Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Atomic resolution structure of a lysine-specific endoproteinase from *Lysobacter enzymogenes* suggests a hydroxyl group bound to the oxyanion hole

Peter Asztalos,[a]*‡ Astrid Müller,[a] Werner Hölke,[a] Harald Sobek[a] and Markus G. Rudolph[b]*‡

[a]Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany, and [b]Molecular Design and Chemical Biology, F. Hoffmann-La Roche, Grenzacher Strasse 124, 4070 Basel, Switzerland

‡ Correspondence regarding enzymology should be directed to PA and correspondence regarding structural biology to MGR.

Correspondence e-mail: peter.asztalos@roche.com, markus.rudolph@roche.com

Received 7 January 2014
Accepted 14 April 2014

PDB references: LysC, wild type, 4nsy; K30R mutant, 4nsv

Lysobacter enzymogenes lysyl endoproteinase (LysC) is a trypsin-type serine protease with a high pH optimum that hydrolyses all Lys-Xaa peptide bonds. The high specificity of LysC renders it useful for biotechnological purposes. The K30R variant of a related lysyl endoproteinase from *Achromobacter lyticus* has favourable enzymatic properties that might be transferrable to LysC. To visualize structural differences in the substrate-binding sites, the crystal structures of wild-type and the K30R variant of LysC were determined. The mutation is located at a distance of 12 Å from the catalytic triad and subtly changes the surface properties of the substrate-binding site. The high pH optimum of LysC can be attributed to electrostatic effects of an aromatic Tyr/His stack on the catalytic aspartate and is a general feature of this enzyme subfamily. LysC crystals in complex with the covalent inhibitor N^{α}-*p*-tosyl-lysyl chloromethylketone yielded data to 1.1 and 0.9 Å resolution, resulting in unprecedented precision of the active and substrate-binding sites for this enzyme subfamily. Error estimates on bond lengths and difference electron density indicate that instead of the expected oxyanion a hydroxyl group binds to the partially solvent-exposed oxyanion hole. Protonation of the alkoxide catalytic intermediate might be a recurring feature during serine protease catalysis.

1. Introduction

Stable serine proteases are used industrially in diverse areas ranging from medical applications to food processing and laundry. For example, *Achromobacter lyticus* protease I has been used to convert pig insulin to human insulin by specific removal and replacement of amino acid 30 in the insulin B-chain (Morihara & Ueno, 1991) or as a step during the total synthesis of human insulin (Tofteng *et al.*, 2008). Biochemical applications such as primary-structure analysis, peptide mapping, in-gel digestion, cleavage of fusion proteins and protein synthesis require specific proteases that are resilient to denaturation even in the presence of significant concentrations of denaturants and that operate under a broad range of conditions (Kuhlman *et al.*, 2009). Lysine residues are frequently encountered on the surface of proteins and are suitable positions for fragmentation reactions. Endoproteinases that cleave C-terminally to lysine residues have been isolated from *A. lyticus* (Masaki *et al.*, 1981), *Pseudomonas aeruginosa* (Elliott & Cohen, 1986) and *Lysobacter enzymogenes* (Jekel *et al.*, 1983). All of these proteases are currently used in protein fragmentation and modification reactions, although some of their properties could be improved. To reduce autodegradation during prolonged incubation times and to increase the activity and stability of the protease, the

lysine residues at positions 30, 49, 106, 155 and 203 in *A. lyticus* protease I were mutated. Compared with the wild type, the K30R variant of *A. lyticus* protease I has a higher efficiency for Lys-Xaa peptide-bond cleavage in (for example) insulin, an increased performance when immobilized on a water-insoluble carrier and augmented protein stability owing to diminished autoproteolysis (Asser *et al.*, 1997). The crystal structure of *A. lyticus* protease I has been determined (Tsunasawa *et al.*, 1989), showing that the lysine-residue positions are mostly solvent-exposed and are thus amenable to autohydrolysis.

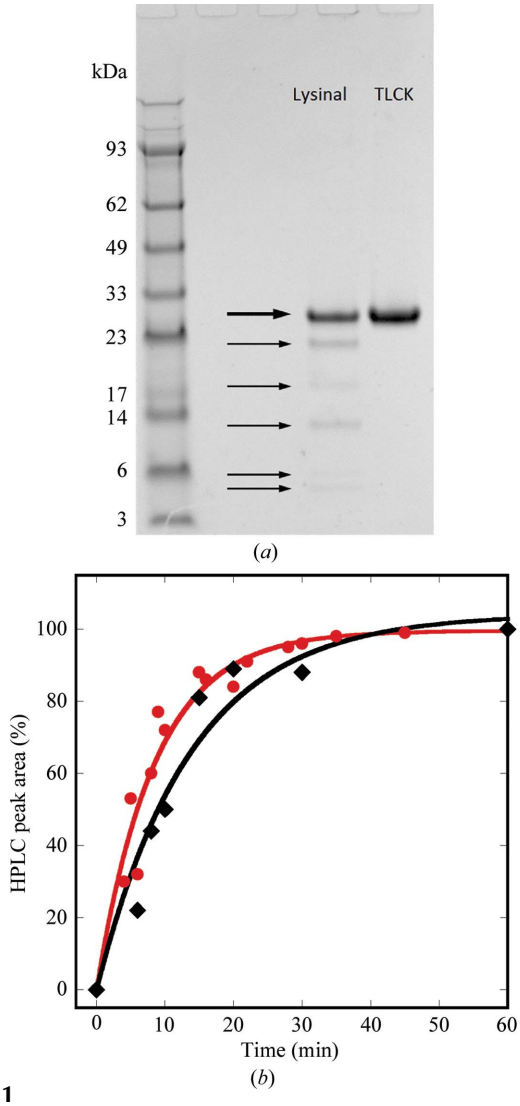

Figure 1
Proteolysis catalysed by LysC. (*a*) LysC autoproteolysis can be inhibited by TLCK but not by lysinal. 508 µ*M* wild-type LysC was incubated with a 1.2-fold stoichiometric excess of covalent inhibitor for 4 h at 20°C. Both inhibitors should bind covalently to LysC. Only TLCK inhibits LysC, whereas degradation products with apparent molecular weights of approximately 5, 6, 13, 18 and 22 kDa accumulate in the lysinal-treated sample. (*b*) Time course for the formation of peptide 1 from cleavage of melittin by wild-type LysC (black diamonds) and the K30R variant (red circles). Data were fitted using an equation describing a single exponential first-order reaction, resulting in rate constants k_{obs} for the wild type and the variant of 0.0727 and 0.1175 min^{-1}, respectively.

Another potent lysine-specific endoproteinase is lysyl endoproteinase from *L. enzymogenes*, also termed LysC, which has 78% identity (89% similarity) to *A. lyticus* protease I over the entire sequence. LysC catalyses the hydrolysis of peptide bonds at the carboxy-terminal side of lysine residues. The nature of the amino acid in the Lys-Xaa peptide is irrelevant and can include proline. LysC exhibits an unusual pH optimum around pH 9 and retains >10% activity at pH 6 and pH 12. The enzyme withstands 7 *M* urea without loss of activity. A drop in activity to 50% is observed in the presence of 0.1% sodium dodecyl sulfate (SDS), 0.3% SDS, 0.5 *M* NaCl or 0.5 *M* guanidinium chloride (Kuhlman *et al.*, 2009). Inhibition by high concentrations of cations was attributed to competition with the charged lysyl side chain of the substrates. Interestingly, 1.5-fold activation was observed with 0.01% SDS, and the presence of 20% acetonitrile leads to a 2.5-fold increased activity. While N^{α}-*p*-tosyl-lysyl chloromethylketone (TLCK) inhibits LysC, leupeptin (*N*-acetyl-leucyl-leucyl-argininal) does not (Kuhlman *et al.*, 2009), in line with the specificity of LysC for lysine but not arginine.

No crystal structure exists for *L. enzymogenes* LysC, and no structural information on any stabilized lysine variant is available for any related protease. To compare LysC with *A. lyticus* protease I and to obtain the first insights into the stabilizing properties of lysine variants, the crystal structures of wild-type and K30R LysC were determined. Crystals were obtained in complex with the covalent inhibitor TLCK, lending insight into substrate specificity for lysine. In contrast to cysteine proteases, serine proteases bind TLCK with two covalent bonds to their active site. The resulting adduct mimics the tetrahedral intermediate that occurs twice during peptide-bond hydrolysis. Although >2500 unique serine protease crystal structures are available in the PDB, there are only a few high-resolution structures of analogues of the tetrahedral intermediate that show clear density, such as a hemiketal in elastase (Tamada *et al.*, 2009), a phosphonate in D-Ala-D-Ala peptidase (Silvaggi *et al.*, 2003) and a boronate in α-lytic protease (Fuhrmann *et al.*, 2006). The atomic resolution data for the TLCK complexes allow the estimation of bond-length uncertainties and, together with the electron density that is visible for some of the inhibitor H atoms, conclusions on the protonation state of the tetrahedral intermediate in the oxyanion hole. A comparison with other high-resolution protease structures shows that a hydroxyl group or an alkoxide can alternatively bind to the oxyanion holes in proteases.

2. Materials and methods

2.1. Cloning, protein production and purification, and characterization

Wild-type and K30R pPICZaA-EndoLysC DNA constructs encoding a C-terminal His tag with a GSG linker were transformed into *Pichia pastoris* KM71H. Chromosomal integration was confirmed by colony PCR. Zeocin resistance was used as a selection marker. The integrated copy number of the pPICZaA-EndoLysC wild-type or K30R DNA constructs

was increased by repeated transformation, where G418 resistance was used as a selection marker in addition to zeocin. During fermentation, LysC was secreted into the extracellular medium. Biomass was removed by centrifugation and the filtered supernatant was applied onto a chelating column (Mitsubishi; Resindion Sepabeads FP-IDA 400) that was equilibrated with 0.2 *M* $ZnCl_2$ pH 5.5. Nonspecifically bound material was eluted with 50 m*M* Tris–HCl pH 7.5, 0.2 *M* NaCl followed with the same buffer including 5 m*M* imidazole. Proteins were eluted with a linear imidazole gradient over ten column volumes to 50 m*M* Tris–HCl pH 7.5, 0.2 *M* NaCl, 0.5 *M* imidazole. Active fractions (see below) were pooled and a purity of >95% was confirmed by SDS–PAGE (Fig. 1*a*). LysC was dialysed against 50 m*M* HEPES–NaOH pH 8.0, concentrated using a 5 kDa PES Biomax membrane (Millipore) and filtered using a 0.2 µm filter membrane (Sartorius). Monodispersity was tested by gel-permeation chromatography on a Superdex 75 column (GE Healthcare). The final enzyme preparations were stored at −20°C.

The proteolytic activity of wild-type and K30R LysC was measured in a continuous and time-dependent manner using Tos-Gly-Pro-Lys-4-nitroaniline (Chromozym PL, Roche Applied Science) as a chromogenic substrate, which has an extinction coefficient of 10 400 M^{-1} cm^{-1} at a wavelength of 405 nm after hydrolysis. Measurements were performed at 25°C in 25 m*M* Tris–HCl pH 7.7, 1 m*M* EDTA. The reaction was initiated by addition of enzyme (1.3–2.7 mU ml^{-1}) to the reaction mixture and the change in absorbance at 405 nm was followed. A discontinuous assay was used to monitor proteolysis of the model substrate melittin, where the products were quantified by reversed-phase HPLC. Melittin is a 26-mer peptide with the sequence GIGAVLK↓VLTTGLPALISWI-K↓RK↓RQQ that contains three cleavage sites for LysC (marked by arrows). 0.05 U ml^{-1} enzyme was added to 116 µ*M* melittin (Roche Applied Science) dissolved in 25 m*M* Tris–HCl pH 8.5, 1 m*M* EDTA and incubated at 37°C. Samples were taken at different time points and the reaction was stopped by transfer to 100°C for 5 min. The peptides resulting from the cleavage were separated by reversed-phase HPLC (Nucleosil C18) and the signals were integrated. The peak areas of the products were normalized to the corresponding peptide signal after complete cleavage, which was set to 100%. Peptide 1 (GIGAVLK) was used as marker for the cleavage reaction. A single exponential equation was fitted to the data, yielding k_{obs} for wild-type LysC and the K30R variant.

2.2. Crystallization and data collection

Wild-type and K30R LysC in 50 m*M* HEPES–NaOH pH 8.0 were concentrated to 14.6 and 9.8 mg ml^{-1}, respectively. TLCK was dissolved in water at 100 m*M* and mixed with LysC at a slightly overstoichiometric ratio of 1.2 with a negligible change in volume. Crystallization trials were performed at 20°C in a sitting-drop vapour-diffusion setup with a ratio of protein:precipitant of 1:1 and a drop volume of 0.5 µl. Crystals of the wild-type TLCK complex were obtained from an unbuffered solution of 0.2 *M* $CaCl_2$, 20% PEG 3350. The K30R LysC–TLCK complex was crystallized from 0.1 *M* Tris–HCl pH 8.5, 0.2 *M* Li_2SO_4, 25% PEG 3350. Prior to data collection, the crystals were directly flash-cooled from their mother liquor by hyperquenching (Warkentin & Thorne, 2007). Data were collected at 100 K on beamline PX-II at the Swiss Light Source using a PILATUS 6M detector and were integrated and scaled with *XDS* and *XSCALE*, respectively (Kabsch, 2010). The resolution cutoff was chosen at

Table 1
Data-collection and refinement statistics.

Unless noted otherwise, values in parentheses are for the highest resolution shell.

	Wild-type LysC (PDB code 4nsy)	K30R LysC (PDB code 4nsv)
Wavelength (Å)	1.0	0.7
Resolution range (Å)	37.6–1.10 (1.14–1.10)	42.8–0.90 (0.93–0.90)
100% criterion† (Å)	1.16	0.93
Rotation range/increment (°)	180/0.25	180/0.25
Mosaicity (°)	0.34	0.30
Space group	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a = 39.6, b = 135.8, c = 45.6, β = 97.2	a = 39.6, b = 135.6, c = 45.6, β = 98.0
Unique reflections	182416 (16073)	339560 (32998)
Multiplicity	3.3 (2.6)	3.4 (3.3)
Completeness (%)	94.8 (83.4)	97.1 (94.6)
R_{meas}‡	0.119 (1.16)	0.083 (1.30)
$R_{p.i.m.}$‡	0.066 (0.68)	0.044 (0.69)
$CC_{1/2}$‡	1.00 (0.40)	1.00 (0.36)
Average $I/\sigma(I)$	7.0 (1.0)	9.7 (1.1)
Wilson B (Å^2)	7.9	5.8
$\langle\lvert E^2-1\rvert\rangle$‡§	0.741 (0.736/0.541)	0.690
Mean $\langle L^2\rangle$‡§	0.314 (0.333/0.200)	0.291
Resolution range (Å)	34.0–1.10 (1.11–1.10)	39.5–0.90 (0.91–0.90)
No. of reflections	182401 (4780)	339441 (10255)
R_{cryst}¶ (%)	16.4 (33.1)	13.9 (36.5)
R_{free}¶ (%)	20.8 (36.6)	15.8 (38.6)
No. of residues	530	528
No. of waters	965	937
No. of TLCK molecules	2	2
No. of ions††	4	4
Coordinate error‡‡ (Å)	0.15	0.11
Phase errors‡‡ (°)	23.2	16.9
R.m.s.d., bond lengths (Å)	0.007	0.009
R.m.s.d., angles (°)	1.20	1.41
Ramachandran plot§§ (%)		
Core region	96.5	96.0
Allowed region	3.5	4.0
Disallowed region	—	—
MolProbity score¶¶	1.40	0.97
Clashscore¶¶	2.74	0.63
$\langle B\rangle$ (Å^2)		
Protein	10 ± 5	7 ± 3
Water	27 ± 11	21 ± 9
TLCK	9 ± 2	7 ± 2
Ions††	19 ± 8	40 ± 8

† The 100% criterion was calculated using *SFTOOLS* (Winn *et al.*, 2011) and represents the resolution in Å of a 100% complete hypothetical data set with the same number of reflections as the measured data. ‡ *E*-values, *L*-values (acentric reflections) and *R* factors were calculated using *PHENIX* (Zwart *et al.*, 2008). *R* values and $CC_{1/2}$ are defined in Diederichs & Karplus (1997) and Karplus & Diederichs (2012), respectively. § Values in parentheses are the expected values for untwinned data and for perfectly twinned data, respectively. ¶ $R_{cryst} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where F_{obs} and F_{calc} are the structure-factor amplitudes from the data and the model, respectively. R_{free} is R_{cryst} calculated using a 5% test set of structure factors. †† Ions are 2 × Cl^- and 2 × Ca^{2+} for the wild type and 3 × SO_4^{2-} plus Cl^- for the K30R mutant. ‡‡ Calculated using *PHENIX* (Zwart *et al.*, 2008). §§ Calculated using *Coot* (Emsley *et al.*, 2010). ¶¶ The *MolProbity* score should approach the high-resolution limit (Chen *et al.*, 2010). Clashscore is defined as the number of unfavourable all-atom steric overlaps of ≥0.4 Å per 1000 atoms (Word *et al.*, 1999).

$CC_{1/2}$ > 0.3 in the high-resolution shell, which incidentally corresponds to $I/\sigma(I) \simeq 1$ (Table 1). According to a heuristic definition (Sheldrick, 1990; Dauter *et al.*, 1995), atomic resolution data should extend to at least 1.2 Å and at least 50% of data in the outer resolution shell should have $I/\sigma(I) > 2$, which is true for the K30R complex [$I/\sigma(I)$ = 8 in the 1.18–1.14 Å resolution shell, 97% completeness] but is not quite true for the wild-type LysC data [$I/\sigma(I)$ = 1.7 in the 1.20–1.17 Å resolution shell, 90% completeness]. Indexing and integration of the isomorphous data sets was possible in a primitive monoclinic setting. Systematically weak reflections along b^* ($k \neq 2n$) from data processed in space group $P2$ indicated a screw axis along this direction; hence, space group $P2_1$ was assigned. Assuming two LysC molecules in the asymmetric unit, the Matthews coefficient (Matthews, 1968) is 2.1 Å^3 Da^{-1} with a solvent content of 41%.

2.3. Phasing and refinement

Molecular replacement using *Phaser* (McCoy *et al.*, 2007) was performed using *A. lyticus* protease I (Tsunasawa *et al.*, 1989) as the search model, which displays 78% sequence identity (89% similarity) to *L. enzymogenes* LysC over 263 residues. The model was rebuilt in *Coot* (Emsley *et al.*, 2010) and refined with *PHENIX* (Zwart *et al.*, 2008) employing anisotropic *B* factors for non-H atoms and optimizing the weights for geometry and *B*-factor restraints. The binding mode of the ligand was clear from the difference electron density of >10 r.m.s.d. per atom (Fig. 6*b*). The covalently bound inhibitor structure was derived from unreacted TLCK and was energy-minimized in *MOLOC* (Gerber & Müller, 1995). Restraints for these coordinates were generated with *PHENIX* and were applied throughout refinement. Error estimates for bond lengths and angles of the K30R structure were derived by full-matrix inversion refinement against all diffraction data (no test set) using *SHELXL* (Sheldrick, 2008). The model from *PHENIX* was first minimized to convergence against the least-squares target, followed by a single round of matrix inversion. The crystallographic *R* value for the resulting model is 12.1% for $F > 4\sigma$ and 14.9% for all data. While these values are similar to the refinement statistics reported in Table 1, they do not coincide owing to the different bulk-solvent protocols that are implemented in *PHENIX* and *SHELXL*.

3. Results and discussion

3.1. Purification and analysis of recombinant *L. enzymogenes* LysC

Wild-type and the K30R variant of *L. enzymogenes* LysC were produced in *P. pastoris* and purified *via* a single chelating

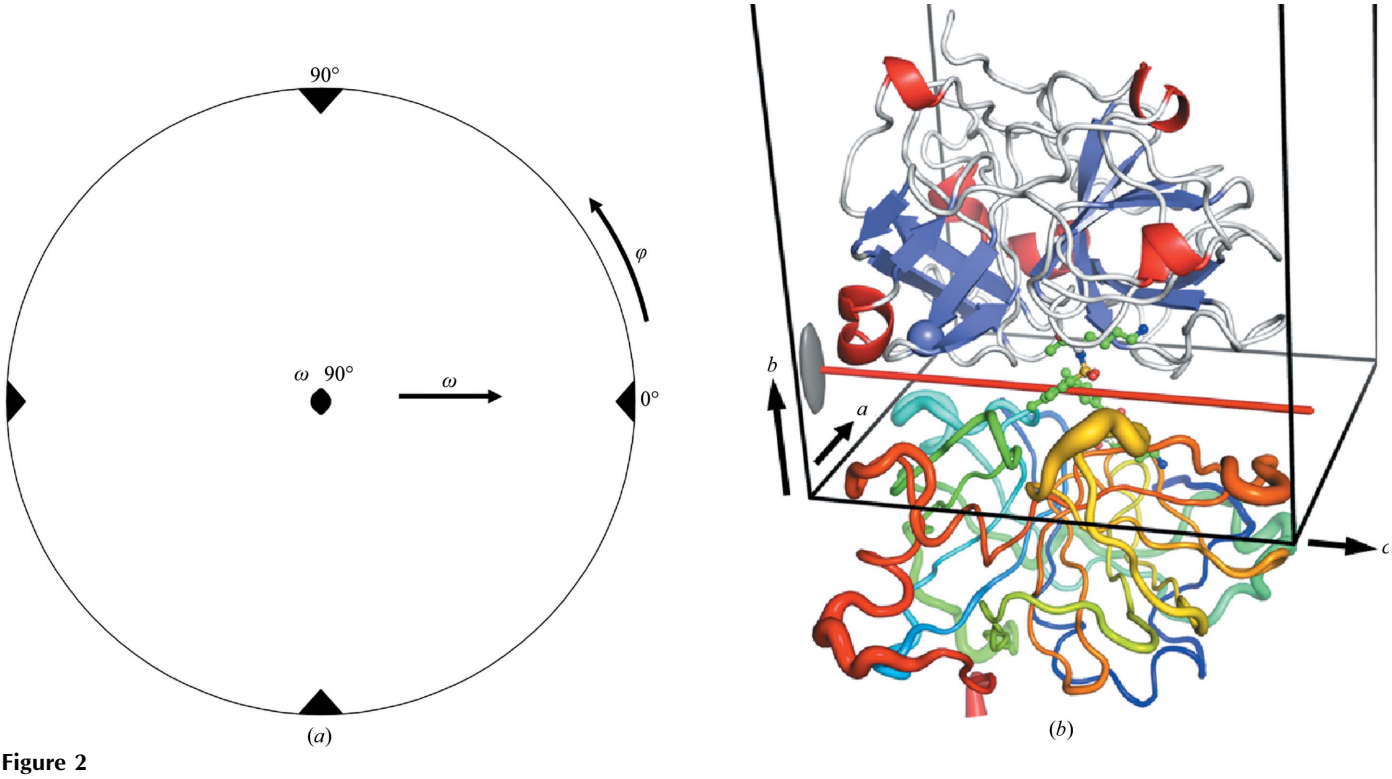

Figure 2
Self-rotation function and asymmetric unit of *L. enzymogenes* LysC. (*a*) The $\kappa = 180^\circ$ section of the self-rotation function calculated in space group $P2$ to 3 Å resolution and contoured at 30% of the origin peak height shows orthorhombic pseudo-symmetry. (*b*) The monoclinic asymmetric unit of LysC. The top protomer is coloured according to secondary structure and shows the independent β-barrels (blue) of the trypsin fold. The bottom protomer is rainbow-coloured from the N-terminus (blue) to the C-terminus (red) and the thickness of its C^α trace varies linearly according to the relative *B*-factor distribution. The two protomers are related by a twofold NCS axis parallel to the *c* axis (red line), *i.e.* perpendicular to the *b* axis, that generates a third twofold axis along the *a* direction. As the NCS and *b* axes do not intersect, an orthorhombic asymmetric unit with one molecule is not feasible. This is also reflected by the β angle of 97° that prohibits an orthogonal coordinate system.

chromatographic step. The functionality of both enzymes was established by a chromogenic activity assay (see §2). Crystallization of LysC in the apo form was unsuccessful; hence, inhibitor complexes were tried. Leupeptin (*N*-acetyl-leucyl-leucyl-argininal) is an actinomycete-derived tripeptide derivative that covalently inhibits several cysteine and serine proteases including papain, cathepsin B, trypsin, plasmin and kallikrein by the formation of a hemi-(thio)acetal. However, other serine proteases such as thrombin, α-chymotrypsin and LysC (Kuhlman *et al.*, 2009) are insensitive to leupeptin. By analogy to leupeptin, it was hoped that lysinal (*N*-acetyl-leucyl-valyl-lysinal; *i.e.* tripeptide LVK with a C-terminal aldehyde) would achieve inhibition of LysC. Lysinal is related to leupeptin but carries a lysyl moiety that should fit into the substrate-binding site of LysC. However, lysinal also has no inhibitory effect towards LysC, as determined by the self-degradation products of LysC after incubation with lysinal (Fig. 1*a*). This result contrasts with the observation that similar tripeptide-lysinals were found to strongly inhibit *A. lyticus* protease I with nanomolar K_i values (Masaki *et al.*, 1992). The apparent molecular weights of the LysC autodegradation products are approximately 5, 6, 13, 18 and 22 kDa, which could be explained by self-cleavage at positions Lys48 or Lys49 (5 and 22 kDa), Lys203 (6 kDa), Lys30 and Lys155 (13 kDa), and Lys30 and Lys 203 (18 kDa). No further analysis of the degradation products was performed. Four of the five lysine positions in LysC are conserved in the sequence of *A. lyticus* protease I and all five are located on the surface of the protein, thus enabling the autodegradation of these proteases. Autoproteolysis of LysC might also explain the inability to obtain co-crystals with inhibitory peptides or even with the known covalent inhibitor phenylmethyl sulfonyl-fluoride (Kuhlman *et al.*, 2009). TLCK, however, is a potent inhibitor of LysC as determined by the lack of any self-degradation (Fig. 1*a*). Kinetic characterization using the short modified tripeptide Tos-Gly-Pro-Lys-4-nitroaniline as a chromogenic substrate showed no significant difference between wild-type and K30R LysC (data not shown). By contrast, activity data obtained using the longer model substrate melittin (26 residues) showed an influence of the K30R mutation (Fig. 1*b*). While for wild-type LysC a rate constant of $k_{obs} = 0.0727$ min^{-1} was detected, the K30R variant displayed a 1.6-fold increased k_{obs} of 0.1175 min^{-1}. It is possible that, in contrast to the longer melittin, the tripeptide does not contact all parts of the LysC substrate-binding site, especially in the vicinity of position 30 (see below). In order to address potential structural differences between wild-type and K30R LysC, the respective TLCK complexes were crystallized and diffraction data were collected to better than 1.1 Å resolution.

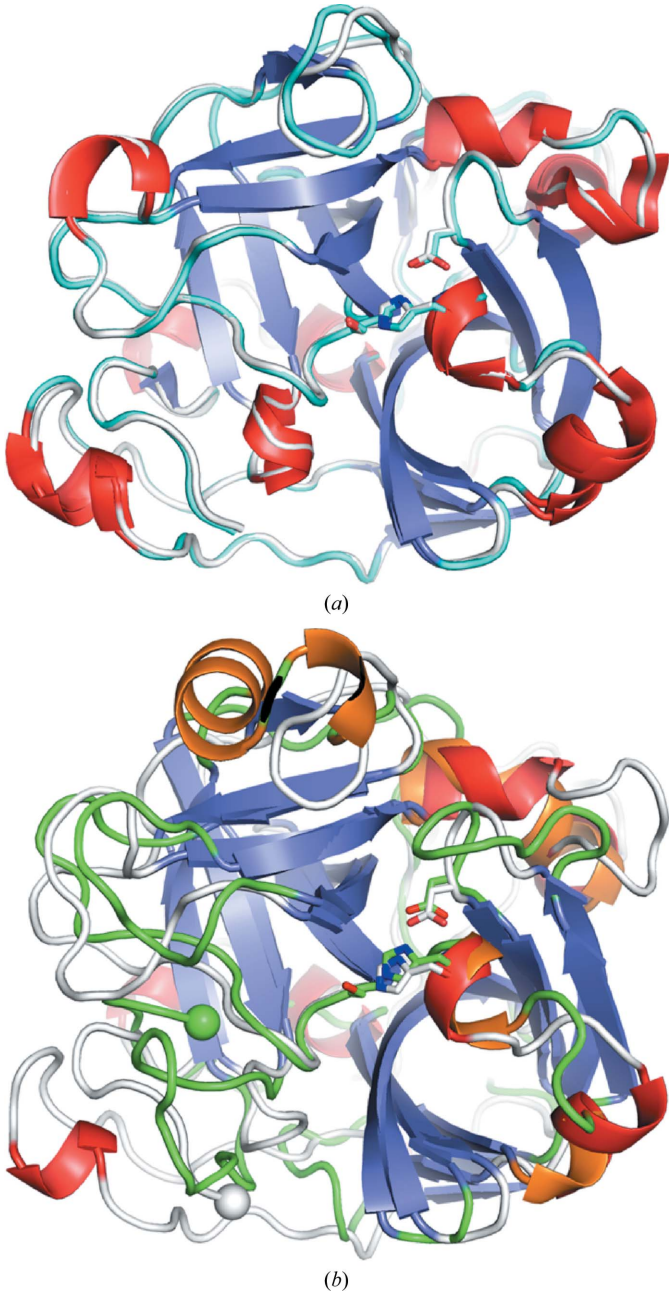

Figure 3
Comparison of LysC with its closest structural homologues. (*a*) Superposition with *A. lyticus* protease I (PDB entry 1arb) displayed with cyan loop regions. (*b*) Superposition with *Fusarium oxysporum* trypsin (PDB entry 1try; green loop regions, orange helices; Rypniewski *et al.*, 1995) reveals a core of 155 C^{α} atoms including the two β-barrels that superpose with an r.m.s.d. of 2.0 Å. Large differences are apparent in the N-terminal parts (N-termini marked as spheres) and in the loop regions that connect the β-strands, most notably an additional α-helix in trypsin that is absent in LysC. *L. enzymogenes* LysC is coloured as in Fig. 2(*b*). The catalytic triads in both panels are shown as stick models. In LysC, the triad consists of Ser194, His57 and Asp113. The canonical nomenclature of Ser195, His57 and Asp102 would have required insertion codes, which was not performed for LysC.

3.2. LysC substrate-binding site and differences between wild-type and K30R LysC

The diffraction data for LysC were indexed in a primitive monoclinic setting. Based on a packing density of 2.1 Å^3 Da^{-1}, two molecules of 260 residues per asymmetric unit were estimated. A self-Patterson map was featureless, but a self-rotation function calculated from data reduced in space group *P*2 showed additional twofold axes perpendicular to the crystallographic twofold and spaced 90° apart, indicating

pseudo-orthorhombic symmetry (Fig. 2*a*). From packing considerations, this higher metric symmetry is in principle possible for LysC since it would lead to a single molecule in the orthorhombic asymmetric unit. However, the β angle of 97° (Table 1) does not support this. After molecular replacement, a twofold NCS axis parallel to the *c* axis but that does not intersect with the crystallographic *b* axis was found to relate the two protomers (Fig. 2*b*). The models for the TLCK complexes of wild-type and K30R LysC were refined to $R_{\rm free}$ values of 20.8 and 15.8%, respectively, and display excellent stereochemistry (Table 1). LysC adopts the trypsin fold of two topologically equivalent six-stranded β-barrels (Figs. 2*b* and 3). The two protomers in the asymmetric unit of wild-type and K30R LysC superpose with r.m.s.d.s of 0.10 and 0.14 Å, respectively. The asymmetric units of wild-type and K30R LysC superimpose with an r.m.s.d. of 0.08 Å. Thus, except for the mutated residue, the two structures may be considered to be identical and all further discussion here relates to the K30R variant owing to its higher resolution (0.9 Å compared with 1.1 Å for the wild type).

According to domain alignment (Holm *et al.*, 2008), the closest member of known structure is *A. lyticus* protease I, with an r.m.s.d. of 0.5 Å over the entire sequence (263 residues; *DALI* score of 50; Fig. 3*a*). Interestingly, the next nearest structural neighbours are much less related to LysC. Trypsin, chymotrypsin, V8 protease and enteropeptidase superimpose with LysC with an r.m.s.d. of 2.7 Å over 220 residues and all have significant *DALI* scores of 19 (where <5 indicates dissimilarity), but all are unsuitable as models for molecular-replacement searches. Therefore, LysC, *A. lyticus* protease I and *P. aeruginosa* endopeptidase, the structure of which has not yet been determined, seem to constitute their own subfamily within the trypsin-like proteases. Notable differences between LysC and trypsin include large deviations in the first 30 N-terminal residues (left side of Fig. 3*b*) and an additional α-helix in trypsin that is a loop region in LysC (top of Fig. 3*b*). As might be expected, the secondary-structure elements that contribute the catalytic triad are conserved.

The lysine moiety of TLCK binds to the so-called S1 specificity pocket in the substrate-binding site of LysC and is a surrogate for the peptide substrate. A surface channel borders the S1 pocket that is partially occupied by TLCK and into which a peptide can be modelled (Fig. 4*a*). The peptide conformation agrees well with those of related protease–peptide complexes (Wilmouth *et al.*, 2001; Kügler *et al.*, 2012; Lechtenberg *et al.*, 2013; PDB entries 1hax, 4bxw and 4a1t) and includes several possible hydrogen bonds between the main chain of LysC and the substrate (Fig. 4*a*). The C-terminal half of the substrate-binding site is bordered by Pro191 and Trp35. The indole side chain of Trp35 separates residue position 30 from the substrate channel; hence, there should not be a direct effect of the K30R mutation on substrate binding in LysC. Yet, the K30R variant of the *A. lyticus* protease I displays higher cleavage efficiency compared with the wild type (Asser *et al.*, 1997) and melittin is hydrolysed 1.6-fold more efficiently by the K30R variant of *L. enzymogenes* LysC compared with the wild type. Comparison of the environments of the two structures at position 30 reveals both Lys30 and Arg30 to pack with their side chains onto Trp35 and to form van der Waals contacts with Ile65. Owing to the larger area of the Arg30 guanidinium group, this hydrophobic sandwich seems to be more stable than in the case of Lys30. The tips of both side chains are solvent-exposed and order water molecules. However, Lys30 has only a single ordered water molecule bound to the N^{ζ} atom, whereas Arg30 binds to three water molecules. In addition, the guanidinium group of Arg30 forms two extra hydrogen bonds to the main-chain carbonyl O atom of Gly59, while Lys30 does not directly bind to Gly59. The full hydrogen-bonding potential of Arg30 is used. Thus, although the conformational differences are small, the

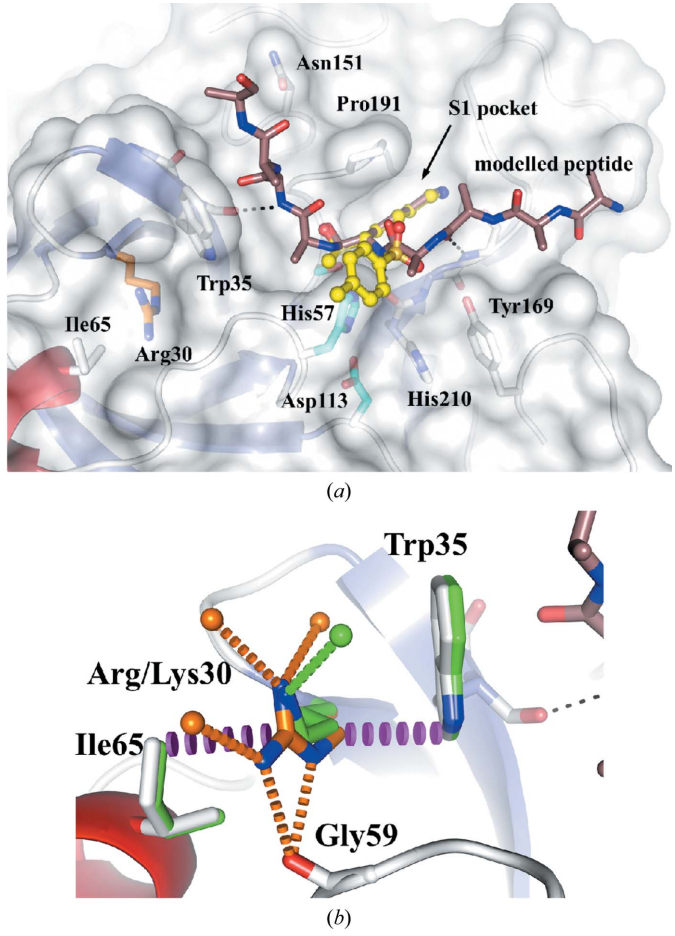

Figure 4
Substrate-binding sites in wild-type and K30R LysC. (*a*) Surface representation of the LysC K30R protein part and replacement of the TLCK inhibitor with a model peptide. The S1 pocket binds the lysine side chain of the substrate (or TLCK inhibitor). A channel on the surface is lined by residues Asn151, Trp35, Pro191, His210 and Tyr169, suggesting a likely trace of the peptide. The side chains of the catalytic triad are shown as cyan stick models. Arg30 (orange) does not directly partake in substrate binding but abuts on Trp35. (*b*) A close-up view of position 30 reveals little conformational difference between the wild-type and K30R structures. Both Lys30 and Arg30 are in van der Waals contact with Ile65 and Trp35 (magenta dashes). Whereas Lys30 in wild-type LysC does not form hydrogen bonds to the protein, arginine in the K30R LysC variant forms two hydrogen bonds to the carbonyl group of Gly59. In addition, only a single water molecule is directly bound to Lys30, while Arg30 binds to three water molecules.

stronger hydrophobic sandwich of Arg30 between Ile65 and Trp35 and the additional hydrogen-bond interactions of its side chain would argue for a stabilizing effect of the K30R variant compared with wild-type LysC. Such a stabilizing effect might explain the increased efficiency of the K30R variant towards larger substrates compared with wild-type LysC. Small substrates such as Tos-Gly-Pro-Lys-4-nitroaniline that do not reach the binding site near position 30 are not distinguished by the two enzymes. Larger substrates such as melittin will bind close to position 30 and are hydrolysed more efficiently by the K30R variant (Fig. 1*b*).

The geometry of the S1 pocket allows conclusions on the substrate specificity of LysC for lysine. The narrow S1 pocket is bordered by main-chain amide groups and thus can only accommodate linear side chains (Fig. 5*a*). Specific interactions are formed with the tip of the side chain, imposing a lower limit on its length. Arginine and lysine are the only side chains that are long enough to reach the bottom of the S1 pocket. The discrimination of LysC against arginine is established by Asp225, which reaches from the side into the S1 pocket and is sterically incompatible with the branched guanidinium group of arginine (Fig. 5*b*). Interestingly, an aspartate at the same location in fiddler crab collagenase is quite important for arginine substrate binding, but here the S1 pocket is more voluminous, allowing both basic amino acids to bind (Tsu *et al.*, 1997). Thus, a combination of S1 pocket volume and placement of acidic charge determines arginine preference in different proteases.

The lysine moiety of TLCK engages in similar interactions as in *A. lyticus* protease I, *i.e.* the ammonium group forms three hydrogen bonds in a tripod-like arrangement with the side chain of Asp225 and the main-chain carbonyl groups of Thr189 and Ser214. Asp225 ensures electroneutrality in the S1 pocket when substrate is bound. The LysC substrate-binding mode in the S1 pocket differs significantly from the arginine recognition present in a number of serine proteases, including trypsin and the coagulation cascade proteases thrombin, factor VIIa and factor Xa, in which an aspartate at the bottom of the S1 pocket forms a bidentate hydrogen bond to the guanidinium group. While having a narrow entrance that selects slender side chains, the S1 pocket in the arginine-specific proteases is larger at the bottom to accommodate not only the bulky guanidinium group of arginine but also water molecules that interact with it (Fig. 5). For trypsin, mutagenesis of glycine residues 216 and 226 in the S1 pocket to alanine shifted the relative substrate preference to arginine or lysine, respectively. Displacement of water from the S1 pocket and interactions of the introduced methyl group with the substrate were held to be responsible for this mutation-specific effect (Craik *et al.*, 1985). In conclusion, the hydrogen-bonding pattern for the tip of the side chain, the volume restriction at the bottom of the S1 pocket and the counter-selection of arginine by Asp225 ensure the lysine specificity of LysC.

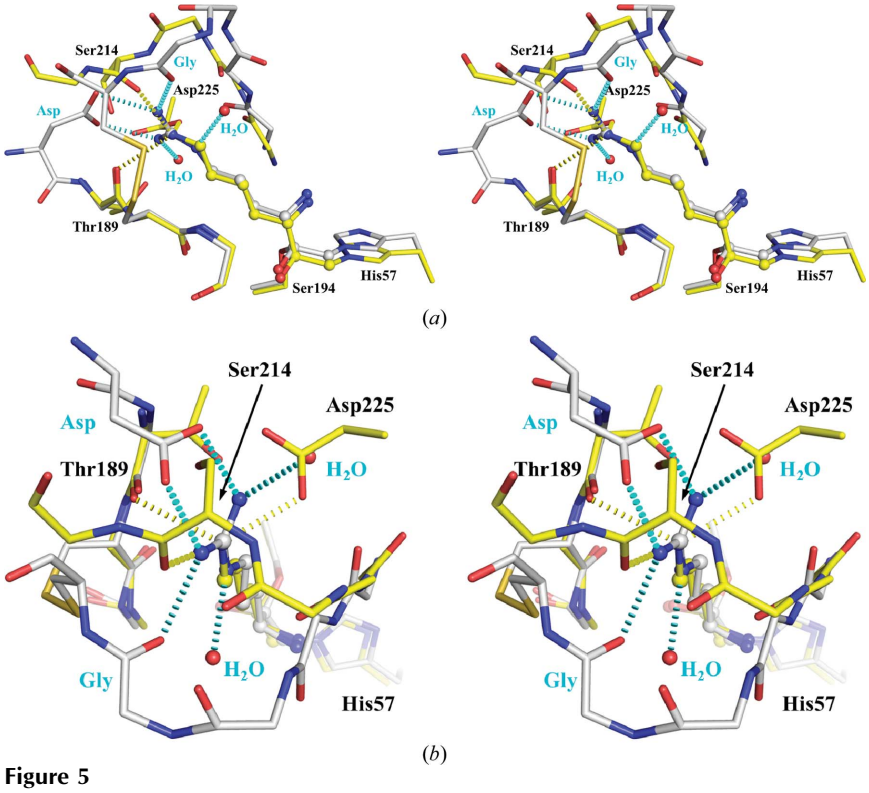

Figure 5
Substrate specificity of LysC and discrimination against arginine. (*a*) Cross-eyed stereoview of the superposition of the S1 pockets of LysC K30R (yellow C atoms) with the Phe-Pro-Arg-CH_2Cl–meizothrombin complex (Papaconstantinou *et al.*, 2008; grey C atoms; PDB entry 3e6p). For clarity, only the lysine and arginine side chains of the inhibitors are shown. Three hydrogen bonds between the ammonium group of lysine and LysC are drawn as yellow dashed lines. The arginine in the meizothrombin complex forms hydrogen bonds (cyan dashed lines) with an aspartate and a main-chain carbonyl group and with two water molecules. (*b*) Cross-eyed stereoview from the bottom of the S1 pocket. Asp225 in LysC would clash with the guanidinium group of an arginine, thus selecting for lysine.

3.3. TLCK binding to *L. enzymogenes* LysC

Chloromethyl ketones such as TLCK (Fig. 6*a*) are well known covalent inhibitors of cysteine and serine proteases alike. Two covalent bonds are formed with serine proteases, one each to the active-site serine and histidine. In LysC, a tetrahedral inhibition complex with a tertiary alkoxy or alcohol function (see below) is formed after reaction with TLCK (Fig. 6*b*). The TLCK complexes of wild-type and K30R LysC are virtually identical around the inhibitor (not shown). A conceivable reaction sequence is attack of Ser194 followed by independent nucleophilic substitution of the Cl atom by His57. Alternatively, a kinetically controlled ring closure of the initially formed alkoxy anion to an oxirane might occur (Darzens reaction) followed by nucleophilic ring opening from His57. The torsion angles of the resulting inhibitor complex are energetically favourable apart from the sulfonamide bond in TLCK (upper arrow in Fig. 6*c*). Usually, sulfonamides prefer an *X*—

S—N—H torsion angle of 180° such that the amide H atom bisects the O—S—O angle in a *cis*-like conformation. In the LysC–TLCK complex, this torsion angle is 77° (Fig. 6*c*), a conformation that is not populated in the Cambridge Structural Database (CSD) of small-molecule structures. The preferred torsion angle cannot be reached owing to steric hindrance of the tosyl group by Pro191. The energetic penalty for the distortion is partially offset by packing of the tosyl group on His57 (Fig. 6*c*) and by crystal contacts with the other protomer of the crystallographic dimer (Fig. 2*b*). A similar situation is present in the *A. lyticus* protease I–TLCK complex (Tsunasawa *et al.*, 1989), which conserves the proline and also imposes crystal-packing restraints onto the tosyl group, resulting in a torsion angle of 61° (green in Fig. 6*c*; PDB entry 1arc). Unfortunately, the TLCK entity in 1arc was somewhat imprecisely parameterized: the tosyl group is not planar, the carbonyl group is intact, not tetrahedral as it should be, and the electrophilic C atom is located at an impossibly close distance of 1.95 Å to the active-site Ser194 O^{γ} atom (lower arrow in Fig. 6*c*). No diffraction data from 1989 that would enable re-refinement of the coordinates are available. Sadly, a search for the tetrahedral substructure of a reacted chloromethyl ketone in the PDB revealed other physically impossible examples for serine proteases (PDB entries 2b8o, 2fir and 4bxw) and a cysteine protease (PDB entries 4eha, 4ehd, 4ehf, 4ehh, 4ehk, 4ehl and 4ehn). From deposited Fourier coefficients we found a single example of a chemically plausible chloromethyl ketone adduct in the 2.1 Å resolution complex of meizothrombin, an active form of thrombin after single cleavage of pro-thrombin, crystallized in complex with the chloromethylated tripeptide Phe-Pro-Arg-CH_2Cl (Papaconstantinou *et al.*, 2008). As a result, to our knowledge the LysC–TLCK structures discussed here currently represent the only complexes of a chloromethyl ketone with a serine protease at atomic resolution.

3.4. The oxyanion hole in LysC can also be an 'alco-hole'

The concept of the oxyanion hole in proteases and other hydrolases, which stabilizes a developing negative charge at the carbonyl O atom during attack by a negatively charged nucleophile (Knowles, 1991), bears resemblance to the P-loop of Walker A-type nucleotide-binding proteins (Smith & Rayment, 1996). In both cases, amide bond NH groups orient their H atoms towards the O atom whose negative charge is to be stabilized. In the case of the P-loop-containing proteins, Mg^{2+} (Smith & Rayment, 1996) and the positive end of a helix dipole (Myllykoski *et al.*, 2012; Herzberg *et al.*, 1992) further stabilize the negative charge on bound phosphate groups. While a helix dipole in subtilisin helps to activate the catalytic serine (Hol, 1985), neither a counterion nor a helix dipole are present in LysC and trypsin-type proteases. The question arises whether the NH groups alone are sufficient to stabilize an oxyanion or whether a hydroxyl group might indeed ensue after protonation of the tetrahedral intermediate. Using the high-resolution LysC–TLCK complex, an attempt was made to distinguish between a hydroxyl group and an alkoxy anion by means of bond-length estimates and difference map electron density for H atoms.

The LysC K30R–TLCK model was refined to convergence using *PHENIX* under the assumption of a deprotonated O atom, as expected for the tetrahedral intermediate of peptide hydrolysis bound to the oxyanion hole. However, an $mF_o - DF_c$ electron-density map showed peaks of >5.5 r.m.s.d. for

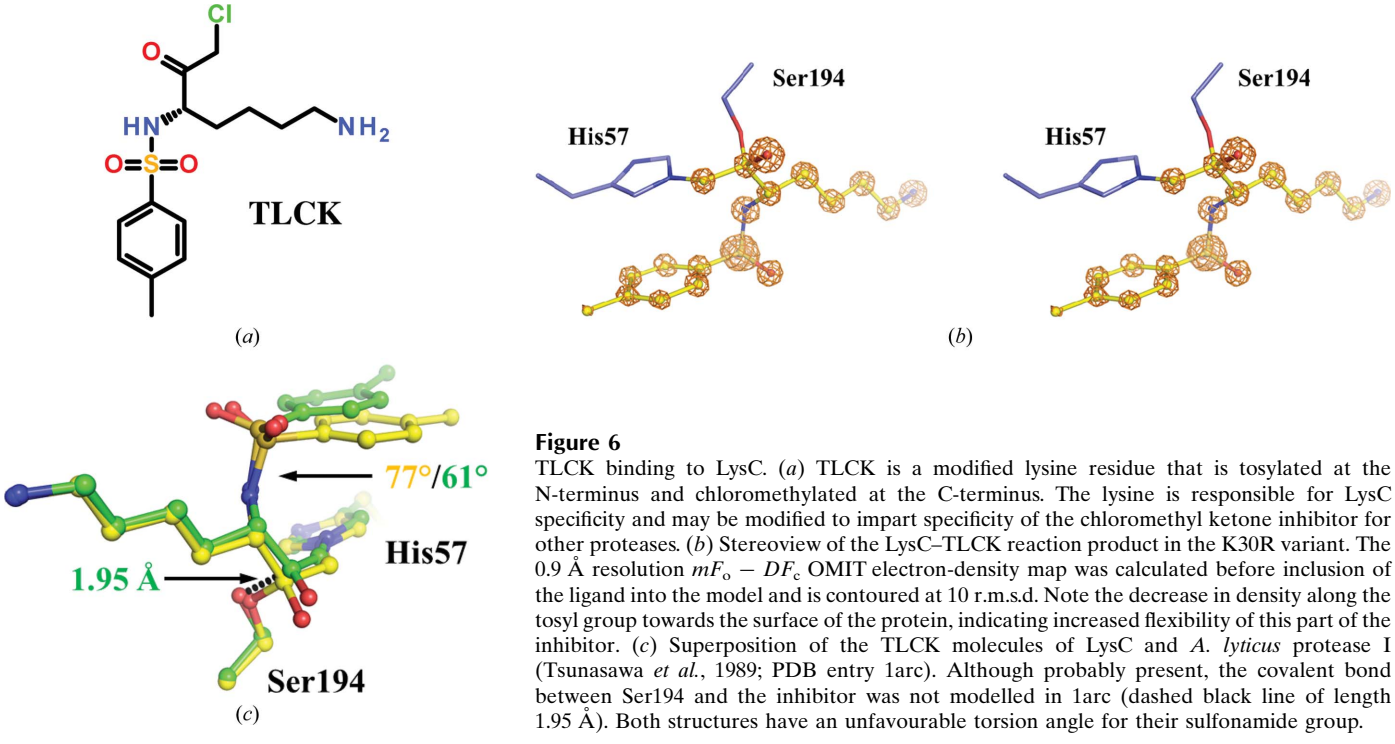

Figure 6
TLCK binding to LysC. (*a*) TLCK is a modified lysine residue that is tosylated at the N-terminus and chloromethylated at the C-terminus. The lysine is responsible for LysC specificity and may be modified to impart specificity of the chloromethyl ketone inhibitor for other proteases. (*b*) Stereoview of the LysC–TLCK reaction product in the K30R variant. The 0.9 Å resolution $mF_o - DF_c$ OMIT electron-density map was calculated before inclusion of the ligand into the model and is contoured at 10 r.m.s.d. Note the decrease in density along the tosyl group towards the surface of the protein, indicating increased flexibility of this part of the inhibitor. (*c*) Superposition of the TLCK molecules of LysC and *A. lyticus* protease I (Tsunasawa *et al.*, 1989; PDB entry 1arc). Although probably present, the covalent bond between Ser194 and the inhibitor was not modelled in 1arc (dashed black line of length 1.95 Å). Both structures have an unfavourable torsion angle for their sulfonamide group.

both complexes in the asymmetric unit at the same position near the former carbonyl O atom that are compatible with the presence of an H atom (Fig. 7*a*). Electron density for two more H atoms is visible at the *ortho* positions of the tosyl group, underscoring the quality of the diffraction data. The presence of density for these three H atoms at this resolution is notable: aromatic C—H bonds are shorter than alkyl C—H bonds owing to the asymmetric distribution of π electrons in Hückel systems. Likewise, the polar H atoms at hydroxyl groups tend to exhibit rotational disorder unless restrained by hydrogen bonds. These H atoms must therefore be ordered exceptionally well, in contrast to most other H atoms of TLCK, for which no electron density above the noise level is observed. In summary, the electron density at the O atom in the oxyanion hole strongly supports the presence of a hydroxyl group.

The presence of a hydroxyl group *versus* an alkoxide should also manifest itself in the corresponding C—O bond length. Using full-matrix refinement in *SHELXL* to estimate errors in bond lengths and angles for TLCK, a mean C—O distance of approximately 1.384 Å was derived for the two independent LysC K30R–TLCK complexes (precise numbers are given in Fig. 7*b*). A search in the CSD for sp^3-hybridized C atoms in acyclic hemiacetals and their deprotonated alkoxides yielded mean bond lengths of 1.397 ± 0.015 Å (n = 60) and 1.351 ± 0.033 Å (n = 15), respectively. The alkoxides have shorter C—O bond lengths than the hydroxyl groups in hemiacetals. The mean bond lengths of the hemiacetals correspond well to the experimental data for the TLCK complex. Since the error estimates in bond lengths are <0.02 Å (Fig. 7*b*), bond-length considerations support a hydroxyl group bound to the oxyanion hole in the TLCK complex.

In conclusion, both lines of evidence, electron density for ordered H atoms and bond-length error estimates, support a model in which the tetrahedral intermediate during peptide hydrolysis could be protonated. While an oxyanion will always form as a first step during nucleophilic attack of serine at the carbonyl group of the cleavable peptide bond, it can be rapidly protonated from bulk solvent. The oxyanion hole is not only solvent-accessible in the LysC–TLCK complex (Fig. 7) but also in the peptide-bound form in other proteases (model in Fig. 4*a*; Wilmouth *et al.*, 2001; Kügler *et al.*, 2012; Lechtenberg *et al.*, 2013), enabling facile transfer of protons to and from the intermediate. The oxyanion hole therefore may not need to stabilize a highly energetic alkoxide with a pK_a of >16, but serves to hydrogen-bond both electron lone pairs of the hydroxyl O atom while the third valence is pointing into, and is protonated by, solvent. A similar hydroxyl group in the oxyanion hole has been observed for a boronate ester inhibitor complex, which is also a mimic of the tetrahedral intermediate of amide hydrolysis. In the 0.9 Å resolution crystal structure of α-lytic protease from *L. enzymogenes* reacted with a peptidyl boronic acid, electron density for an H atom was observed at the O atom of the boronate adduct, indicating the presence of a hydroxyl group in the oxyanion hole (Fuhrmann *et al.*, 2006). Compared with alkoxides, boronic acids and their tetrahedral esters have rather low pK_a values of between 7 and 9. The presence of a hydroxyl group in the boronate adduct of α-lytic protease would therefore indicate a strong propensity of the oxyanion to retrieve a proton. This propensity would be even stronger in a more basic alkoxide of pK_a >16. By contrast, neutron structure analysis of elastase in complex with a highly activated trifluoromethyl ketone peptidomimetic (FR130180) did not reveal a deuteron in the nuclear maps, consistent with the presence of a genuine oxyanion (Tamada *et al.*, 2009). Although rotational disorder can account for the absence of (nuclear) density, there are two hydrogen bonds from the alkoxide to NH groups in the oxyanion hole, which efficiently prevent the alkoxide C—O bond from rotating. It is quite likely that the pK_a of the trifluoromethyl hemiketal adduct is low enough for it to function as a Brønsted acid in aqueous solution, leaving a stable oxyanion.

3.5. Modulation of the pH optimum by an aromatic stack

In serine proteases, the N$^{\varepsilon}$ atom of His57 in the triad serves as the general base for hydrogen abstraction from the nucleophilic serine. The pH optima for the three lysine-specific proteases LysC, *A. lyticus* protease I and *P. aeruginosa* endopeptidase are in the alkaline range, typically between pH 8 and pH 9. In addition to the catalytic triad (Ser194, His57 and Asp113), two more residues have been identified as crucial for the elevated pH optimum of *A. lyticus* protease I.

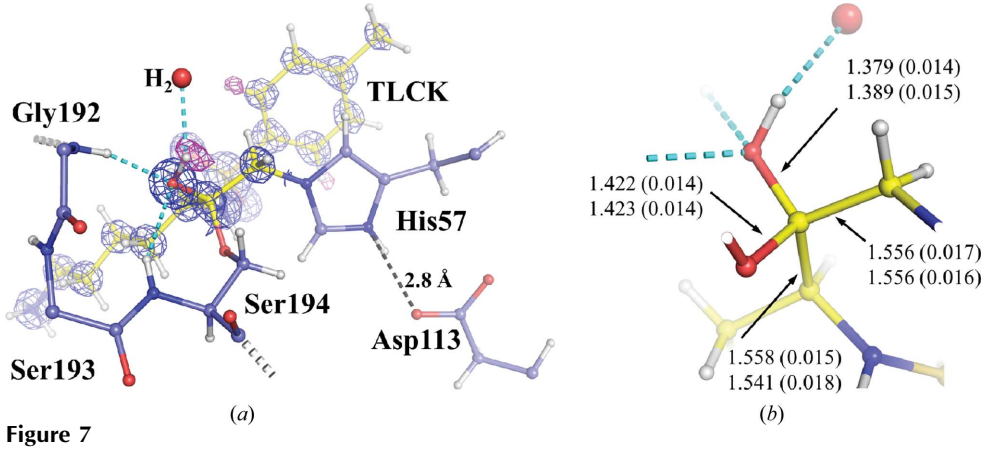

Figure 7
A hydroxyl group is present in the LysC oxyanion hole. (*a*) $mF_o - DF_c$ H-atom OMIT map of TLCK shows 4 r.m.s.d. difference map peaks (magenta) at some, but not all, expected H-atom positions. Difference density is apparent at the *ortho* positions of the tosyl group and at the O atom of the tetrahedral adduct. The $2mF_o - DF_c$ map contoured at 3 r.m.s.d. is drawn as a blue mesh. (*b*) Bond-length error estimates from matrix-inversion refinement using *SHELXL*. The two values refer to the independent molecules in the asymmetric unit. Errors are given in parentheses.

Trp169 and His210 stack on top of Asp113 and seem to modify the pK_a value of the acid by an electrostatic mechanism (Shiraki, Norioka, Li & Sakiyama, 2002; Shiraki, Norioka, Li, Yokota *et al.*, 2002). In *L. enzymogenes* LysC with its pH optimum of 8.5–8.8 the corresponding residues are Tyr169 and His210 (Fig. 8*a*). Thus, while His210 is retained, another aromatic side chain (Tyr169) can exert the same effect on the pH optimum. Interestingly, His210 is not conserved in *P. aeruginosa* endopeptidase, which has a pH optimum of 8–9 (Elliott & Cohen, 1986). Here, tyrosine residues are present at positions 169 and 210. The high pH optimum of these three proteases is thus owing to an aromatic stack of variable constitution that packs on top of Asp113. In trypsin, which has a neutral pH optimum, the stack is interrupted by the presence of a serine residue at position 210 (Fig. 8*b*). Mutagenesis of *A. lyticus* protease I has demonstrated that serine and alanine can substitute for His210 without loss of catalytic efficiency, but

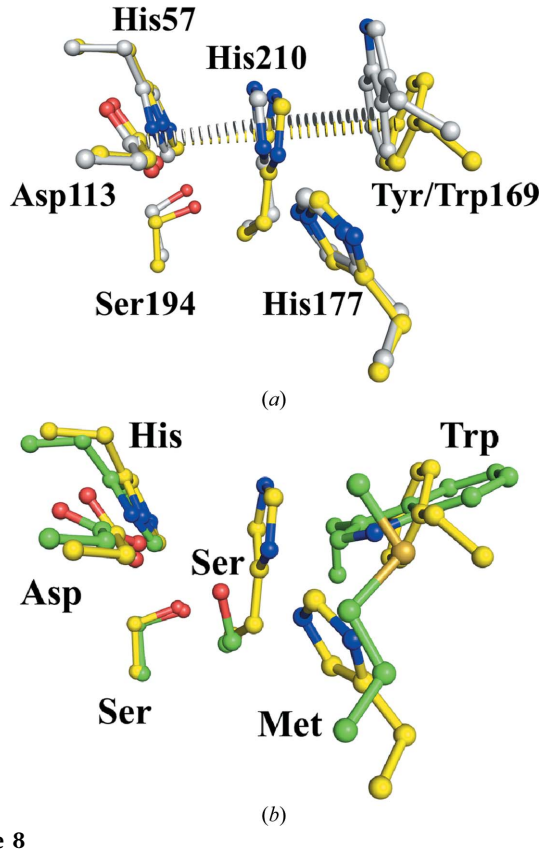

Figure 8
The aromatic stack that is responsible for the high pH optimum of LysC. (*a*) The residues in LysC (yellow) and *A. lyticus* protease I (grey) that border the catalytic triad (left side) and form an aromatic stack are compared. The van der Waals distances between His210 and both Asp113 and residue position 169 (Trp in *A. lyticus* protease I and Tyr in LysC) are all 3.7–3.8 Å. The stacks are not entirely coplanar, as would be expected for maximal interaction. A second histidine (His177) is at hydrogen-bonding distance from the stacking His210 and could further influence the pK_a of the catalytic His57. (*b*) A comparison of LysC with *F. oxysporum* trypsin (PDB entry 1try; green; Rypniewski *et al.*, 1995) shows the absence of the stack in trypsin-like serine proteases. While a tryptophan is present at the location of LysC Tyr169, a serine replaces His210. Also the second histidine of LysC (His177) is replaced by a methionine.

the pH optimum decreases by two units to pH 6.3 (Shiraki, Norioka, Li, Yokota *et al.*, 2002). As expected from sequence comparisons between the *A. lyticus*, *P. aeruginosa* and *L. enzymogenes* enzymes, residue position 169 should be aromatic to retain the high pH optimum (but see the exception given below). Indeed, with the exception of histidine, all other aromatic side chains at position 169 retain the high pH optimum in *A. lyticus* protease I (Shiraki, Norioka, Li, Yokota *et al.*, 2002).

The question arises as to what the electronic mechanism for the increase in pH optimum by the aromatic stack could be. Isolating one face of the charged His57/Asp113 hydrogen bond from solvent by the aromatic stack will increase Coulomb interactions owing to a decrease in the local dielectric constant. For this isolation, other large hydrophobic amino acids should suffice, and indeed the Trp169Leu mutant in *A. lyticus* protease I still retains its high pH optimum, albeit with a fourfold reduced catalytic efficiency (Shiraki, Norioka, Li, Yokota *et al.*, 2002). The increased electron density at the catalytic triad owing to the stacking might further increase the pK_a of His57, possibly beyond 12 (Lin *et al.*, 1998). The stacking and the short contact between His57 and Asp113 (see below) increase the electron density of His57, thereby increasing the pK_a value of the N^{ε} atom. Trypsin-type and chymotrypsin-type serine proteases with this His57 environment typically have neutral pH optima, with His57 displaying a pK_a of ∼7.5 in the unperturbed state. This pK_a value can increase to values between 10.3 and 12.1, as measured for chymotrypsin in complex with tetrahedral hemiketal adducts of peptidyl trifluoromethyl ketones (Lin *et al.*, 1998). If this analogy holds true for LysC, the basicity of His57 (owing to the combined influence of the aromatic stack and the short hydrogen bond to Asp113) will eventually be too high to retain a deprotonated imidazole as the catalytic base under neutral conditions, and consequently the physiological pH optimum of the enzyme will shift to the alkaline range. NMR titration experiments on His57 similar to those performed for chymotrypsin (Lin *et al.*, 1998) should shed light on this issue.

3.6. No evidence for a low-barrier hydrogen bond in LysC

An early neutron study identified the catalytic histidine as the terminal proton acceptor in trypsin (Kossiakoff & Spencer, 1980), indicating that the role of the catalytic aspartate is to stabilize the imidazolium ion. The charge relay (not proton relay) between the imidazolium and aspartate ions of the catalytic triad of serine proteases often gives rise to an unusually short hydrogen bond ranging between 2.48 and 2.77 Å (Tamada *et al.*, 2009; Liebschner *et al.*, 2013). A debate ensued whether this distance is best explained by a short ionic hydrogen bond (SIHB) with strict separation of the H atom (Fuhrmann *et al.*, 2006) or by a low-barrier hydrogen bond (LBHB) where the H atom is shared midway between two non-H atoms of comparable pK_a value (Cleland *et al.*, 1998; Lin *et al.*, 1998; Katona *et al.*, 2002; Ohnishi *et al.*, 2013). While NMR data on chymotrypsin (Frey *et al.*, 1994), the ultrahigh-resolution (0.78 Å) crystal structure of *Bacillus lentis* subtilisin

(PDB entry 1gci; Kuhn *et al.*, 1998) and a 0.95 Å resolution crystal structure of porcine pancreatic elastase (Katona *et al.*, 2002) indicated sharing of the H atom between His and Asp to some degree, other high-resolution crystal structures (0.8–1.3 Å) and other NMR data did not. For instance, a NMR study on *cis*-urocanate, a model compound for the His-Asp dyad, found no evidence for an LBHB based on chemical shifts (Ash *et al.*, 1997). Neutron diffraction studies on porcine pancreatic elastase explicitly located the deuteron at the imidazole side chain (Tamada *et al.*, 2009), as did ultrahigh-resolution (0.83 Å) X-ray crystal structures of α-lytic protease for the H atom (Fuhrmann *et al.*, 2004, 2006).

In the LysC protomers the distances between the His57 $N^{\delta1}$ and Asp112 $O^{\delta2}$ atoms are 2.76 and 2.77 Å, respectively (Fig. 7*a*). The distances approach the van der Waals limit of 2.7 Å, but are at the upper end of the range observed for short hydrogen bonds (Tamada *et al.*, 2009). During catalysis, this hydrogen bond might be further compressed to <2.7 Å, giving rise to either an SIHB or an LBHB. No electron density for an H atom is visible at 0.9–1.0 Å distance from His57 in the maps. Therefore, either the diffraction data are of insufficient quality or, should a LBHB be present, the H-atom location is disordered owing to a mixture of states when sharing the H atom with Asp113. While the LysC–TLCK complex is a good mimic for a tetrahedral intermediate, it does not convey a true picture of catalysis because of the covalent bond between the catalytic base His57 and the inhibitor. The covalent bond to TLCK relieves the strain that would be present in the active site if a peptidyl tetrahedral intermediate is bound and His57 is at a van der Waals distance to protonate the leaving group. Strong LBHBs are formed when the pK_a values of the donor and the corresponding acid of the acceptor match (Cleland *et al.*, 1998). The inductive effect of an additional covalent bond to His57 is likely to increase its pK_a value, leading to a more pronounced ΔpK_a with Asp113, thus disfavouring an LBHB.

4. Conclusions and outlook

The crystal structure of LysC defines the substrate-specificity determinants of this lysine-specific protease. Somewhat to our surprise, the K30R variant of LysC, which had superior properties compared with the wild type in a related enzyme, does not exhibit marked structural changes. The K30R position is shielded from the substrate-binding site by Trp35 and only subtly alters its properties. From the crystal structure of LysC it is apparent that all of the lysine residues (positions 30, 48, 49, 155 and 203) are solvent-accessible. Thus, the increased performance of the K30R variant might simply reflect the absence of a self-cleavage site. To generate a more stable protease for biotechnological use, mutation of each lysine individually or in combination may prove to be useful. The atomic resolution data allowed the visualization of a hydroxyl group in the oxyanion hole of LysC. This novel possible intermediate along the reaction pathway of proteolysis might also be traversed by other serine proteases.

Tetrahedral intermediates in serine proteases arise twice during proteolysis, firstly after the nucleophilic addition of serine to the substrate and secondly after the attack of the acyl-enzyme intermediate by water. These intermediates were visualized crystallographically many years ago, but many of the coordinates from that time (Rühlmann *et al.*, 1973; Sweet *et al.*, 1974; Matthews *et al.*, 1975; James *et al.*, 1980; Tulinsky & Blevins, 1987) seem to have been lost. Tetrahedral adducts have been produced from reactive aldehydes (Delbaere & Brayer, 1985), ketones (Marquart *et al.*, 1983) and boronic acids (Bone *et al.*, 1987), sometimes to atomic resolution (Fuhrmann *et al.*, 2006). The TLCK complexes described here at atomic resolution provide further insight into the protonation state of the O atom in the oxyanion hole. Our data are in accordance with a hydroxyl group located in the oxyanion hole. Of course, if the lifetime of the oxyanion intermediate is of the order of femtoseconds, the anion is too short-lived for a protonation event to occur. Yet, a lifetime of a few picoseconds should be enough to allow access of a proton to the alkoxide, possibly aided by a Grotthuss-like proton-transfer mechanism that is faster than diffusion-controlled (Cukierman, 2006). Depending on its enzymatic properties, a serine protease may either have a hemiketal or just an oxyanion as an intermediate along the reaction coordinate. Integration of the *A. lyticus* protease I mutagenesis data with the *L. enzymogenes* LysC crystal structure revealed that an aromatic stack on the catalytic Asp113 is likely to be responsible for the unusually high pH optimum of this subfamily of serine proteases. A short hydrogen bond, solvent exclusion and electrostatic effects on the charge-stabilizing aspartate were identified as the main contributors to this effect. This information might be useful in trying to adjust the pH optima of other serine proteases by site-directed mutagenesis.

We thank the staff at SLS beamline PX II and our colleagues at Expose for support during synchrotron data collection, David Borer for SDS–PAGE analysis and Bernd Kuhn for helpful discussions on protease active sites.

References

Ash, E. L., Sudmeier, J. L., De Fabo, E. C. & Bachovchin, W. W. (1997). *Science*, **278**, 1128–1132.
Asser, A., Per, B., Sven, B. & Sven, H. (1997). Patent No. WO 1997033984 A1.
Bone, R., Shenvi, A. B., Kettner, C. A. & Agard, D. A. (1987). *Biochemistry*, **26**, 7609–7614.
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.
Cleland, W. W., Frey, P. A. & Gerlt, J. A. (1998). *J. Biol. Chem.* **273**, 25529–25532.
Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R. & Rutter, W. J. (1985). *Science*, **228**, 291–297.
Cukierman, S. (2006). *Biochim. Biophys. Acta*, **1757**, 876–885.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1995). *Curr. Opin. Struct. Biol.* **5**, 784–790.
Delbaere, L. T. & Brayer, G. D. (1985). *J. Mol. Biol.* **183**, 89–103.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Elliott, B. W. Jr & Cohen, C. (1986). *J. Biol. Chem.* **261**, 11259–11265.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Frey, P. A., Whitt, S. A. & Tobin, J. B. (1994). *Science*, **264**, 1927–1930.
Fuhrmann, C. N., Daugherty, M. D. & Agard, D. A. (2006). *J. Am. Chem. Soc.* **128**, 9086–9102.
Fuhrmann, C. N., Kelch, B. A., Ota, N. & Agard, D. A. (2004). *J. Mol. Biol.* **338**, 999–1013.
Gerber, P. R. & Müller, K. (1995). *J. Comput. Aided Mol. Des.* **9**, 251–268.
Herzberg, O., Reddy, P., Sutrina, S., Saier, M. H. Jr, Reizer, J. & Kapadia, G. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 2499–2503.
Hol, W. G. J. (1985). *Prog. Biophys. Mol. Biol.* **45**, 149–195.
Holm, L., Kääriäinen, S., Rosenström, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.
James, M. N., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. & Bauer, C. A. (1980). *J. Mol. Biol.* **144**, 43–88.
Jekel, P. A., Weijer, W. J. & Beintema, J. J. (1983). *Anal. Biochem.* **134**, 347–354.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Katona, G., Wilmouth, R. C., Wright, P. A., Berglund, G. I., Hajdu, J., Neutze, R. & Schofield, C. J. (2002). *J. Biol. Chem.* **277**, 21962–21970.
Knowles, J. R. (1991). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **332**, 115–121.
Kossiakoff, A. A. & Spencer, S. A. (1980). *Nature (London)*, **288**, 414–416.
Kügler, J., Schmelz, S., Gentzsch, J., Haid, S., Pollmann, E., van den Heuvel, J., Franke, R., Pietschmann, T., Heinz, D. W. & Collins, J. (2012). *J. Biol. Chem.* **287**, 39224–39232.
Kuhlman, P. A., Chen, R., Alcantara, J. & Szarka, S. (2009). *Bioprocess Int.* **7**, 28–38.
Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.
Lechtenberg, B. C., Murray-Rust, T. A., Johnson, D. J., Adams, T. E., Krishnaswamy, S., Camire, R. M. & Huntington, J. A. (2013). *Blood*, **122**, 2777–2783.
Liebschner, D., Dauter, M., Brzuszkiewicz, A. & Dauter, Z. (2013). *Acta Cryst.* D**69**, 1447–1462.
Lin, J., Cassidy, C. S. & Frey, P. A. (1998). *Biochemistry*, **37**, 11940–11948.
Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). *Acta Cryst.* B**39**, 480–490.
Masaki, T., Tanabe, M., Nakamura, K. & Soejima, M. (1981). *Biochim. Biophys. Acta*, **660**, 44–50.
Masaki, T., Tanaka, T., Tsunasawa, S., Sakiyama, F. & Soejima, M. (1992). *Biosci. Biotechnol. Biochem.* **56**, 1604–1607.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T. & Kraut, J. (1975). *J. Biol. Chem.* **250**, 7120–7126.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Morihara, K. & Ueno, Y. (1991). *Biotechnol. Bioeng.* **37**, 693–695.
Myllykoski, M., Raasakka, A., Han, H. & Kursula, P. (2012). *PLoS One*, **7**, e32336.
Ohnishi, Y., Yamada, T., Kurihara, K., Tanaka, I., Sakiyama, F., Masaki, T. & Niimura, N. (2013). *Biochim. Biophys. Acta*, **1834**, 1642–1647.
Papaconstantinou, M. E., Gandhi, P. S., Chen, Z., Bah, A. & Di Cera, E. (2008). *Cell. Mol. Life Sci.* **65**, 3688–3697.
Rühlmann, A., Kukla, D., Schwager, P., Bartels, K. & Huber, R. (1973). *J. Mol. Biol.* **77**, 417–436.
Rypniewski, W. R., Dambmann, C., von der Osten, C., Dauter, M. & Wilson, K. S. (1995). *Acta Cryst.* D**51**, 73–85.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Shiraki, K., Norioka, S., Li, S. & Sakiyama, F. (2002). *J. Biochem.* **131**, 213–218.
Shiraki, K., Norioka, S., Li, S., Yokota, K. & Sakiyama, F. (2002). *Eur. J. Biochem.* **269**, 4152–4158.
Silvaggi, N. R., Anderson, J. W., Brinsmade, S. R., Pratt, R. F. & Kelly, J. A. (2003). *Biochemistry*, **42**, 1199–1208.
Smith, C. A. & Rayment, I. (1996). *Biophys. J.* **70**, 1590–1602.
Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974). *Biochemistry*, **13**, 4212–4228.
Tamada, T., Kinoshita, T., Kurihara, K., Adachi, M., Ohhara, T., Imai, K., Kuroki, R. & Tada, T. (2009). *J. Am. Chem. Soc.* **131**, 11033–11040.
Tofteng, A. P., Jensen, K. J., Schäffer, L. & Hoeg-Jensen, T. (2008). *Chembiochem*, **9**, 2989–2996.
Tsu, C. A., Perona, J. J., Fletterick, R. J. & Craik, C. S. (1997). *Biochemistry*, **36**, 5393–5401.
Tsunasawa, S., Masaki, T., Hirose, M., Soejima, M. & Sakiyama, F. (1989). *J. Biol. Chem.* **264**, 3832–3839.
Tulinsky, A. & Blevins, R. A. (1987). *J. Biol. Chem.* **262**, 7737–7743.
Warkentin, M. & Thorne, R. E. (2007). *J. Struct. Funct. Genomics*, **8**, 141–144.
Wilmouth, R. C., Edman, K., Neutze, R., Wright, P. A., Clifton, I. J., Schneider, T. R., Schofield, C. J. & Hajdu, J. (2001). *Nature Struct. Biol.* **8**, 689–694.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S. & Richardson, D. C. (1999). *J. Mol. Biol.* **285**, 1711–1733.
Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C. & Adams, P. D. (2008). *Methods Mol. Biol.* **426**, 419–435.