Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The high-resolution structure of (+)-*epi*-biotin bound to streptavidin

Isolde Le Trong,[a] Dimitri G. L. Aubert,[b] Neil R. Thomas[b] and Ronald E. Stenkamp[a]*

[a]Departments of Biological Structure and Biochemistry, Biomolecular Structure Center, University of Washington, Seattle, WA 98195, USA, and [b]Centre for Biomolecular Sciences, School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, England

Correspondence e-mail: stenkamp@u.washington.edu

Received 14 November 2005
Accepted 1 April 2006

PDB References: (+)-*epi*-biotin–streptavidin complex, 2gh7, r2gh7sf; 2f01, r2f01sf.

(+)-*Epi*-biotin differs from (+)-biotin in the configuration of the chiral center at atom C2. This could lead to a difference in the mode of binding of (+)-*epi*-biotin to streptavidin, a natural protein receptor for (+)-biotin. Diffraction data were collected to a maximum of 0.85 Å resolution for structural analysis of the complex of streptavidin with a sample of (+)-*epi*-biotin and refinement was carried out at both 1.0 and 0.85 Å resolution. The structure determination shows a superposition of two ligands in the binding site, (+)-biotin and (+)-*epi*-biotin. The molecules overlap in the model for the complex except for the position of S1 in the tetrahydrothiophene ring. Differences in the conformation of the ring permits binding of each molecule to streptavidin with little observable difference in the protein structures at this high resolution.

1. Introduction

The tight binding of biotin to streptavidin ($K_d \simeq 10^{-13}$ *M*; Green, 1963) is the basis for its use in biotechnical applications where tight coupling of molecules is involved. Protein-engineering and synthetic chemistry techniques allow the attachment of these molecules to cells/surfaces/molecules that can then be directed and bound to specific sites containing the other member of the binding duo (for examples, see Schechter *et al.*, 1996; Hirsch *et al.*, 2002; Roberts *et al.*, 2004; Schiestel *et al.*, 2004). Complexes containing streptavidin and biotin or other ligands have also served as model systems for the study of protein–ligand interactions (Chilkoti *et al.*, 1995; Reznik *et al.*, 1998; Freitag *et al.*, 1999; Hyre *et al.*, 2000).

To aid in engineering specific and controlled binding activities, extensive investigations of potential ligands for streptavidin have shown that the protein can bind a range of ligands in addition to biotin. Peptides (Katz & Cass, 1997; Korndoerfer & Skerra, 2002) and a number of biotin derivatives and analogs (Weber *et al.*, 1994; Katz, 1997; Pazy *et al.*, 2002) have been studied to determine how the binding affinity for these ligands changes as different functional groups are utilized in forming the protein–ligand complex.

We have initiated a study of the interactions between biotin stereoisomers and streptavidin (Aubert *et al.*, 2006). Here, we report the crystal structure of streptavidin in complex with (+)-*epi*-biotin. These two molecules differ in the configuration at C2, the chiral C atom in the tetrahydrothiophene ring. The goal of our crystallographic studies was to determine how this biotin derivative binds to streptavidin.

© 2006 International Union of Crystallography
Printed in Denmark – all rights reserved

Table 1
Summary of diffraction data statistics.

Values in parentheses are for the last shell.

Data set	I	II
Data-collection device	R-AXIS II	SSRL
Maximum resolution (Å)	1.50 (1.53–1.50)	0.85 (0.86–0.85)
Unique reflections	29420	189859
Average redundancy	4.3	7.2
Completeness (%)	80.1 (10.1) [95.5 to 1.7 Å resolution]	95.0 (83.0)
R_{merge} (%)	0.070 (0.184)	0.045 (2.585)
$\langle I\rangle/\langle\sigma(I)\rangle$	22.1 (4.2)	55.5

Table 2
Detailed diffraction data statistics for data set II.

Resolution shells (Å)	No. of reflections	Completeness (%)	$\langle I\rangle$	$\langle$Error$\rangle$	R_{merge}
50.00–2.31	10354	99.1	1320772.7	13935.0	0.024
2.31–1.83	10098	99.3	389148.4	6917.2	0.060
1.83–1.60	9969	98.8	127285.5	2918.4	0.130
1.60–1.45	9652	96.4	62429.8	2517.2	0.063
1.45–1.35	9803	97.4	37145.9	1829.4	0.083
1.35–1.27	9703	97.0	27656.3	1427.8	0.097
1.27–1.21	9663	96.7	21778.8	1191.9	0.106
1.21–1.15	9601	96.2	19316.5	1100.7	0.110
1.15–1.11	9542	95.9	14429.3	908.3	0.127
1.11–1.07	9531	95.6	9708.4	715.9	0.155
1.07–1.04	9476	95.1	5790.1	570.4	0.211
1.04–1.01	9406	94.8	4030.3	508.7	0.269
1.01–0.98	9414	94.5	2764.0	457.4	0.359
0.98–0.96	9347	94.1	1767.9	429.2	0.508
0.96–0.94	9290	93.8	1307.5	418.5	0.664
0.94–0.92	9255	93.3	941.7	407.6	0.876
0.92–0.90	9242	92.9	649.6	393.3	1.028
0.90–0.88	9129	92.6	461.1	379.9	1.364
0.88–0.86	9184	92.1	311.6	367.2	1.828
0.86–0.85	8210	83.0	187.3	399.3	2.585
Overall	189859	95.0	110148.4	1984.5	0.045

2. Experimental

2.1. Crystallization

Cocrystals of the streptavidin–*epi*-biotin complex were obtained using vapor-diffusion methods. Solutions of streptavidin (10 mg ml^{-1} in 10 m*M* Tris–HCl pH 7.0) and ligand (20-fold excess) were combined and incubated for 5 d at 277 K. Sitting drops were equilibrated against a reservoir solution containing 38% saturated ammonium sulfate, 0.1 *M* sodium acetate pH 4.5, 0.2 *M* sodium chloride.

2.2. Data collection and processing

Two diffraction data sets were used in the analysis of these crystals. The first (data set I) was collected on an in-house R-AXIS II system. The second (data set II) was collected at SSRL beamline 11-1 with a MAR Research image-plate scanner. Two sets of frames made up these data sets. A low-resolution quick scan yielded reflections to 2.0 Å resolution. The highest resolution reflections were obtained from a second slower scan where a large aluminium beamstop was used to avoid overflows on the image plate.

For both the R-AXIS II and SSRL data collections, the crystals were mounted in 35% glycerol in a loop and flash-frozen at 100 K in a nitrogen stream. Each data set was processed with *DENZO* and merged and scaled with *SCALEPACK* (Otwinowski & Minor, 1997). Data-set statistics are shown in Tables 1 and 2. The R-AXIS setup for data set I limited its effective resolution. The highest shells contained reflections collected in the corners of the detector and thus are less complete than the lower resolution shells. If the highest resolution obtained in the corner of the detector is 1.5 Å, the resolution at the edge of the detector is about 1.7 Å. This is the resolution where the completeness of the shells drops and this accounts for the overall completeness of 80% for the data set.

The effective resolution of data set II is also less than the maximum limit reported owing to the large number of weak reflections in the outer resolution shells. This contributed to the high R_{merge} values in the highest resolution shells and their lower $I/\sigma(I)$. Nevertheless, there were measurable intensities at high resolution, so all of the reflections were used in the refinement.

The unit-cell parameters obtained for data set II are a = 46.39, b = 93.81, c = 104.45 Å. The symmetry of the diffraction pattern indicates that the crystals are orthorhombic and the systematic absences are consistent with space groups $I222$ or $I2_12_12_1$. V_M (Matthews, 1968) is 1.88 Å^3 Da^{-1} for two streptavidin subunits in the asymmetric unit. This corresponds to 34.5% solvent content in the crystal. One of the molecular twofold axes in the streptavidin tetramer (point group $D2$) is aligned with one of the crystallographic twofold rotation axes.

2.3. Structure solution and refinement

The crystals of the complex are isomorphous with those of wild-type streptavidin bound to biotin (PDB code 1mk5), space group $I222$. A model of streptavidin based on the 1mk5 structure, but without biotin, solvent molecules or alternate amino-acid side-chain conformations, was used as the starting point for the refinement. The structure refinement was carried out using the program *SHELXL*97 (Sheldrick & Schneider, 1997), its auxiliary program *SHELXPRO* and the graphics program *XtalView* (McRee, 1999).

The initial refinement and model building made use of the 1.5 Å resolution data set (data set I), but after data set II became available, refinement continued with only that data set at two resolution limits (see Table 3). The initial model contained no ligand or solvent molecules. Ligand and solvent molecules were added to the model based on their fit to α_A-weighted (Read, 1986) $|F_o|$ and $\big||F_o| - |F_c|\big|$ electron-density maps. The structural model, including anisotropic atomic displacement parameters for the protein atoms and heteroatoms, was refined against F^2. All parameters were refined simultaneously and 5% of the data were used for the calculation of R_{free} (Brünger, 1993). All reflections were used in refinement. R values for reflections with $|F_o| > 4\sigma(|F_o|)$ were calculated for comparison purposes only. Distance, planarity and chiral volume restraints were applied, as were anti-bumping restraints. The target values for 1–2 and 1–3 distances

Table 3
Refinement statistics.

Resolution (Å)	10.0–1.00	10.0–0.85
No. of unique reflections	118725	180225
No. of reflections with $\|F_o\| > 4\sigma(\|F_o\|)$	99921	125109
No. of parameters	18050	18050
No. of restraints	22250	22250
R_{work} (all reflections)	0.141	0.155
R_{work} [$\|F_o\| > 4\sigma(\|F_o\|)$]	0.132	0.134
R_{free} (all reflections; 5% of total)	0.152	0.173
R_{free} [$\|F_o\| > 4\sigma(\|F_o\|)$]	0.143	0.151
No. of protein atoms	1854	1854
No. of solvent atoms	273	273
No. of heteroatoms	52	52
Mean B value (isotropic) (Å^2)		
Main-chain atoms	14.0	14.1
Side-chain atoms	15.9	16.0
Solvent and heteroatoms	33.5	32.9
R.m.s. deviations from ideal values		
Bond lengths (Å)	0.015	0.018
Ramachandran plot statistics (%)		
Residues in most favored region	89.7	90.7
Residues in additional region	10.3	9.3
Residues in generously allowed region	0.0	0.0

Table 4
Isotropic temperature factors for the biotin ligands before generation of the disordered model.

	B value (Å^2)	
Atom	Subunit A	Subunit B
S1	24.77	21.66
C2	15.06	17.35
C3	11.92	12.55
C4	12.15	11.51
C5	13.25	14.03
N1′	14.42	11.30
C2′	15.70	12.22
O2P	14.00	13.43
N3′	12.14	13.09
C6	14.10	16.04
C7	14.66	14.86
C8	18.95	18.30
C9	15.23	17.70
C10	17.36	21.61
O11	17.64	18.64
O12	18.07	18.20

were based on the Engh & Huber (1991) study. Similarity restraints were applied for isotropic and anisotropic displacement parameters throughout the refinement. H atoms were added to the model at calculated positions in the final refinement cycles. *XtalView* (McRee, 1999) and *RASTER3D* (Merritt & Bacon, 1997) were used to produce the figures. The anisotropic displacement parameters were monitored using the program *PARVATI* (Merritt, 1999*a*,*b*).

Application of the validation program *SFCHECK* (Vaguine *et al.*, 1999) indicates that the optical resolution limit for the higher resolution data set is 1.03 Å, so refinement was carried out at 1.0 Å resolution. This structural model has been deposited in the Protein Data Bank with PDB code 2gh7 after being validated using *PROCHECK* (Laskowski *et al.*, 1993).

To assess whether the reflections between 1.0 and 0.85 Å resolution introduce noise and error into the refinement, refinement was also carried out using data to 0.85 Å resolution. Refinement statistics for this are included in Table 3 and they indicate little difference between the resulting models. 37% of the reflections between 1.0 and 0.85 Å resolution had $I > 2\sigma(I)$. These 25 303 reflections indicate how significant data are still available in resolution shells where the overall data set statistics are not favorable. The model resulting from the 0.85 Å resolution refinement has been deposited in the

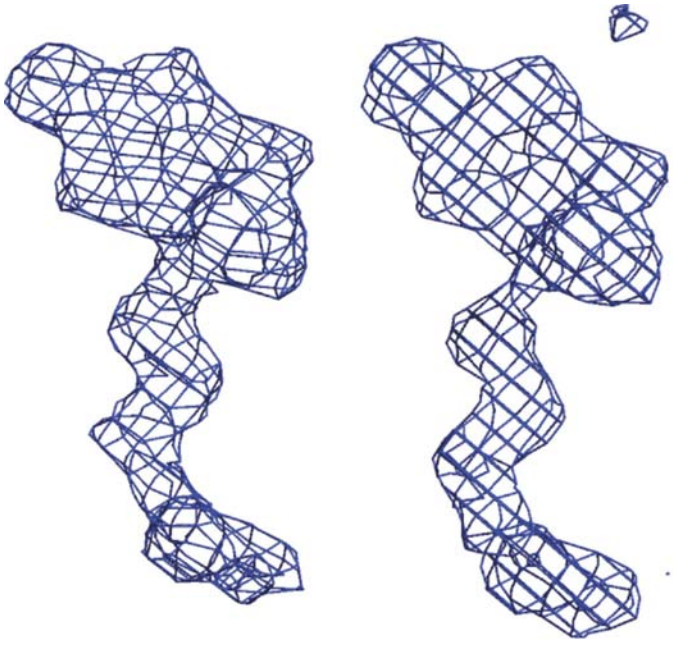

Figure 1
Difference electron density in the two biotin-binding sites in the asymmetric unit at 1.51 Å resolution contoured at the 2σ level. The map is a σ_A-weighted $\|F_o| - |F_c\|$ map calculated using model phases. No ligand atoms were present in the model for this calculation.

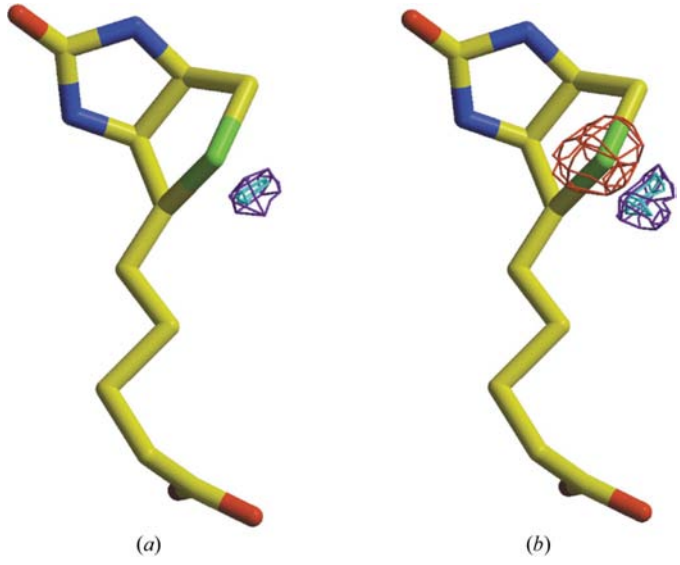

Figure 2
(*a*) Difference electron density for one binding site after refinement including the biotin molecule in the model. (*b*) Difference electron density after making the thermal parameter of the S atom consistent with those of the neighboring atoms. Negative densities are shown in red (-4σ) and positive in blue (4σ) and cyan (5σ).

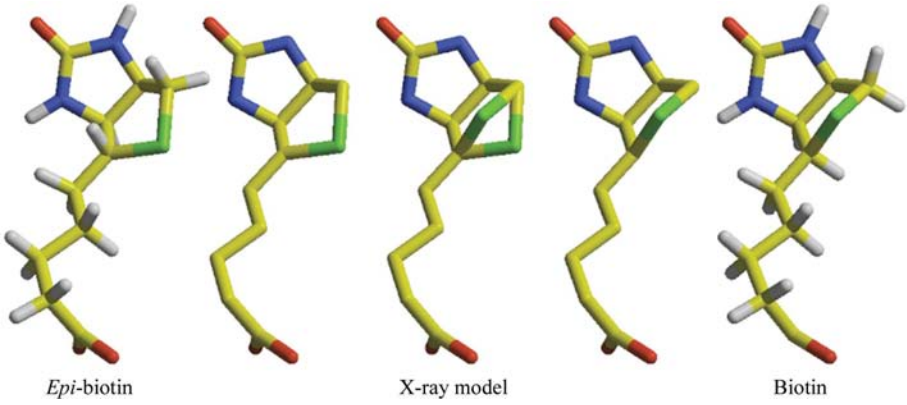

Figure 3
Summary figure showing the superposition of biotin in the *endo* conformation and *epi*-biotin in the *exo* conformation to form the observed structural model for the bound ligand.

Protein Data Bank with PDB code 2f01, again after being validated using *PROCHECK*.

3. Results and discussion

The difference electron density showing the biotin ligand in both subunits in the asymmetric unit is displayed in Fig. 1. The density can be fitted with the model for (+)-biotin found in the wild-type core streptavidin structure (1mk5), but after refinement of this initial model a small positive peak near the biotin S atom was apparent in the difference electron density (see Fig. 2*a*). Examination of the isotropic thermal parameters for the biotin ligand (see Table 4) showed that the sulfur thermal parameters were considerably higher than those of the other atoms in the ligand. The sulfur thermal parameters were set to the average of those of the atoms to which it was bonded and the resulting difference density is shown in Fig. 2(*b*). This density could be accounted for by two partially occupied positions for the sulfur. It is interesting to see that an increase in the *B* values for the S atoms of 5–10 Å^2 can account for an occupancy difference of 0.3–0.6. This adjustment in *B* values was sufficient to remove any negative density in the difference map shown in Fig. 2(*a*).

Various attempts to include additional components in the disorder model, especially in the valeric acid side chain, failed to refine well, probably owing to a lack of spatial resolution for the disordered atoms. The simplest, best behaved model was that shown in the center of Fig. 3. Refinement of the occupancies of the alternate locations for the S atoms resulted in occupancies of 0.62 and 0.38 in one subunit and 0.71 and 0.29 in the second. Fig. 4 shows the atom-naming scheme and anisotropic atomic displacement parameters for one of the biotin molecules in the asymmetric unit.

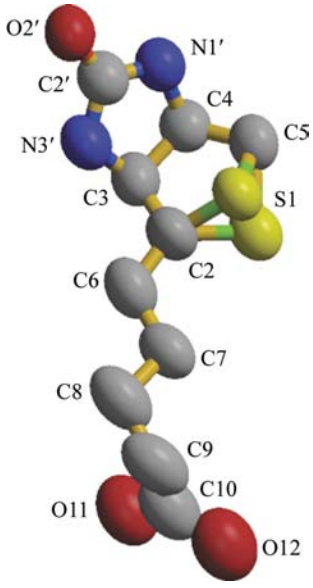

Figure 4
Anisotropic atomic displacement parameters (50% ellipsoids) and atom labels for one of the biotin models in the asymmetric unit.

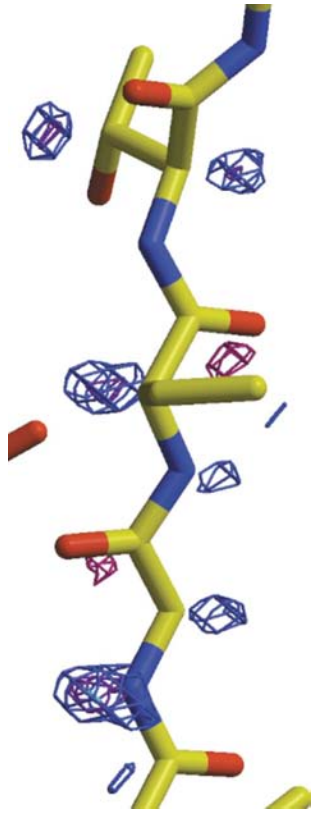

Figure 5
Difference electron-density map showing the positive peaks associated with H atoms on a β-strand at the center of the molecule. Positive contours are shown at 3σ (blue), 4σ (purple) and 5σ (cyan); negative contours are shown at -3σ (red). The $|F_o| - |F_c|$ map was calculated using phases and $|F_c|$ for the model just prior to the addition of H atoms at their calculated positions.

The alternate position for the sulfur, if the ligand were purely (+)-biotin, would require additional density for the alternate conformation of the valeric acid side chain. Conversion from the *endo* to the *exo* conformation of the ring would lead to a large change in the position of C6, the first atom in the side chain. Alternatively, the second sulfur position would be consistent with the presence of (+)-*epi*-biotin with an *exo* ring conformation. The model with two sulfur positions and a single side-chain conformation is best interpreted as being a superposition of (+)-*endo*-biotin and (+)-*exo*-*epi*-biotin as shown in Fig. 3. Sulfur position *A* (occupancy of 0.62) corresponds to (+)-*endo*-biotin, while position *B* (occupancy of 0.38) corresponds to (+)-*exo*-*epi*-biotin.

The streptavidin sample used in this study also provided crystals for another study in which no ligand was found in the binding site. Thus, the likely source of the (+)-biotin seen here was the (+)-*epi*-biotin sample (Aubert *et al.*, 2006).

The presence of (+)-*exo*-*epi*-biotin does not cause any significant change in the protein structure. The structure of this complex is superimposable with the isomorphous 1mk5 structure, the biotin complex of wild-type core streptavidin. The S atom from *epi*-biotin is close to two side chains (Thr90 and Leu110) which might move slightly to make room for the larger atom. However, these side chains show evidence of multiple conformations in both the (+)-biotin and (+)-*epi*-biotin complexes, so it is not clear whether the differences between the two structures are associated with the identity of the ligand or simply reflect differences in the model building and refinement.

There are no changes in the position of the carboxylate of the valeric acid tail of biotin, but this is expected since the switch of the two ring conformations places the C6 atom in the same position for both biotin molecules comprising the disordered model. However, the region near the biotin molecule is relatively open and if *epi*-biotin took on the *endo* conformation there is sufficient flexibility in the valeric acid tail to permit the carboxylate group to continue to participate in the hydrogen bonds found in the (+)-biotin complex.

The high-resolution diffraction data available for this complex permit a rather clear description of the disordered biotin molecules bound to the protein (see Fig. 3). In addition, near the end of the refinement, H atoms were clearly apparent for a well ordered part of the protein (see Fig. 5). Addition of H atoms at calculated positions improved the accuracy of the model and removed the peaks from the difference maps. A further gauge of the accuracy of the data-collection protocols can be obtained from the difference electron-density map calculated after six cycles of refinement of the model with the biotin ligands removed. In addition to resolved atoms in the ligand, the peak heights for the difference density are consistent with the differences in scattering power for the C, N, O and partially occupied S atoms. The peaks in the first biotin molecule for atoms in the ureido moiety are 12.5, 18.7, 18.0, 22.5 and 33.6σ for atoms C2′, N1′, N3′, O2′ and S1 (0.6 occupancy), respectively. For the second biotin, the corresponding peak heights are 16.9, 19.8, 21.6, 24.5 and 37.7σ.

Further studies of the conformational properties of (+)-*epi*-biotin will investigate whether the molecule prefers the *exo* or *endo* conformation in solution. It appears that either conformation would fit in the protein's binding site for biotin, so further work will be needed to determine whether the protein selects molecules with the *exo* conformation or whether that is the conformation preferred by most of the (+)-*epi*-biotin molecules.

Our expectation was that the difference in configuration at C2 for (+)-biotin and (+)-*epi*-biotin would yield a different structure for the (+)-*epi*-biotin complex. It is interesting that the *endo*/*exo* change in the ring conformation allows (+)-*epi*-biotin to overlap so closely with (+)-biotin in streptavidin's binding site. The high resolution of this structure analysis provides a clear picture of the two molecules bound to the protein.

This work was supported by NIH grant DK-49655. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences. We thank Ethan Merritt for assistance with data collection. We are also grateful for funding from The University of Nottingham for a PhD studentship to DGLA and and the Royal Society for a University Research Scholarship to NRT.

References

Aubert, D. G. L., Wilson, C., Blake, A. J. & Thomas, N. R. (2006). Submitted.
Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Chilkoti, A., Tan, P. H. & Stayton, P. S. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 1754–1758.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Freitag, S., Le Trong, I., Klumb, L. A., Chu, V., Chilkoti, A., Stayton, P. S. & Stenkamp, R. E. (1999). *Biomol. Eng.* **16**, 13–19.
Green, N. M. (1963). *Biochem. J.* **89**, 599–609.
Hirsch, J. D., Eslamizar, L., Filanoski, B. J., Malekzadeh, N., Haugland, R. P., Beechem, J. M. & Haugland, R. P. (2002). *Anal. Biochem.* **308**, 343–357.
Hyre, D. E., Le Trong, I., Freitag, S., Stenkamp, R. E. & Stayton, P. S. (2000). *Protein Sci.* **9**, 878–885.
Katz, B. A. (1997). *J. Mol. Biol.* **274**, 776–800.
Katz, B. A. & Cass, R. T. (1997). *J. Biol. Chem.* **272**, 13220–13228.
Korndoerfer, I. P. & Skerra, A. (2002). *Protein Sci.* **11**, 883–893.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. (1999*a*). *Acta Cryst.* D**55**, 1997–2004.
Merritt, E. A. (1999*b*). *Acta Cryst.* D**55**, 1109–1117.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pazy, Y., Kulik, T., Bayer, E. A., Wilchek, M. & Livnah, O. (2002). *J. Biol. Chem.* **277**, 30892–30900.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Reznik, G. O., Vajda, S., Sano, T. & Cantor, C. R. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 13525–13530.
Roberts, J. L., Moretti, P. A. B., Darrow, A. L., Derian, C. K., Vadas, M. A. & Pitson, S. M. (2004). *Anal. Biochem.* **331**, 122–129.
Schechter, B., Arnon, R., Freedman, Y. E., Chen, L. M. & Wilcheck, M. (1996). *J. Drug Test.* **4**, 171–179.
Schiestel, T., Brunner, H. & Tovar, G. E. M. (2004). *J. Nanosci. Nanotechnol.* **4**, 504–511.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Weber, P. C., Pantoliano, M. W., Simons, D. M. & Salemme, F. R. (1994). *J. Am. Chem. Soc.* **116**, 2717–2724.