from the typical conformations. It can be seen that a wide variety of loop shapes are available for design.

The design process utilizes known antibody structures with varying L1 loop lengths. Much can be learned about alternative loops through sequence examination. To assist selection of the new loop, we identified and grouped L1 loops of varying length from known crystal structures. Figure 3 shows a comparison of the sequences of a given length in sequence logo form. The multiple sequence alignment highlights the positions of relatively conserved hydrophobic residues which face inward to support the loops (black, larger letters). The wildtype antibody has an L1 loop length of 10 residues (top). Two gaps in the alignment, where additional residues insert,

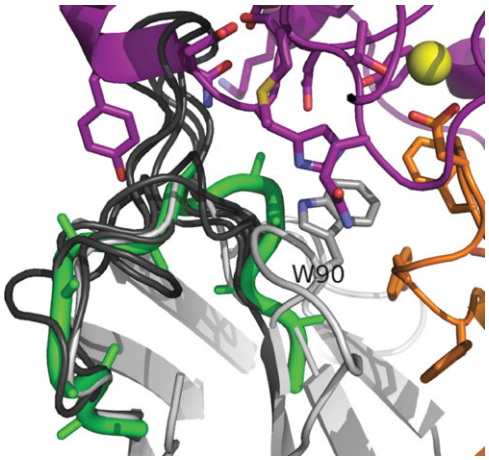

Fig. 2. A comparison of the effect of CDR L1 loop length on the expected loop conformation. Alternative loops with various lengths are shown in dark gray and green. Only the 12 residue loop (green tube, two-residue insertion) avoids severe steric clashes and seems to improve space-filling on the interface. Trp90 apparently clashes with the longer L1 loop and is mutated in many of the designs. The antigen is shown in purple, the wildtype light chain (1MHP) in light gray, the heavy chain in orange and the metal ion in yellow.

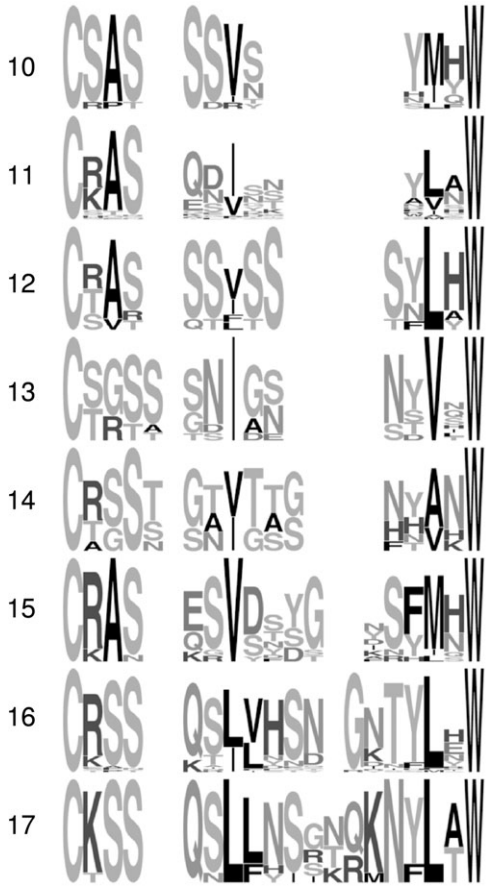

Fig. 3. Sequence logos (Crooks *et al*., 2004) showing the variation in sequence for a given loop length. The multiple alignment shows gaps where the extra residues present in the longer loops are inserted. Relatively conserved hydrophobic residue positions are colored black.

can be seen. One or two additional residues (length 11 or 12) add length to fill in the gap near the C-terminus of the loop. The result is a small bulge in the loop near w90 as seen in Fig. 2. At 13 residues, the additional residue instead adds

Table I. Summary of constructed CDR L1 loop length variants and sequences having +2 loop lengths in known crystal structures

Mutant	Antibody loop sequences		Other mutations	Affinity (nM)
	CDR L1	CDR L2		
Wildtype	CSASSSVN-HMFW	LTSNLASG	NA	10
LAC004	CSASSSVNSSAMFW	LTSNLASG	–	–
LAC006	CSASSSVNSSAMFW	LTSNLASG	W92I	–
LAC007	CSASSSVNSSALFW	LTSNLASG	W92M	–
LAC008	CSASSSVNSSALFW	LTSNLASG	W92I, F99W	–
LAC029	CSASSSVN-HMFW	LTSNLASG	W90I	600
LAC030	CSASSSVN-HMFW	LTSNLASG	W90M	–
LAC031	CSASSSVNSSALFW	LTSNLASG	W92I	120
LAC032	CSASSTVNSSALFW	LTSNLASG	W92I	–
LAC033	CSASSSVSSSALFW	LTSNLASG	W92I	–
LAC035	CSASSSVNSSALFW	STSNLASG	W92I	400
1AIF	CSVSSSISSSNLHW	GTSNLASG	NA	NA
1ORQ	CRARSSVSSSYLHW	STSNLASG	NA	NA
1ORS	CTASSSVSSSYLHW	STSNLASG	NA	NA
1KEN	CSASSTITSSFLYW	STSNLASG	NA	NA
1FIG	CRASSSVSSTYLHW	STSNLASG	NA	NA
35C8	CTASSSVSSSNLHW	STSNLASG	NA	NA
1IQD	CRASQSFSSSYLAW	GASTRATG	NA	NA

Residues 23–34 of the wildtype form the loop plus one conserved residue at head and tail. Examples of loop length 12 structures from the Protein Data Bank are given in the second portion of the table. L2 CDR loop sequences plus one proceeding residue position are given to show the variation at the critical L51 position.

length at the beginning of the loop and gives the overall loop a helical shape (see Fig. 2) unique to this length. A similar bulge in the beginning of the loop is present in the 14 residue loop, but absent for the remainder of lengths up to 17. The additional residues in loop length 15–17 come at the end of the loop and form a hairpin which protrude into the antigen in the back of Fig. 2. Variations in the loop shapes are partially rationalizable from the multiple sequence alignment.

Design rationale

The most attractive loop replacement involved a switch from loop length 10 to 12. Examples of loop length 12 sequences and their corresponding PDB ID's are given in Table I. The CDR L1 loop of the anti-VLA1 antibody was lengthened by two residues near the middle. A small number of variations were designed and constructed to investigate the effects of modifying the surrounding residues and to mitigate design failure for stability reasons. The sequences are given in Table I and the construction is described later in this section. An important change at Trp90 to a smaller residue was made to accommodate the loop lengthening. Note that the numbering of this position changes to 92 when a longer loop is inserted. Figures 1 and 2 show that Trp90 could sterically inhibit formation of the desired bulge where the two residues are inserted in the L1 loop.

In the second round of design, the crystal structure provided valuable information leading to stabilization of the fold. The sidechain of Leu51 in the designed sequence is rotated almost 180° in the crystal structure relative to the 1MHP structure (see Results). A change of this magnitude suggests a steric clash and alternative amino acids were sought at this position to stabilize the desired loop structures. Most of the L2 loop

sequences for the length 12 L1 loop donor structures are all identical to the wildtype except for the amino acid at the position preceding the L2 loop (Leu49 in wildtype, Kabat position 41). The consensus residue at this position is serine (Table I). Examination of the 1ORS structure (for example) provided a good solution to the problem and a Leu51Ser mutation was made in LAC031 to produce LAC035 and introduce a hydrogen bond between the L1 and L2 loops.

Structure preparation and simulation details

The starting complex structure at 2.8 Å was obtained from the Protein Data Bank (1MHP chains B, X and Y) and chains X and Y were relabeled as H and L. Asparagine, glutamine and histidine sidechain flips and histidine protonation states were corrected using suggestions generated by the WHATIF software (Vriend, 1990). The N-terminus of the light chain was patched with an acetyl group to compensate for the missing first residue in the 1MHP structure. The constant domains from the antibody's heavy and light chains were removed and the C-termini patched with N-methylamide groups. The structure was minimized using harmonic constraints (10 kcal/mol/$Å^2$ on all heavy atoms) and the CHARMM22 forcefield to remove major steric clashes.

To incorporate some backbone variation, design calculations were done on seven models. Each model was constructed by fitting the L1 loop from one of the crystal structures listed in Table I to the 1MHP structure at Ser24, Ala25, Met41 and Phe42 backbone positions. The L1 loop from the 1MHP structure was then eliminated and the substitute inserted using the minimization procedure described above. The DEZYMER package from the Hellinga lab (Benson *et al*., 2000; Looger and Hellinga, 2001; Wisz and Hellinga, 2003) was used to vary the identity of all L1 loop residues while allowing the rotamer state of neighboring residues outside the loop to change. As in our previous work (Clark *et al*., 2006), low energy rotamer choices were sought with the software as recommendations and final residue identity choices were made using our structural judgment. Information from the multiple models was incorporated in a consensus fashion for the final designs.

Molecular dynamics simulations using explicit water were performed for both the wildtype (1MHP) complex and for a model of the complex with the LAC031 sequence. The purpose was to investigate the stability of the complexes by allowing them to relax for 2 ns at 300 K in the NVT ensemble. A periodic cubic box sized approximately 75 Å × 66 Å × 61 Å and containing the complex was filled with TIP3 water plus sodium and chlorine counterions using VMD (Humphrey *et al*., 1996). The simulations were run using CHARMM22 potentials and the particle-mesh Ewald technique implemented in NAMD (Phillips *et al*., 2005). Each Fab was truncated at the end of the variable domain and the resulting C-termini were constrained to simulate the presence of the remainder of the Fab.

Protein production

Escherichia coli expressed His-tagged Fab fragments of anti-VLA1 and the described variants were used for this study. The plasmid consisting of the Fab fragment (Abraham *et al*., 2004), a 6-His tag and OmpA and PhoA periplasmic localization signals in a bicistronic arabinose-inducible vector was constructed. Amino acid substitutions were introduced into the anti-VLA1 Fab by QuikChange site-directed mutagenesis (Stratagene). Expression of the antibody variants was carried out in super broth using induction with 0.002% arabinose. Cells were pelleted, resuspended in 30 mM Tris, 20% sucrose with gentle mixing, repelleted and resuspended in 5 mM MgSO4 with vigorous mixing. The cells were again pelleted and the supernatant periplasmic lysate fraction was filtered.

Purification and expression

The various His-tagged mutants of anti-VLA1 were purified by passing the periplasmic supernatant over a 1 ml HisTrap HP nickel chelate column (Pharmacia) which had been equilibrated with 50 mM sodium phosphate 300 mM NaCl 20 mM imidazole 0.05% Tween 20 pH 8.0. After 10 column washings, the Tween 20 was removed from the buffer and then the imidazole concentration was increased to 250 mM for elution. The appropriate sample peak was collected and dialyzed into a 1% PBS buffer pH 7.0 plus 0.02% sodium azide. The proteins were analyzed by PAGE, UV scan, mass spectrometry and size exclusion chromatography (SEC)/light scattering to look for aggregates and multimers. In preparation for crystallization, the LAC031 protein was purified by SEC on a Superdex 200 (Amersham) column using a 20 mM Tris, 100 mM NaCl, pH 7.5 buffer. The result was concentrated to 9.8 mg/ml using a Centricon-10 ultrafiltration device (Millipore). Aliquots were flash frozen in liquid nitrogen and thawed as needed for crystallization trials.

Recombinant $\alpha_1\beta_1$-integrin I-domain (humanized rat) was expressed in *E.coli* as a GST-fusion protein. The I-domain was cleaved from the purified GST-fusion with thrombin and further purified as described previously (Gotwals *et al*., 1999).

For crystallography, Fabs are expressed using the strain W3110ara, a Δara derivative of W3110. A 100 ml culture containing the expression plasmid, at a density of >109 cells/ml, was used to inoculate (1:100) a 5 l fermenter with fermentation medium containing 2% fructose and 100 μg of ampicillin. The fermentation culture was grown overnight at a pH of 7.0 with the dissolved oxygen maintained at 30% by fructose titration. The culture was induced at 15 ODs with a final concentration of 0.02% arabinose. At the time of induction, 250 ml of induction medium (Amisoy, 80 g/l; yeast extract, 20 g/l; L-proline 12 g/l; L-leucine, 12 g/l; tryptophan, 6 g/l) was added. The culture was harvested 3 h after induction at an OD600 of about 25–30.

Fluorescence measurements

Tryptophan fluorescence experiments were done in 2 ml cuvettes using a 20 μM solution of protein. Samples were excited at 290 nm and emission was measured at 300–400 nm on a SIM AMINCO Bowman Series 2 Luminescence Spectrometer attached to a temperature-controlled waterbath. The emission spectrum has one peak around 325 nm and a second much broader peak around 340 nm which extends past 360 nm. As the temperature is increased, the entire emission spectrum attenuates and the second peak begins to red-shift around 60°C. Buried and exposed tryptophans fluorescence at ca. 340 nm and 355 nm, respectively (Royer, 1995). Denaturation via tryptophan exposure can be quantified using the ratio of emission at 355–340 nm. Temperature course experiments were done by equilibrating for 15 min, measuring fluorescence then resetting the waterbath temperature to the next

temperature point. White fibrous precipitate was eventually observed in all samples after sustained exposure to temperature higher than approximately 55°C. All temperature course experiments were done in triplicate to assess reproducibility under these non-equilibrium irreversible denaturation conditions.

Affinity measurements

To estimate the fold change in affinity of mutant anti-VLA1 proteins, we used a competition ELISA. In this assay, GST I-domain fusion protein was coated onto an ELISA plate. A dilution series of anti-VLA1 Fab samples was incubated with 10 nM biotin anti-VLA1 Fab on the plate for 2 h at room temperature in HEPES buffer with 150 mM NaCl, 1 mM $MgCl_2$, 0.05% Tween-20 and 1% BSA. The plate was washed and the amount of biotinylated anti-VLA1 Fab bound was determined using streptavidin HRP as a secondary assay. The fold change in affinity versus wildtype was determined by comparing the EC50 of binding to wildtype Fab measured on the same plate.

To measure the solution-phase affinity, we employed the KinExA 3000 (Sapidyne Instruments, Boise, Idaho). Polystyrene beads were coated with GST I-domain fusion protein by passive adsorption. Purified anti-VLA1 Fab fragment is flowed through the column to bind to the I-domain on the bead. The Fab is detected with a secondary anti-mouse IgG F(ab′)2 fragment specific antibody conjugated with the fluorescent dye Cy5. A dilution series of soluble I-domain protein with 3 h equilibration is used. The amount of free anti-VLA1 Fab that remains in solution is determined by the intensity of the fluorescence signal. A non-linear regression curve fit gives a Kd value.

Crystallization and crystallography

Crystallization was performed using 1.0 μl of the SEC-purified LAC031 protein solution plus 1.0 μl of crystallization solution in a hanging drop over 1 ml of the crystallization solution. Crystals formed within 2 days using a pH 6.5 solution of 16–20% PEG 3350, 0.1M sodium cacodylate (NaCaco), 0.2M $MgCl_2$ at 18C. The final crystal quality and sized were greatly enhanced using microseeding techniques, moderate cooling (10°C) and a reduction of precipitant concentration (10–12% PEG 3350).

An initial data set for LAC031 was collected on a rotating anode source and the structure was solved using constant and variable domains from a related structure (PDB: 2B2X) (Clark *et al*., 2006) as separate probes during molecular replacement utilizing the program AMoRe (Navaza, 1994). Subsequently, this solution was used for refinement of the higher resolution data set collected at beamline X29 at the National Synchrotron Light Source and presented here. Model building was performed using O (Jones *et al*., 1991), and simulated annealing protocols in CNX (Brunger *et al*., 1998), utilizing bulk solvent corrections and anisotropic B-factor scaling protocols, were used for refinement of the final structure. Data collection and refinement statistics are listed in Table II. The structure has been deposited in the Protein Data Bank with identifier 3EOT.

Results

Production of all loop replacement variants was successful. All constructs expressed and gave single bands at ca. 47 kDa on SDS–PAGE gels after purification. The loop changes did not destabilize the protein enough to prevent expression. Affinities for all of the constructs in Fab format were tested in primary assay using ELISA. Unfortunately, none of the loop change designs improved the affinity of the antibody with the VLA1 antigen. All loop replacement variants had affinities which approached or exceeded 100 times lower than the wildtype and were outside the reliable range of our ELISA measurements.

Table II. Crystallography data collection and refinement statistics

	Anti-VLA1 antibody LAC031
Data collection	
Space group	P2(1)2(1)2
Cell dimensions	
a, *b*, *c* (Å)	83.30, 132.46, 41.69
α, *β*, *γ* (°)	90.0, 90.0, 90.0
Resolution (Å)	50–1.8 (1.86–1.8)[a]
R_{sym}^{b}	0.058 (0.215)
$I/\sigma I$	39.0 (4.5)
Completeness (%)	93.5 (68.1)
Redundancy	6.7 (5.3)
Refinement	
Resolution (Å)	35.0–1.8
No. of reflections	40 525
R_{work}/R_{free}^{c}	0.243/0.286
No. of atoms	
Protein	3249
Water	190
RMS deviations	
Bond lengths (Å)	0.00430
Bond angles (°)	0.878

[a]Numbers in parentheses refer to outer resolution shell.
[b]$R_{sym} = \sum |I - < I > | / \sum I$, where I is the integrated intensity of a given reflection.
[c]$R_{value} = \sum |F(obs) - F(calc)| / \sum F(obs)$, where F(calc) represents the structure factor amplitudes obtained from back transformation of the model. Ten percent of the data was used in calculating R_{free}.

Initially, two constructs deemed most important for design analysis were selected and their affinities were measured using the solution phase KinExA assay. LAC029 is identical to wildtype, but has a W90I mutation intended to create space for the lengthened L1 loop. It can be seen that the W90I mutation by itself substantially disrupts the binding, decreasing the affinity from 10 to 600 nM. LAC031 includes the W90I mutation plus the loop lengthening changes (Table I). Some of the affinity is recovered in the full LAC031 variant, but is still more than 10x lower than wildtype (120 versus 10 nM).

The LAC031 antibody Fab exists in equilibrium with a dimer form. At room temperature in PBS buffer and concentrated to 1.7 mg/ml, it is 21% dimer as determined by SEC. At 10 mg/ml, it is ca. 30% dimer. Dimerization is reversible upon dilution with k_{off} ca. 17 h. We initially suspected that the dimerization was related to the loss of affinity. The dimerization could directly affect the binding if the binding interface is disrupted. Alternatively, the dimerization could simply be a symptom of CDR destabilization which would also decrease binding affinity by requiring more entropically unfavorable ordering upon binding.

In an effort to learn more about the designed loop replacement, LAC031 was crystallized (see Materials and Methods). The protein in Fab form diffracted to 1.9 Å with an R_{free} of

Fig. 4. Light chain portion overview of the Fab LAC031 crystal structure showing domain swapping. The loop involved in the domain swapping is shown as tube cartoons. Tryptophan residues, two of which are buried under the mobile loop, are shown as red spheres. The placement of the antigen (purple) from the original 1MHP structure is shown in transparency to demonstrate that the dimerization would inhibit or obviate binding by blocking the paratope.

28.6%. Crystallization samples were prepared from a monomer peak by SEC purification, but the crystal structure is clearly a dimer. Figure 4 shows the surprise dimer form of the Fab where the loop (residues Leu51–Ser68) which is C-terminal to the L2 CDR loop is flipped away from the framework to make similar contacts to a crystallographically related Fab. This is a form of domain-swapping where only a single loop swaps. Most of the contacts on the new dimer interface are exactly as they were in the wildtype crystal structure. The new interface contacts [often called the 'open' interface (Bennett *et al*., 1995; Liu and Eisenberg, 2002)] are only made by residues Ser68 on the C-terminal end of the swapping loop and residues Leu51–Ser53 on the N-terminal side. Residues Tyr50 and Asn54 are also displaced from their analogous positions in the wildtype.

In dimer form, LAC031 would be expected to have very low affinity for the VLA1 antigen. The domain-swapping positions the paired monomer to sterically block the CDR binding site (Fig. 4). The joined VL domain from the other chain would need to move to make way for binding. However, because less than 20% of the protein is present as dimer under assay condition, this cannot be the main cause for the affinity decrease.

Figure 5 shows the structure and a comparison to the wildtype and intended design. In most areas, the intended and final structures fall within the gray tubes, indicating there is no substantial difference between the wildtype and the design. It can be seen that the intended L1 loop structure is similar to that in the dimer crystal structure despite the movement of the domain-swapping loop which forms the nearby L2 loop. In other areas, there are differences. The L1

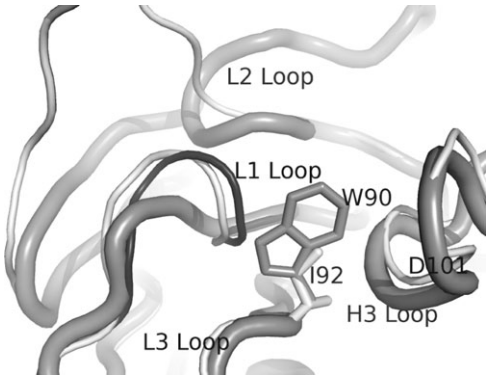

Fig. 5. Detail of the CDR L1 loop region showing that the intended loop conformation (black) is similar to the crystal structure (white). The conformation of the wildtype structure (1MHP) is shown as thicker gray tubes. A shift is also seen in the position of the H3 loop in the LAC031 relative to the wildtype. The tryptophan at position 90 in the wildtype is shown compared to the smaller isoleucine in the designed structure at position 92.

loop takes the desired conformation, thereby establishing that it is possible to replace single CDR loops.

Differences in structure around the L3 and H3 loops are potentially important. Figure 5 also shows the Trp90 to Ile92 mutation. Recall that this substitution was done to eliminate potential steric conflict with the bulge in the new longer L1 loop. The mutation or some other factor has changed the conformation of the H3 loop. It cannot be excluded that the difference is due to the antigen-bound wildtype versus unbound LAC031. Changes in the H3 loop may be critical because it carries the Asp101 amino acid which coordinates a metal ion on the antigen side of the interface. Previous work has shown that approximately two orders of magnitude affinity are lost when the metal binding is disrupted by mutation or metal depletion (Karpusas *et al*., 2003).

Perhaps the most interesting area of the dimer crystal structure is the L2 loop region where the backbone adopts a different conformation to produce the domain swapping. Residues Leu51–Ser53 in the dimer structure have identical residue types in the wildtype, but make contacts which differ substantially. They form part of the dimer interface (see Fig. 6B). The most substantial change is at Leu51, which is facing outward in the wildtype (Leu49, Fig. 6A) structure. It rotates 180° inward to form part of the hydrophobic core in the dimer structure (Fig. 6B). Such a large change suggests that the residue at this position could substantially affect the tendency for the molecule to dimerize. There are no obvious steric clashes between Leu51 and its neighbors in the model design (not a dimer). It is possible that the hydrophobic character of the leucine is sufficient to bury it in the core and induce the domain swapping.

Tryptophan fluorescence experiments were conducted as a function of temperature to assess the state of the loop-swapping domain in solution. There are two tryptophan residues under the mobile loop which would be exposed if the loop was flipped out (Fig. 4). LAC029 (W90I) was used as the control because it contains the same number of tryptophans as the tested loop replacement design (LAC031). Figure 7 shows that the loop replacement design has a more gradual transition to the unfolded state. It begins to expose tryptophans at around 55°C, a lower temperature than the control.

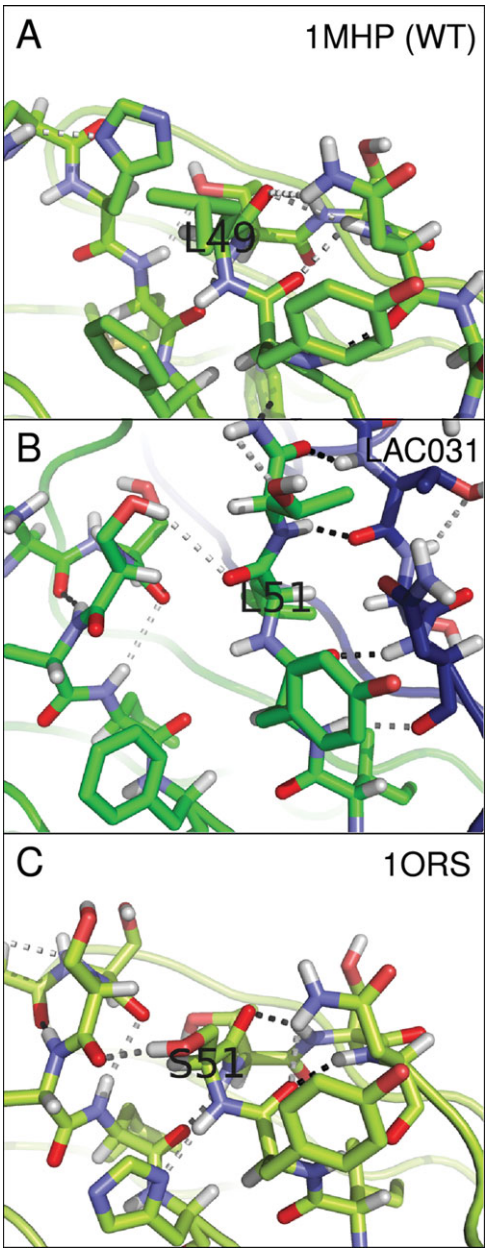

Fig. 6. Examination of the region around Leu51/Ser51 in the designed (**B**) structure shows that the Leu51 residue is flipped 180° relative to the analgous Leu49 residue in the wildtype (**A**) structure. In the 1ORS structure, there is a serine at position 51 which forms a stabilizing sidechain-backbone hydrogen bond with the L1 loop. All structures are shown in the same orientation and fitted to the wildtype (A) 1MHP structure. Hydrogen bonds were calculated using the Rosetta package (ROSETTA software package, 2006) and are darker for the stronger bonds.

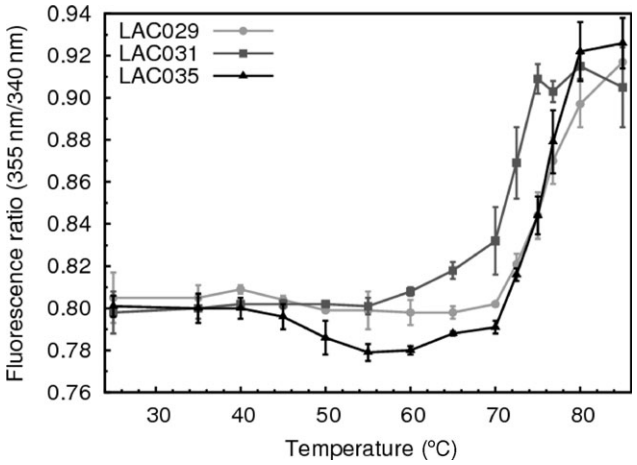

Fig. 7. Tryptophan exposure as a function of temperature shows that the loop replacement variant (LAC031) has an uncooperative folding transition which begins around 55°C. Error bars are at one standard deviation based on triplicate experiments.

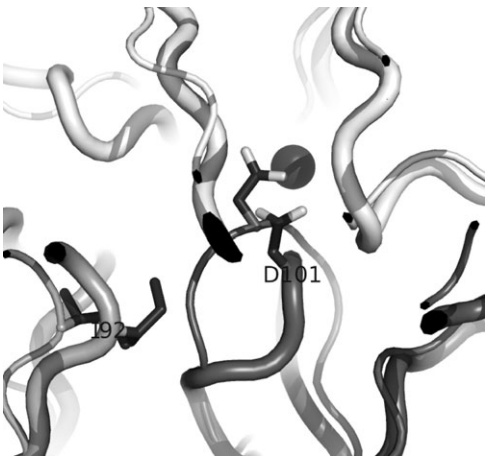

Fig. 8. Before and after a 2 ns MD simulation of the designed model complex (LAC031). The starting structure is displayed as thicker tubes. The H3 loop which bears the Asp101 residue has rotated substantially during the course of the simulation. Destabilization of the H3 loop may be the cause of the decreased affinity for the designed antibody.

The second round design (LAC035) corrects the unfolding at lower temperatures. It also restores the cooperativity of the unfolding transition to what is seen for the control variant. In contrast to the LAC031 design, the improved variant has no discernable oligomer peaks during an SEC purification. The apparent connection here between greater cooperativity and elimination of dimerization may be a general phenomenon (Clark, 2005).

A pair of two nanosecond molecular dynamics simulations were performed in explicit water to investigate the stability of the designed, loop replaced variable domains. The simulations were not long or numerous enough to sample ensemble dynamics and should only be interpreted qualitatively. One trajectory was started from the 1MHP crystal structure coordinates as a control. Regions of the designed structure where the final conformation simulation differs substantially from the initial suggest local instabilities. Perhaps the most strikingly different region is that surrounding the H3 loop. Figure 8 shows this region in detail. After only 2 ns, the H3 loop has deformed and rotated substantially. While the H3 loop is relatively short at 10 amino acids, it is glycine rich (GFGDGGYFDV) and therefore flexible. Movement in the region near the Trp92Ile mutation may be related. The wildtype simulation does not show substantial H3 loop movement.

Discussion

The successful replacement of the L1 loop with a longer loop from another crystal structure demonstrates that this is a feasible strategy for rational design of antibodies. Crystallization of the designed antibody fragment showed that the loop takes the desired conformation. It also showed a

surprise domain swapped structure in which the outermost loop C-terminal to the L2 loop is displaced and swapped with an adjacent Fv domain. A single amino acid change in another round of design corrected the domain swapping, but did not result in increased affinity for the antigen.

Lessons learned

It is important to analyze the affinity loss of the designed antibody relative to the wildtype antibody. Recall that the secondary goal was to increase affinity by adding additional contacts across the antibody–antigen interface (see Fig. 1). The domain swapping of the loop replacement design would likely prevent or inhibit antigen binding. However, the dimerization only occurs to a significant extent at concentrations higher than those used during affinity measurements. From the tryptophan fluorescence studies, we know that the domain-swapping loop is not detached from the framework at temperatures lower than 55°C. The reason for the affinity loss must therefore be more subtle than larger tertiary changes.

One strong possibility for the failure of the design is the destabilization of the CDR H3 loop structure. The H3 loop carries the aspartate residue which binds the Mn or Mg ion on the antigen side. This interaction is crucial since removal of the metal ion reduces the affinity about 100-fold (Karpusas *et al*., 2003). A single molecular dynamics simulation in explicit water shows that after only 2 ns the H3 loop has rotated substantially (Fig. 8). A control simulation of the wildtype complex showed marginal movement of the H3 loop, suggesting that the Trp92Ile mutation may be responsible for the rearrangement. The larger tryptophan may be required to brace the base of the H3 loop.

Aspects of domain-swapping in protein design

Lessons can also be learned from the domain-swapping. Dimerization of designed proteins is generally undesirable. The example given in this work shows how easily the balance between correctly folded monomer and domain-swapped dimer can be upset. Toggling between the hydrophobic leucine at position 51 and serine is sufficient to eliminate dimerization, as shown by the absence of multimer peaks in the SEC chromatogram. This fragility is consistent with reports of domain swapping in protein L (Kuhlman *et al*., 2001; O'neill *et al*., 2001) and protein G (Frank *et al*., 2002) mutants. During the design process, one must be careful to maximize productive contacts even if it requires changing residues outside the replacement loop. Destabilization of the desired structure likely contributes to domain-swapping.

The domain swapping phenomena may be related to aggregation and the formation of insoluble precipitate. Formation of domain-swapped interfaces has been suggested to be one possible aggregation mechanism (Klafki *et al*., 1993; Janowski *et al*., 2001; Liu *et al*., 2001). Design techniques to avoid domain swapping may therefore be useful in designing aggregation-resistant proteins. Strategies for designing against aggregation have been suggested based on observations in crystallized structures and should be tested (Richardson and Richardson, 2002).

Domain swapping in antibody constructs has been seen previously and documented in published crystal structures. Full variable domain antibody domain-swapping is most common (Perisic *et al*., 1994; Pei *et al*., 1997; Calarese *et al*., 2003). It may even have applications, for example, in enlarging the antibody–antigen interface (Calarese *et al*., 2003). The type of loop-only domain swapping observed in our work has been seen previously only in a camelid antibody (Spinelli *et al*., 2004).

Domain swapped conformations may be predictable and therefore could potentially be avoided via negative design. Ding *et al*. (Ding *et al*., 2006) have described a set of molecular dynamics simulations using a Go-like potential model where they are able to find dimer structures which match the domain-swapped crystal structures qualitatively. They are also able to predict hinge regions where partial unfolding can lead to domain swapping based on a connectivity model.

There may be applications for designing reversible concentration-dependent dimerization into existing proteins. Controlled oligomerization may hinder the formation of aggregates by burying aggregation-prone regions or blocking unfolding routes. In a related example, immobilized enzymes deactivated much more slowly than soluble enzymes in a number of test cases (Schellenberger and Ulbrich, 1989). Biotherapeutics are typically stored at high concentrations for extended periods. Both conditions may contribute to aggregation. If a protein were designed to reversibly oligomerize at high concentrations, then it would be partially protected from degradation, yet it would recover its monomer functional state during the dilution process before injection.

Perspective and future applications

The success of this work in establishing that single loops can be replaced should help encourage protein engineers to expand designs into sequence length and conformation space. At the same time, the domain-swapped crystal structure nicely illustrates what can happen at an atomic level when a design fails to maintain stability. Expansion of this loop replacement technique to more difficult, less characterized scaffolds could be pursued with a combined computational and experimental approach. In silico screening to improve contacts coupled with an *in vitro* library approach will maximize the chances of successful design.

Acknowledgements

This work was made possible through support from the Biogen Idec postdoctoral fellow program. We thank Dr Martin McMillan and the beamline support staff at X29 at the National Synchrotron Light Source at Brookhaven National Laboratories for data collection. Advice from Graham Farrington and Fred Taylor is gratefully acknowledged. Support from Scott Glaser and Adrian Whitty is also appreciated.

References

Abraham,W.M., *et al*. (2004) *Am. J. Respir. Crit. Care Med.*, **169**, 97–104.
Almagro,J.C. (2004) *J. Mol. Recognit.*, **17**, 132–143.
Almagro,J.C., Quintero-hernandez,V., Ortiz-leon,M., Velandia,A., Smith,S.L. and Becerril,B. (2006) *J. Mol. Recognit.*, **19**, 413–422.
Ambroggio,X.I. and Kuhlman,B. (2006) *J. Am. Chem. Soc.*, **128**, 1154–1161.
Batori,V., Koide,A. and Koide,S. (2002) *Protein Eng.*, **15**, 1015–1020.
Ben-horin,S. and Bank,I. (2004) *Clin. Immunol.*, **113**, 119–129.
Bennett,M.J., Schlunegger,M.P. and Eisenberg,D. (1995) *Protein Sci.*, **4**, 2455–2468.
Benson,D.E., Wisz,M.S. and Hellinga,H.W. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 6292–6297.
Bocola,M., Schulz,F., Leca,F., Vogel,A., Fraaije,M.W. and Reetz,M.T. (2005) *Adv. Synth. Catal.*, **347**, 979–986.
Brunger,A.T., *et al*. (1998) *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.

Calarese,D.A., *et al*. (2003) *Science*, **300**, 2065–2071.
Clark,L.A. (2005) *Protein Sci.*, **14**, 653–662.
Clark,L.A., *et al*. (2006) *Protein Sci.*, **15**, 949–960.
Collis,A.V., Brouwer,A.P. and Martin,A.C. (2003) *J. Mol. Biol.*, **325**, 337–354.
Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) *Genome Res.*, **14**, 1188–1190.
Dahiyat,B.I. and Mayo,S.L. (1997) *Science*, **278**, 82–87.
DeLano,W. (2002) *DeLano Scientific*, http://www.pymol.org.
Ding,F., Prutzman,K.C., Campbell,S.L. and Dokholyan,N.V. (2006) *Structure*, **14**, 5–14.
Fellouse,F.A., *et al*. (2007) *J. Mol. Biol.*, **373**, 924–940.
Fox,R.J., *et al*. (2007) *Nat. Biotechnol.*, **25**, 338–344.
Frank,M.K., Dyda,F., Dobrodumov,A. and Gronenborn,A.M. (2002) *Nat. Struct. Biol.*, **9**, 877–885.
Gotwals,P.J., Chi-rosso,G., Ryan,S.T., Sizing,I., Zafari,M., Benjamin,C., Singh,J., Venyaminov,S.Y., Pepinsky,R.B. and Koteliansky,V. (1999) *Biochemistry*, **38**, 8280–8288.
Hu,X., Wang,H., Ke,H. and Kuhlman,B. (2007) *Proc. Natl Acad. Sci. USA*, **104**, 17668–17673.
Humphrey,W., Dalke,A. and Schulten,K. (1996) *J. Mol. Graph.*, **14**, 33–38.
Janowski,R., Kozak,M., Jankowska,E., Grzonka,Z., Grubb,A., Abrahamson,M. and Jaskolski,M. (2001) *Nat. Struct. Biol.*, **8**, 316–320.
Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr. A*, **47**, 110–119.
Karpusas,M., Ferrant,J., Weinreb,P.H., Carmillo,A., Taylor,F.R. and Garber,E.A. (2003) *J. Mol. Biol.*, **327**, 1031–1041.
Klafki,H.W., Pick,A.I., Pardowitz,I., Cole,T., Awni,L.A., Barnikol,H.U., Mayer,F., Kratzin,H.D. and Hilschmann,N. (1993) *Biol. Chem. Hoppe Seyler*, **374**, 1117–1122.
Korkegian,A., Black,M.E., Baker,D. and Stoddard,B.L. (2005) *Science*, **308**, 857–860.
Kuhlman,B. and Baker,D. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 10383–10388.
Kuhlman,B., O'neill,J.W., Kim,D.E., Zhang,K.Y.J. and Baker,D. (2001) *Proc. Natl Acad. Sci. USA*, **98**, 10687–10691.
Kuhlman,B., Dantas,G., Ireton,G.C., Varani,G., Stoddard,B.L. and Baker,D. (2003) *Science*, **302**, 1364–1368.
Lamminmaki,U., Pauperio,S., Westerlund-karlsson,A., Karvinen,J., Virtanen,P.L., Lovgren,T. and Saviranta,P. (1999) *J. Mol. Biol.*, **291**, 589–602.
Lantto,J. and Ohlin,M. (2002) *J. Biol. Chem.*, **277**, 45108–45114.
Leader,B., Baca,Q.J. and Golan,D.E. (2008) *Nat. Rev. Drug Discov.*, **7**, 21–39.
Lippow,S.M. and Tidor,B. (2007) *Curr. Opin. Biotechnol.*, **18**, 305–311.
Liu,Y. and Eisenberg,D. (2002) *Protein Sci.*, **11**, 1285–1299.
Liu,Y.S., Gotte,G., Libonati,M. and Eisenberg,D. (2001) *Nat. Struct. Biol.*, **8**, 211–214.
Looger,L.L. and Hellinga,H.W. (2001) *J. Mol. Biol.*, **307**, 429–445.
Macbeath,G., Kast,P. and Hilvert,D. (1998) *Science*, **279**, 1958–1961.
Mathonet,P., Deherve,J., Soumillion,P. and Fastrez,J. (2006) *Protein Sci.*, **15**, 2323–2334.
Mooers,B.H., Datta,D., Baase,W.A., Zollars,E.S., Mayo,S.L. and Matthews,B.W. (2003) *J. Mol. Biol.*, **332**, 741–756.
Morea,V., Lesk,A.M. and Tramontano,A. (2000) *Methods*, **20**, 267–279.
Navaza,J. (1994) *Acta Cryst. A*, **50**, 157.
O'neill,J.W., Kim,D.E., Johnsen,K., Baker,D. and Zhang,K.Y.J. (2001) *Structure*, **9**, 1017–1027.
Park,H.S., Nam,S.H., Lee,J.K., Yoon,C.N., Mannervik,B., Benkovic,S.J. and Kim,H.S. (2006) *Science*, **311**, 535–538.
Pei,X.Y., Holliger,P., Murzin,A.G. and Williams,R.L. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 9637–9642.
Perisic,O., Webb,P.A., Holliger,P., Winter,G. and Williams,R.L. (1994) *Structure*, **2**, 1217–1226.
Persson,H., Lantto,J. and Ohlin,M. (2006) *J. Mol. Biol.*, **357**, 607–620.
Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E., Villa,E., Chipot,C., Skeel,R.D., Kale,L. and Schulten,K. (2005) *J. Comp. Chem.*, **26**, 1781–1802.
Richardson,J.S. and Richardson,D.C. (2002) *Proc. Natl Acad. Sci. USA*, **99**, 2754–2759.
ROSETTA software package. (2006) version 2.2.0. University of Washington, Seattle, available from http://www.rosettacommons.org/.
Royer,C.A. (1995) *Fluorescence spectroscopy*. **Vol. 40**. Humana Press, Totowa, NJ, pp. 65–89.
Schellenberger,A. and Ulbrich,R. (1989) *Biomed. Biochim. Acta*, **48**, 63–67.
Sood,V.D. and Baker,D. (2006) *J. Mol. Biol.*, **357**, 917–927.
Spinelli,S., Desmyter,A., Frenken,L., Verrips,T., Tegoni,M. and Cambillau,C. (2004) *Febs. Lett.*, **564**, 35–40.
Vriend,G. (1990) *J. Mol. Graph.*, **8**, 52–56.
Wildt,R.M.D., Venrooij,W.J.V., Winter,G., Hoet,R.M. and Tomlinson,I.M. (1999) *J. Mol. Biol.*, **294**, 701–710.
Wisz,M.S. and Hellinga,H.W. (2003) *Proteins*, **51**, 360–377.

Received June 1, 2008; revised October 4, 2008; accepted November 4, 2008

Edited by Feng Ni