Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of the SBP-Tag–streptavidin complex reveals a novel helical scaffold bridging binding pockets on separate subunits

Isabelle H. Barrette-Ng, Sau-Ching Wu, Wai-Mui Tjia, Sui-Lam Wong and Kenneth K. S. Ng*

Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada

Correspondence e-mail: ngk@ucalgary.ca

Received 23 November 2012
Accepted 25 January 2013

PDB Reference: streptavidin complex with SBP-Tag, 4jo6

The 38-residue SBP-Tag binds to streptavidin more tightly ($K_d \simeq 2.5$–4.9 nM) than most if not all other known peptide sequences. Crystallographic analysis at 1.75 Å resolution shows that the SBP-Tag binds to streptavidin in an unprecedented manner by simultaneously interacting with biotin-binding pockets from two separate subunits. An N-terminal HVV peptide sequence (residues 12–14) and a C-terminal HPQ sequence (residues 31–33) form the bulk of the direct interactions between the SBP-Tag and the two biotin-binding pockets. Surprisingly, most of the peptide spanning these two sites (residues 17–28) adopts a regular α-helical structure that projects three leucine side chains into a groove formed at the interface between two streptavidin protomers. The crystal structure shows that residues 1–10 and 35–38 of the original SBP-Tag identified through *in vitro* selection and deletion analysis do not appear to contact streptavidin and thus may not be important for binding. A 25-residue peptide comprising residues 11–34 (SBP-Tag2) was synthesized and shown using surface plasmon resonance to bind streptavidin with very similar affinity and kinetics when compared with the SBP-Tag. The SBP-Tag2 was also added to the C-terminus of β-lactamase and was shown to be just as effective as the full-length SBP-Tag in affinity purification. These results validate the molecular structure of the SBP-Tag–streptavidin complex and establish a minimal bivalent streptavidin-binding tag from which further rational design and optimization can proceed.

1. Introduction

The tetrameric bacterial protein streptavidin (SAV) is one of the most commonly used protein reagents in biotechnology because of its intrinsically high level of binding specificity and affinity (Green, 1990; Wilchek & Bayer, 1990; Laitinen *et al.*, 2007; O'Sullivan *et al.*, 2012). In addition to the cognate ligand biotin, SAV binds strongly to a number of peptide ligands. Peptide ligands are particularly useful in biotechnology because short 'tag' sequences can be engineered into recombinant proteins to provide a means of specifically binding SAV. Several peptide sequences that can serve as SAV-binding tags have been identified, primarily through library-screening (Devlin *et al.*, 1990; Lam *et al.*, 1991) and *in vitro* selection (Wilson *et al.*, 2001) approaches. The molecular structural basis underlying the recognition of peptide sequences by SAV has been established through crystallographic studies of complexes formed between SAV and first-generation peptides bearing the HPQ motif (Weber *et al.*, 1992; Katz, 1995). The second-generation peptides named *Strep*-Tag (K_d = 37 µM; Schmidt *et al.*, 1996), *Strep*-Tag II (K_d = 14 µM; Voss & Skerra, 1997; Korndörfer & Skerra, 2002) and NanoTag (K_d = 4 nM;

Lamla & Erdmann, 2004; Perbandt *et al.*, 2007) were subsequently developed and shown to have higher binding affinities to SAV. However, the *Strep*-Tag peptide only binds well if the terminal carboxylate group is exposed, thus limiting the use of these tags at the C-termini of engineered proteins. The NanoTag peptide only interacts well if the N-terminus of the peptide is exposed and if the N-terminal Met is formylated, again limiting the possible applications of this tag. *Strep*-Tag II does not need to be located at the end of a chain, but its modest binding affinity limits its usefulness in applications requiring tighter binding (*e.g.* the immobilization of proteins and peptides to biosensors).

Crystallographic studies have revealed how short peptide sequences mimic aspects of the natural ligand biotin to form interactions with the biotin-binding pocket of SAV (Schmidt *et al.*, 1996; Voss & Skerra, 1997; Korndörfer & Skerra, 2002; Perbandt *et al.*, 2007). Although some of these peptides have successfully been used in biotechnological applications, the best binding affinities for SAV are still modest in comparison with the cognate ligand biotin (K_d = 50 f*M*; Green, 1990; Hyre *et al.*, 2006). Because the rates of association and dissociation of the best peptide-tag sequences currently available are often not ideal for applications such as protein purification, the design or identification of new peptide sequences with higher binding affinities, higher rates of association and lower rates of dissociation is highly desirable.

The origin of the SBP-Tag derives from an mRNA display/*in vitro* selection procedure (Wilson *et al.*, 2001) that was used to identify 88-mer peptides with higher binding affinities for SAV. One of the tightest-binding peptides originally identified through this procedure (SB19; K_d = 10 n*M*) was used as a starting point to identify a much shorter 38-residue deletion mutant (SB19-C4) with a slightly higher binding affinity (K_d = 2.5–4.9 n*M*; Keefe *et al.*, 2001; Wilson *et al.*, 2001). This deletion mutant was subsequently renamed the SBP-Tag and shown to have sufficient binding affinity to SAV for use as an affinity tag in the purification of recombinant proteins (Keefe *et al.*, 2001). Since it was first described, many applications of the SBP-Tag have been reported, including use in the tandem affinity purification (TAP) of protein complexes (Bürckstümmer *et al.*, 2006; Li *et al.*, 2011). To better understand the molecular basis of recognition of the SBP-Tag by SAV, as well as to provide a molecular structural basis for the design of affinity-tag sequences with superior binding properties, we have crystallized the SBP-Tag with SAV and determined the crystal structure of the complex.

2. Materials and methods

2.1. Crystallization and structure determination

Streptavidin (Roche Biosciences) was dissolved in water [1%(*w*/*v*)] and mixed with SBP-Tag [EZBiolab Inc., Indiana, USA; 1%(*w*/*v*)] at a ratio of 1 mol tetramer:2 mol SBP-Tag. Sparse-matrix crystallization screens identified conditions for the growth of prismatic crystals, which were optimized to form crystals large enough for data measurement. A single crystal (0.2 × 0.1 × 0.05 mm) was grown by hanging-drop vapour diffusion by mixing 2 µl of the streptavidin/SBP-Tag mixture with 2 µl precipitant solution [56% Tacsimate (Hampton Research) pH 7.0, 12%(*w*/*v*) glycerol] and equilibrating the drop against 1 ml precipitant solution. Diffraction data were measured using this crystal on beamline 9-2 at the Stanford Synchrotron Radiation Laboratory and the *HKL* suite was used for indexing, integration and scaling (Otwinowski & Minor, 1997). Molecular-replacement calculations were carried out using *Phaser* with PDB entry 1swe (Freitag *et al.*, 1997) as the search model (Read, 2001). *REFMAC* and *Coot* were used for refinement and model building (Murshudov *et al.*, 2011; Emsley & Cowtan, 2004). *MolProbity* was used to evaluate the geometric quality of the model during refinement (Chen *et al.*, 2010). *Surface Racer* was used to calculate accessible surface area (Tsodikov *et al.*, 2002). Crystallographic statistics are given in Table 1.

Table 1
Crystallographic statistics.

Values in parentheses are for the outermost resolution shell.

Data collection	
Space group	$P4_1$
Unit-cell parameters (Å)	$a = b = 57.52$, $c = 177.55$
Resolution (Å)	60–1.75 (1.80–1.75)
R_{merge}†	0.066 (0.424)
$\langle I/\sigma(I)\rangle$	29.2 (2.5)
Completeness (%)	95.3 (72.6)
Multiplicity	7.6 (4.6)
Refinement	
Resolution (Å)	60–1.75
Unique reflections	52299
R_{work}‡/R_{free}§	0.206/0.242
No. of atoms	4250
Protein	3589
SBP-Tag	388
Water	273
R.m.s. deviations	
Bond lengths¶ (Å)	0.008
Bond angles¶ (°)	1.19
Bond-related *B* factors, main chain†† (Å^2)	1.6
Bond-related *B* factors, side chain†† (Å^2)	3.4
Ramachandran angles‡‡ (%)	
Favoured	97.7
Outliers	0

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th integrated intensity of a given reflection and $\langle I(hkl)\rangle$ is the weighted mean of all measurements of $I(hkl)$. ‡ $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ for the 95% of the reflection data used in refinement. § $R_{free} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ for the 5% of the reflection data excluded from refinement. ¶ Root-mean-square deviations from ideal geometry calculated by *REFMAC* (Murshudov *et al.*, 2011). †† Root-mean-square deviations for refined temperature factors of bonded atoms. ‡‡ Ramachandran plot analysis carried out using *MolProbity* (Chen *et al.*, 2010).

2.2. Biosensor study of the streptavidin–SBP-Tag interaction

Kinetic parameters (k_{on} and k_{off}) of the streptavidin–SBP-Tag interaction were determined using a surface plasmon resonance-based BIAcoreX biosensor (BIAcore Inc.). Two synthetic peptides, FL and S (ordered from Peptide 2.0 Inc., Virginia, USA), were used in this study. FL is a 59-amino-acid peptide (**CGGGGSTSGGSTSGGSTSGGG**MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP) with the full-length 38-residue SBP-Tag sequence located at the C-terminal end of the peptide. The first 21 amino acids

(highlighted in bold) function as a spacer to project the SBP-Tag to the chip surface. In contrast, S is a 45-amino-acid peptide (**CGGGGSTSGGSTSGGSTSGGG**GHVVEGLAGELEQLRARLEHHPQG) carrying the shortened version (residues 11–34) of the SBP-Tag at the C-terminal end. It also has a 21-residue spacer at the N-terminus. The presence of the N-terminal cysteine residues in these peptides allows the immobilization of either FL or S to the second flow cell (FC-2) of a CM5 sensor chip (GE Healthcare Biosciences) *via* the thiol-coupling method according to the manufacturer's instructions. A blank flow cell (FC-1) in each chip was used as a reference cell. Approximately 70 response units of either peptide were immobilized on the chip surface. Different concentrations (30–300 n*M*) of streptavidin (from *Streptomyces avidinii*; Sigma) were passed through the flow cells at a flow rate of 30 µl min^{-1}. The sensor chip was regenerated by a 1.5 min pulse of 3 m*M* D-biotin followed by a 1 min pulse of 10 m*M* HCl, both at a flow rate of 30 µl min^{-1}.

Parameters for binding kinetics were evaluated by fitting the SPR data to two distinct models using the *BIAevaluation* software v.3.2. The 'AB model' (where A is tetrameric SAV, B is SBP-Tag; A + B → AB) assumes that a single SBP-Tag binds to a single dimeric binding site in a single streptavidin tetramer. The AB_2 model (A+ B → AB; AB + B → AB_2) assumes that two SBP-Tags bind to two dimeric binding sites in a single streptavidin tetramer. The residual plots and χ^2 values obtained by fitting the data to these two models show that the AB model describes the data much better than the AB_2 model. As a result, kinetic parameters were estimated using the AB binding model. This analysis does not exclude the possibility that some binding occurs through the AB_2 model under the reported experimental conditions, but the much better fit for the AB model indicates that binding according to the AB_2 model is minor, if present at all.

2.3. Construction of β-lactamase fusions with either the full-length or the short version of the SBP-Tag

Two synthetic genes encoding the full-length and short versions of the SBP-Tag were ordered from Epoch Life Science Inc., Texas, USA. A *Bst*B1 site at the 5′ end and an *Nhe*I site downstream from the stop codon of the SBP-Tag were introduced into these sequences. The sequences were inserted into *Bst*BI- and *Nhe*I-cut pWB980-L55-LytE (Chen *et al.*, 2008) to generate pWB980-BLA-L-FLSBP with the full-length SBP-Tag and pWB980-BLA-L-SSBP with the short version of the SBP-Tag. BLA-L19-FLSBP has a 19-amino-acid linker (IDPAGTSPSTPEGPSTPSN) located upstream of the SBP-Tag. BLA-L21-SSBP has the same 19-amino-acid linker plus two extra glycine residues located upstream of the short version of the SBP-Tag.

2.4. Production of β-lactamase fusions from *Bacillus subtilis* and purification of the fusion proteins using streptavidin-agarose

β-Lactamase fusions were produced from *B. subtilis* WB800[pWB980-BLA-L-FLSBP] (where FLSBP denotes the full-length SBP tag) and WB800[pWB980-BLA-L-SSBP] (where SSBP denotes the short version) cultured at 310 K for 7–8 h in a super-rich medium (Wu *et al.*, 1993) in a shake flask. The culture supernatant containing the secreted proteins was concentrated using ultrafiltration and dialyzed in phosphate-buffered saline (PBS; 0.1 *M* sodium phosphate, 0.15 *M* sodium chloride pH 7.5). The crude sample was loaded onto a column

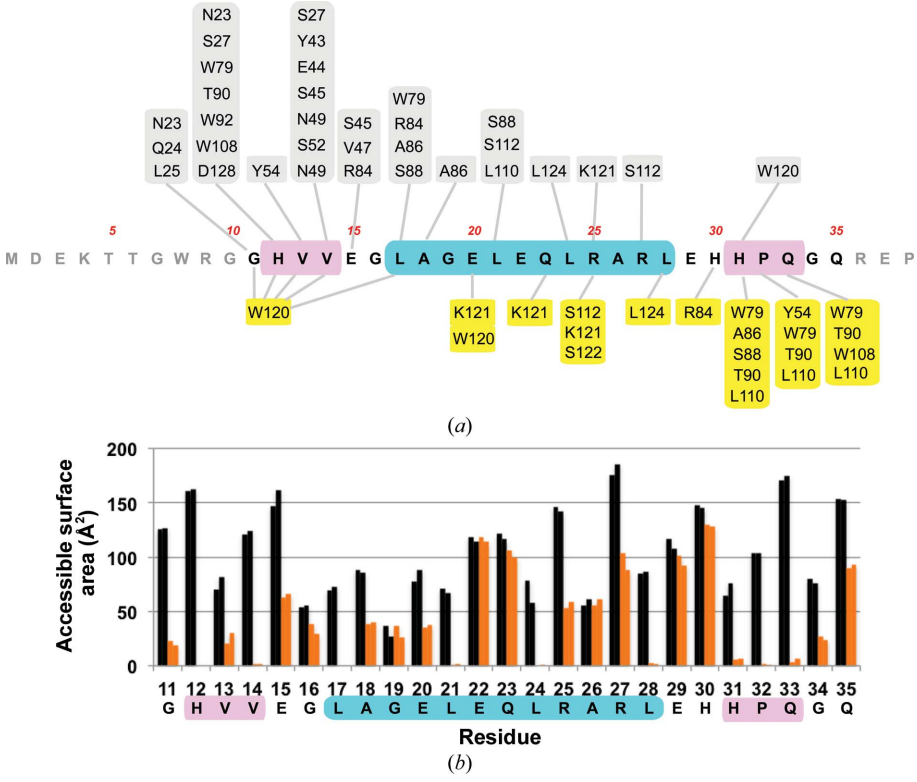

Figure 1
(*a*) Sequence of the SBP-Tag and schematic diagram of its interactions with SAV. Residues highlighted in magenta indicate the HVV and HPQ motifs interacting with residues in the biotin-binding pocket. Residues highlighted in cyan adopt an α-helical conformation. Residues highlighted in grey belong to subunit *A* (or *B*) and residues highlighted in yellow belong to subunit *C* (or *D*) of SAV. Hydrogen-bond and van der Waals interactions were detected using the programs *HBPLUS* (McDonald & Thornton, 1994) and *CONTACT* in the *CCP*4 suite (Winn *et al.*, 2011) and were verified by manual inspection of distances and geometry. (*b*) Total accessible surface area (ASA) for each residue clearly defined by electron density. Black bars indicate the ASA for the two copies of the SBP-Tag in the asymmetric unit (chains *Y* and *Z*) when the SBP-Tag is not bound to SAV. Orange bars indicate the ASA for chains *Y* and *Z* when the SBP-Tag is bound to SAV. If a small difference is found between ASA values of the bound and unbound SBP-Tag, then there is little interaction of the residue with SAV. If there is a large difference, this indicates a more significant interaction between the SBP-Tag and SAV.

containing 300 µl streptavidin-agarose (Sigma). After washing the column with PBS, the bound proteins were eluted with PBS containing 5 m*M* D-biotin.

3. Results

3.1. Overall structure of the SAV–SBP-Tag complex

The structure of the complex of the SBP-Tag with SAV reveals two molecules of SBP-Tag peptide bound to a single SAV tetramer. Each of the two SBP-Tag molecules contains 25 well ordered residues and 13 disordered residues: ten at the N-terminus and three at the C-terminus (Fig. 1). A total of 1494 Å^2 or 56% of the accessible surface area of each SBP-Tag molecule is buried upon binding to SAV. Most surprisingly, most of the SBP-Tag adopts an α-helical conformation, with short tripeptide sequences preceding the N-terminal end of the helix by two residues and following the C-terminal end by two residues reaching into the biotin-binding pockets of SAV protomers belonging to separate dimers of the streptavidin tetramer (Fig. 2). Crystallographic demonstration of a bivalent ligand binding to separate subunits of the SAV tetramer is unprecedented and helps to explain why the SBP-Tag has a higher binding affinity than univalent peptide ligands such as the *Strep*-Tag and has comparable binding affinity to peptide ligands such as the Nano-Tag.

The formation of the tetrameric structure of SAV can be thought of as the assembly of four subunits into two very stable dimers that in turn associate with each other through a less well formed packing interface (Fig. 2). Each of the two dimers in the SAV tetramer thus each present a single biotin-binding pocket on each of two opposing faces of the SAV tetramer. As a result, the SAV tetramer as a whole presents two binding pockets on each of two faces. Each molecule of SBP-Tag in turn binds simultaneously to the two binding pockets presented on a single face of the tetramer.

It is notable that by binding across the twofold rotational symmetry axis relating the two dimers in SAV, a single SBP-Tag molecule breaks the twofold rotational symmetry relating the two dimers to each other. This mode of binding is reminiscent of how linear peptide substrates and inhibitors bind to homodimeric viral aspartic proteinases (Wlodawer & Gustchina, 2000). In retroviral proteinases, many asymmetric inhibitors and substrate analogues have been crystallized with a homodimeric enzyme in which each subunit is related to the other by twofold rotational symmetry. In some cases, the peptide binds in a unique orientation in the crystal, whereas in others the orientation of the bound peptide appears to be averaged over two possibilities, thus complicating the analysis. Each of the two SBP-Tag molecules bound to the SAV tetramer appears to bind predominantly in a single unique orientation, most likely because in this crystal form one of the key crystal-packing interfaces involves contacts between SBP-Tag molecules (Fig. 2; Supplementary Movie 1[1]). This packing interface presumably selects a consistent and unique orientation of the SBP-Tag peptide from the two different possible orientations found in solution for a molecular complex composed of two SBP-Tag molecules and a single SAV tetramer.

One of the most surprising aspects of the SBP-Tag is the α-helical framework that is used to present the binding epitopes to the two ligand-binding pockets of the SAV protomers. Although α-helices are one of the most commonly found secondary-structural elements in proteins, isolated α-helices in ligands are much less common. This may be because the stability of an isolated helix is generally much lower than the stability of helices forming tertiary contacts in natural proteins (Scholtz & Baldwin, 1992; Baldwin & Rose, 1999). One of the few examples of a ligand forming a single helix is the IA3 peptide, which binds to and specifically inhibits proteinase A from baker's yeast (Li *et al.*, 2000). The structure of IA3 bound to its target enzyme reveals a 30-residue α-helix that makes a modest number of contacts with the broad substrate-binding cleft of the yeast aspartic proteinase. In this case, evidence from circular-dichroism and nuclear magnetic resonance spectroscopy indicates that IA3 adopts very little ordered secondary structure when free in solution and only adopts an α-helical structure when bound to proteinase A (Green *et al.*, 2004; Ganesh *et al.*, 2006).

The helical portion of the SBP-Tag may behave in a manner similar to IA3. When only analyzing the amino-acid sequence, the program *AGADIR* predicts ~20% helical content for the central portion of the SBP-Tag peptide that is ordered in the crystal structure (Muñoz & Serrano, 1994). *AGADIR* also predicts the highest levels of residue-level helicity for most of the residues in the helix observed in the SAV–SBP-Tag complex (residues 20–28). The crystal structure of the SAV–SBP-Tag complex also reveals that most of the helical portion of the SBP-Tag does not form extensive contacts with SAV. Instead, this region mostly serves as a scaffold or linker between the short peptide sequences that reach into the biotin-binding pockets. Because most of the helical region does not appear to form extensive interactions with the surface residues of SAV, the crystal structure suggests opportunities for improving the binding affinity of the SBP-Tag by making modifications to the sequence that introduce novel interactions with SAV.

3.2. Recognition of SBP-Tag sequences by SAV

The bulk of the interactions between SAV and the SBP-Tag are focused on the HVV and HPQ tripeptide sequences lying two residues before and two residues after the central α-helix formed by the SBP-Tag (Figs. 1 and 3). The HPQ sequence binds in a manner that is similar to that observed previously in other peptide tags bound to SAV (Weber *et al.*, 1992; Katz, 1995; Schmidt *et al.*, 1996). The tripeptide adopts a 3_{10}-helix secondary structure and mimics the structure of biotin by occupying the biotin-binding pocket in a manner that forms both hydrogen-bonding and hydrophobic interactions with amino-acid residues contributed mostly by one subunit (chains

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: MV5083). Services for accessing this material are described at the back of the journal.

A or *B*) as well as the predominantly hydrophobic side chain of Trp120 from an adjacent subunit (chains *C* or *D*). The HVV tripeptide binds to the biotin-binding pocket in a novel manner in which the side chain of His is buried most deeply into the pocket, whereas the two valine residues interact with the hydrophobic side chains presented by Trp120, Leu25, Tyr43, Trp79 and Tyr54.

Apart from the central tripeptide HVV and HPQ sequences in the SBP-Tag, most of the residues flanking these sequences form a few additional interactions that are likely to enhance binding affinity and specificity. Only a single glycine residue N-terminal to HVV is visible in the electron-density maps, which suggests that residues 1–10 of the original SBP-Tag are likely to extend away from the SAV tetramer and are likely to adopt a primarily disordered structure in solution. Gly11 does lie in a constrained exit channel leading away from the biotin-binding pocket, but does not form many specific interactions with SAV. Glu15 and Gly16, the residues immediately

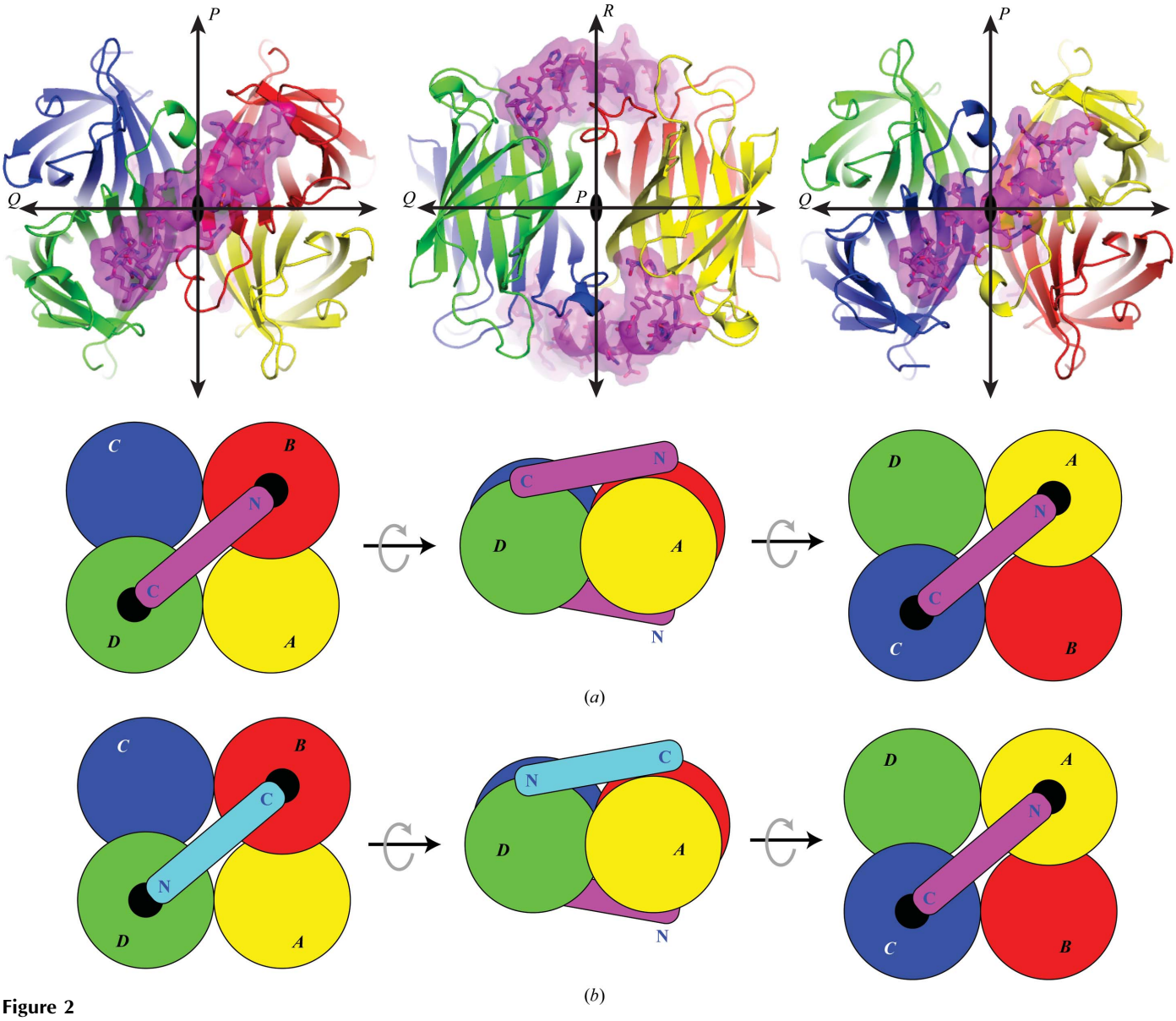

Figure 2
(*a*) Ribbon-diagram representations of views of the SAV tetramer bound to two molecules of the SBP-Tag (ribbon diagrams and stick representations of side chains). The views in the left, centre and right panels are related by 90° rotations about a horizontal axis as shown. Each subunit is labelled *A* (yellow), *B* (red), *C* (blue) or *D* (green). The N-terminus and C-terminus of each of the two molecules of the SBP-Tag (purple) is also shown in blue lettering. The approximate location of the biotin-binding pockets are shown in the schematic diagrams as black circles. The three perpendicular twofold rotation axes for the SAV tetramer are labelled according to Hendrickson *et al.* (1989). The binding of two molecules of the SBP-Tag to the SAV tetramer results in the loss of two of these symmetry elements: only the dyad labelled *Q* is obeyed in the SBP-Tag–SAV complex. (*b*) A second type of SBP-Tag–SAV complex expected to form in solution but presumably selected against during crystallization owing to crystal-packing interactions. The SBP-Tag molecule shown in cyan binds to the tetramer in an orientation opposite to that observed in the crystallized complex, even though the local interactions between this molecule and the two streptavidin subunits are identical to those observed in the crystallized complex. These interactions are also identical to those formed by the other SBP-Tag molecule shown in purple. Note that in this alternate complex only the twofold rotational symmetry defined by axis *P* is obeyed.

following the HVV sequence, lie on the surface of subunit *A*, forming both hydrophobic and hydrogen-bonding interactions with the side chains of residues Ser45, Val47 and Arg84. Leu17 is the first residue of the central helix in the SBP-Tag and is buried in a hydrophobic pocket formed in part by Trp79, Arg84, Ala86 and Ser88, as well as the HVV motif of the SBP-Tag itself. The remaining residues of the helical linker of the SBP-Tag are mostly exposed to solvent, although the side chains of Leu21, Leu24 and Leu28 face SAV and are nearly completely buried; they are thus likely to provide some binding energy through a classical hydrophobic effect as well as van der Waals interactions.

Leu28, which is the last residue in the α-helix of the SBP-Tag, faces towards subunit *C* and is mostly buried in a pocket formed by the surface-exposed Leu110 and Leu124 side chains from subunit *C* as well as Lys121 and Trp120 from subunit *A* and the Arg27 and His31 (the first residue in the HPQ motif) side chains of the SBP-Tag itself. His31, Pro32 and Gln33 adopt a stapled structure very similar to that observed in other HPQ-containing peptides crystallized with streptavidin (Weber *et al.*, 1992; Katz, 1995; Schmidt *et al.*, 1996). Immediately following the HPQ motif, Gly34 forms additional interactions with SAV by packing against Trp120, whereas Gln35 is mostly solvent-exposed and extends away from the surface of SAV.

3.3. Design and testing of the minimal SBP-Tag2

To further test the prediction from the crystal structure of a minimal length for full binding activity in the SBP-Tag, we synthesized a shortened version lacking residues 1–10 and 35–38, which we call the SBP-Tag2 (24 of the 38 residues of the

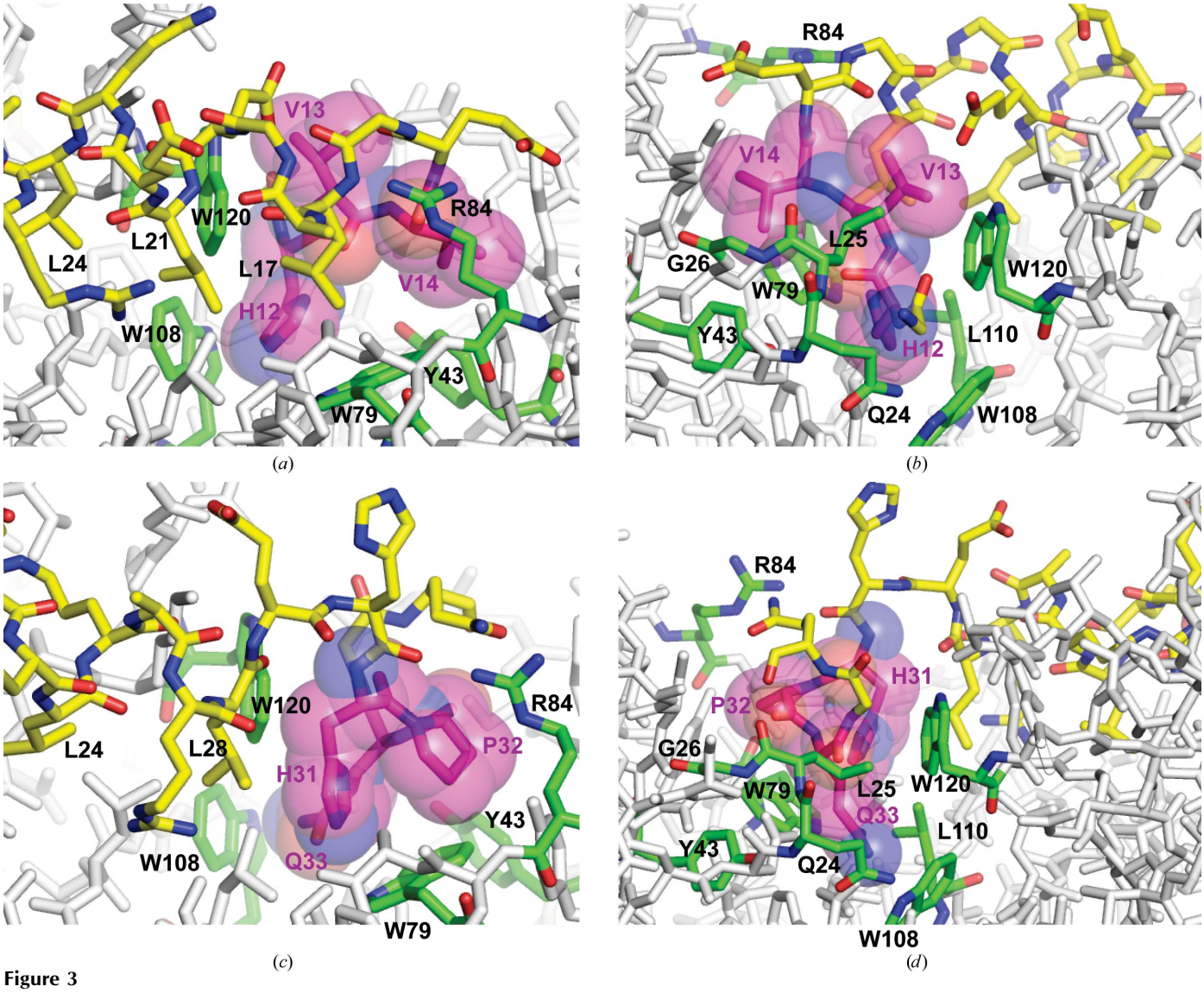

Figure 3
Detailed view of the interactions between (*a*, *b*) the N-terminal end and (*c*, *d*) the C-terminal end of the SBP-Tag and SAV. The views shown in (*a*) and (*c*) are from the 'front' and the views in (*b*) and (*d*) are drawn from the 'back' side of the SBP-Tag. The C atoms in the HVV and HPQ motifs are coloured magenta and shown in semi-transparent space-filling representation. C atoms in other parts of the SBP-Tag are coloured yellow and drawn in stick representation. C atoms of key residues of SAV interacting with the SBP-Tag are coloured green. Other residues of SAV are coloured light grey. N and O atoms are coloured blue and red, respectively.

Table 2
Binding constants for interactions between streptavidin and SBP-Tags.

The *BIAevaluation* software was used to evaluate the fit of the association model (A + B = AB) and dissociation model (AB = A + B) to the experimental data. The goodness of fit to these models was assessed by inspecting the residual plot (which shows the difference between experimental and calculated data), the lag plot and the χ^2 value. The residual (±0.6 RU randomly scattered around the *x* axis) and χ^2 values are well within the noise limit. The points in the lag plot are randomly scattered around the origin. These tests indicate that the model provides a good representation of the data.

	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_d (M)
SBP-Tag	3.17×10^5 $\pm 7.98 \times 10^3$	4.85×10^{-4} $\pm 3.61 \times 10^{-6}$	1.53×10^{-9} $\pm 4.52 \times 10^{-10}$
SBP-Tag2	3.38×10^5 $\pm 8.51 \times 10^3$	4.96×10^{-4} $\pm 3.52 \times 10^{-6}$	1.47×10^{-9} $\pm 4.14 \times 10^{-10}$

SBP-Tag, or 63% of the total length). We added a short linker sequence at the N-terminus for immobilization on thiol-activated carboxymethyl dextran and to increase accessibility by projecting the SBP-Tag2 away from the matrix. The immobilized tag was then used to measure the binding of SBP-Tag2 *versus* SBP-Tag to SAV using surface plasmon resonance. The binding kinetics and binding affinities of both peptides are very similar (K_d = 1.5 nM), thus confirming a key prediction of the structure (Fig. 4 and Table 2).

As a further practical demonstration of the utility of the SBP-Tag2, β-lactamase fusion proteins containing either the SBP-Tag or the SBP-Tag2 at the C-terminus were expressed and purified using streptavidin-agarose (Fig. 5). Both fusion proteins were retained selectively on the affinity matrix and were readily eluted by biotin. Notably, neither fusion protein was detected in the wash fraction. Also, after elution with biotin neither fusion protein was retained on the streptavidin-agarose beads (Fig. 5, lane Bo). The recovery of both fusion proteins was estimated to be at least 85%. This practical application of the SBP-Tag2 for protein purification confirms that the minimal version of the SBP-Tag can functionally replace the original full-length SBP-Tag in binding and purification applications. In addition, this experiment demonstrates that both the short and full-length versions of the SBP-Tag are compatible with the secretion pathway and can be applied to purify not only intracellular proteins but also extracellular proteins.

4. Discussion

4.1. Structure-based analysis of SBP-Tag variants

The surprising structure of the complex of the SBP-Tag with SAV provides a three-dimensional framework for the design of novel peptide ligands with improved binding properties for SAV. Firstly, the structure indicates that residues 1–10 and 36–38 (13 out of 38 residues or 34%) of the original SBP-Tag optimized by deletion analysis do not appear to contact SAV. This observation provides novel insight into the the binding properties of a series of SB19 variants containing specific deletions and site-directed replacements (Wilson *et al.*, 2001). The binding behaviour of the SB19 M1, N1 and N2 variants is particularly informative. Variant M1 replaced PQ from the HPQ motif with GA and resulted in 0.065% binding activity compared with SB19 in a column-binding assay. Our structure indicates the key role of the HPQ motif and this result indicates that the remaining binding interactions are not sufficient to compensate for the loss of this primary site of interaction. Mutant N1 deletes most of the residues in the N-terminal region (residues 3–10), which are also observed to be disordered and do not interact with SAV in our crystal structure,

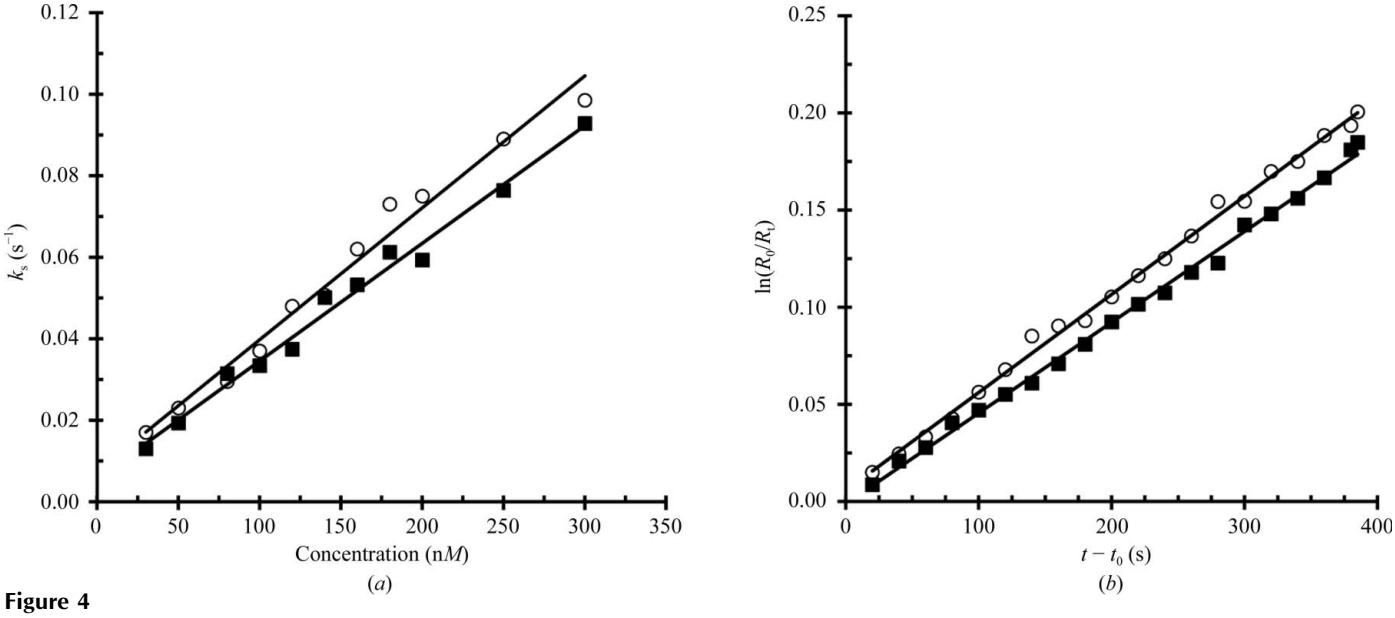

Figure 4
Kinetic analysis of the interaction of streptavidin with SBP-Tags using a BIAcore biosensor. (*a*) Determination of the association rate constant (k_{on}); k_{on} is obtained from the slope of the plot. (*b*) Determination of the dissociation rate constant (k_{off}). k_{off} is obtained from the slope of the plot. k_s is the observed rate constant (Qureshi *et al.*, 2001). R_0 is the response at arbitrary starting time t_0; R_t is the response at time *t*. Black squares are for the SBP-Tag and white circles are for the SBP-Tag2.

and this leads to minimal effects on binding (69% of that of SB19) as expected. The slight reduction in binding may be attributed to steric interference or perhaps to electrostatic repulsion of the Asp residue preceding Gly10 in the N1 mutant, whereas SB19 has the smaller and neutral Gly residue preceding Gly10.

Mutant N2 shows that the deletion of four additional residues (11–14) of mutant N1, including the HVV motif, further reduces binding to 30% of that of SB19 but does not abolish binding. This suggests that the HVV motif contributes to the binding of SB19 but is not as important as the HPQ motif and the remaining interactions. A second explanation that is more consistent with the behaviour of the N3 mutant described below is that the MD residues at the N-terminus of the N2 mutant are able to replace some of the interactions contributed by the HVV motif. The structure of the SAV–SBP-Tag complex supports both of these possibilities. The HVV motif does not seem to fill out the biotin-binding pocket as well as either biotin or the HPQ motif, and a simplistic model suggests that a combination of the hydrophobic and hydrogen-bonding interactions present in the MD sequence replacing HVV in the N2 mutant could contribute sufficient binding interactions to partially compensate for the loss of the HVV motif (Supplementary Fig. S1). In fact, Val13 in the SBP-Tag structure is mostly solvent-exposed and is not likely to contribute as much to binding as His12 and Val14. Further structural analysis of the N2 mutant may help to elucidate whether the MD motif can interact with the biotin-binding pocket in place of HVV as suggested by modelling. The last mutant, N3, in which residues 2–19 are lost, shows very little column-binding activity (0.058% of that of SB19). This observation indicates that the HPQ motif and three-quarters of the helical linker do not contribute sufficient binding affinity to be detected using the column-binding assay, which indicates the importance of the second binding site contributed by HVV and the N-terminal portion of the linker. The interpretation of the binding in this mutant is consistent with the interpretation of binding in the N2 mutant. N2 retains significant binding because we propose that MD can partially replace the role of the HVV motif, but the deletion of the entire HVV motif and N-terminal portion of the linker helix is something that cannot be tolerated.

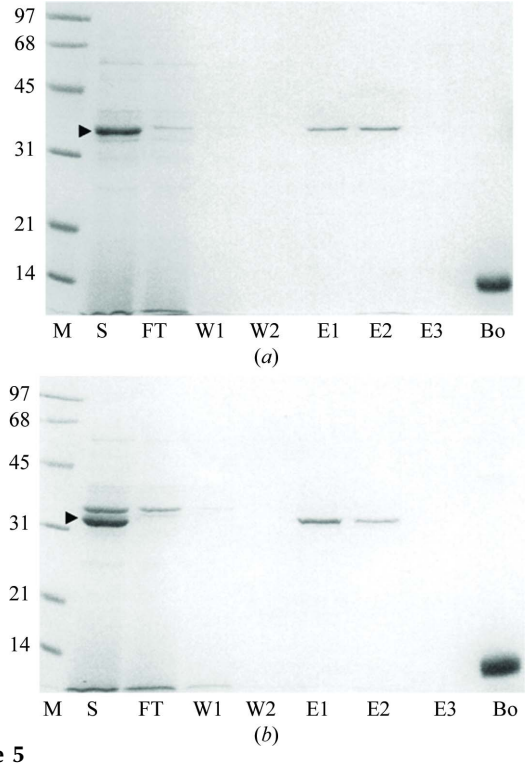

Figure 5
SDS–PAGE showing streptavidin-agarose purification of secreted β-lactamase fusion proteins containing either the SBP-Tag (*a*) or the SBP-Tag2 (*b*) at the C-terminus. Culture supernatant containing β-lactamase fusions was applied onto a streptavidin-agarose column. 10 µl of sample from each fraction (300 µl for each fraction volume) was loaded onto the gel. S, crude sample; FT, flowthrough fraction; W, wash fraction; E, elution fraction; Bo, boiled streptavidin beads after elution of the column-bound β-lactamase fusions. The 13 kDa protein observed in lane Bo is the streptavidin subunit. The positions of the β-lactamase fusions are marked by arrowheads.

Now that we have established a structural framework for the binding of the SBP-Tag to streptavidin, future refinement and possible improvements in binding kinetics and affinity can be more rationally designed. Saturation mutagenesis of the HVV motif may be expected to yield further improvements in binding affinity because this motif does not appear to form optimal interactions with the biotin-binding pocket at present. Mutagenesis targeting specific positions in the helical linker region may also be envisioned to improve interactions between the tag and SAV. Notably, positions extending away from SAV may now also be selectively altered to possibly improve solubility or other properties without affecting the interactions with SAV. The crystal structure also reveals that both the N-terminus and the C-terminus of the SBP-Tag2 are expected to extend freely into solution and are not recognized by SAV. As a result, unlike the *Strep*-Tag and Nano-Tag, the SBP-Tag2 can be inserted at either the N- or C-termini of target proteins. With appropriate spacer peptides at each end of the SBP-Tag2, this tag could also be inserted into the middle of a long loop or another exposed region of a protein.

This research was supported by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to SLW and KN. KN is also supported by the Alberta Glycomics Centre and a Senior Scholar Award from Alberta Innovates Health Solutions. We gratefully acknowledge access to crystal screening and data-collection facilities at the Stanford Synchrotron Radiation Laboratory (SSRL) and the Canadian Light Source (CLS) at various stages during the course of this project. SSRL is a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the US Department of Energy Office of Science by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393) and the National Center for Research Resources (P41RR001209). The CLS is supported by the Natural Sciences and Engineering

Research Council of Canada, the National Research Council Canada, the Canadian Institutes of Health Research, the Province of Saskatchewan, Western Economic Diversification Canada and the University of Saskatchewan.

References

Baldwin, R. L. & Rose, G. D. (1999). *Trends Biochem. Sci.* **24**, 26–33.

Bürckstümmer, T., Bennett, K. L., Preradovic, A., Schütze, G., Hantschel, O., Superti-Furga, G. & Bauch, A. (2006). *Nature Methods*, **3**, 1013–1019.

Chen, C.-L., Wu, S.-C., Tjia, W. M., Wang, C. L. C., Lohka, M. J. & Wong, S.-L. (2008). *Microb. Biotechnol.* **1**, 177–190.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Devlin, J. J., Panganiban, L. C. & Devlin, P. E. (1990). *Science*, **249**, 404–406.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Freitag, S., Le Trong, I., Klumb, L., Stayton, P. S. & Stenkamp, R. E. (1997). *Protein Sci.* **6**, 1157–1166.

Ganesh, O. K., Green, T. B., Edison, A. S. & Hagen, S. J. (2006). *Biochemistry*, **45**, 13585–13596.

Green, N. M. (1990). *Methods Enzymol.* **184**, 51–67.

Green, T. B., Ganesh, O., Perry, K., Smith, L., Phylip, L. H., Logan, T. M., Hagen, S. J., Dunn, B. M. & Edison, A. S. (2004). *Biochemistry*, **43**, 4071–4081.

Hendrickson, W. A., Pahler, A., Smith, J. L., Satow, Y., Merritt, E. A. & Phizackerley, R. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 2190–2194.

Hyre, D. E., Le Trong, I., Merritt, E. A., Eccleston, J. F., Green, N. M., Stenkamp, R. E. & Stayton, P. S. (2006). *Protein Sci.* **15**, 459–467.

Katz, B. A. (1995). *Biochemistry*, **34**, 15421–15429.

Keefe, A. D., Wilson, D. S., Seelig, B. & Szostak, J. W. (2001). *Protein Expr. Purif.* **23**, 440–446.

Korndörfer, I. P. & Skerra, A. (2002). *Protein Sci.* **11**, 883–893.

Laitinen, O. H., Nordlund, H. R., Hytönen, V. P. & Kulomaa, M. S. (2007). *Trends Biotechnol.* **25**, 269–277.

Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M. & Knapp, R. J. (1991). *Nature (London)*, **354**, 82–84.

Lamla, T. & Erdmann, V. A. (2004). *Protein Expr. Purif.* **33**, 39–47.

Li, Y., Franklin, S., Zhang, M. J. & Vondriska, T. M. (2011). *Protein Sci.* **20**, 140–149.

Li, M., Phylip, L. H., Lees, W. E., Winther, J. R., Dunn, B. M., Wlodawer, A., Kay, J. & Gustchina, A. (2000). *Nature Struct. Biol.* **7**, 113–117.

McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.

Muñoz, V. & Serrano, L. (1994). *Nature Struct. Biol.* **1**, 399–409.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

O'Sullivan, V. J., Barrette-Ng, I., Hommema, E., Hermanson, G. T., Schofield, M., Wu, S. C., Honetschlaeger, C., Ng, K. K.-S. & Wong, S.-L. (2012). *PLoS One*, **7**, e35203.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perbandt, M., Bruns, O., Vallazza, M., Lamla, T., Betzel, C. & Erdmann, V. A. (2007). *Proteins*, **67**, 1147–1153.

Qureshi, M. H., Yeung, J. C., Wu, S.-C. & Wong, S.-L. (2001). *J. Biol. Chem.* **276**, 46422–46428.

Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.

Schmidt, T. G., Koepke, J., Frank, R. & Skerra, A. (1996). *J. Mol. Biol.* **255**, 753–766.

Scholtz, J. M. & Baldwin, R. L. (1992). *Annu. Rev. Biophys. Biomol. Struct.* **21**, 95–118.

Tsodikov, O. V., Record, M. T. & Sergeev, Y. V. (2002). *J. Comput. Chem.* **23**, 600–609.

Voss, S. & Skerra, A. (1997). *Protein Eng.* **10**, 975–982.

Weber, P. C., Pantoliano, M. W. & Thompson, L. D. (1992). *Biochemistry*, **31**, 9350–9354.

Wilchek, M. & Bayer, E. A. (1990). *Methods Enzymol.* **184**, 5–13.

Wilson, D. S., Keefe, A. D. & Szostak, J. W. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 3750–3755.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Wlodawer, A. & Gustchina, A. (2000). *Biochim. Biophys. Acta*, **1477**, 16–34.

Wu, X.-C., Ng, S.-C., Near, R. I. & Wong, S.-L. (1993). *Biotechnology*, **11**, 71–76.