Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

High-resolution structure of proteinase K cocrystallized with digalacturonic acid

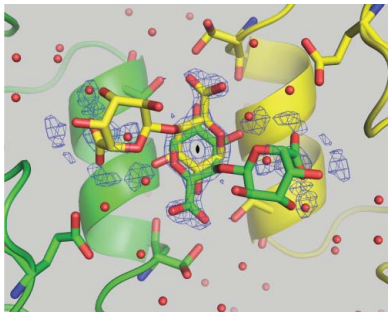
Steven B. Larson,[a] John S. Day,[a] Chieugiang Nguyen,[b] Robert Cudney[b] and Alexander McPherson[a]*

[a]Department of Molecular Biology and Biochemistry, The University of California, Irvine, CA 92697-3900, USA, and [b]Hampton Research, Aliso Viejo, CA 92656-3317, USA

Correspondence e-mail: amcphers@uci.edu

Received 12 September 2008
Accepted 19 January 2009

PDB Reference: proteinase K–digalacturonic acid complex, 3dyb, r3dybsf.



Proteinase K, a subtilisin-like fungal protease, was crystallized from a cocktail of small molecules containing digalacturonic acid (DGA). The crystal structure was determined to 1.32 Å resolution and refined to an *R* factor of 0.158. The final model contained, beside the protein, two calcium ions, 379 water molecules, a molecule of DGA and a partially occupied HEPES molecule. The DGA molecule has one sugar moiety disposed exactly on a crystallographic twofold axis; the second ring was not observed. The DGA molecule is bound to two protein molecules across the twofold axis through hydrogen-bonding networks involving Ser150 and water molecules. One of the calcium-ion sites has not been reported previously. This study further illustrates the involvement of small molecules in the crystallization of macromolecules through their ability to form intermolecular lattice interactions.

1. Introduction

We have reported in previous papers (McPherson & Cudney, 2006; Larson *et al.*, 2007, 2008) on the promotion of the crystallization of macromolecules by mixtures, or 'cocktails', of a variety of conventional small molecules. In nine cases, we were able to show (Larson *et al.*, 2007, 2008) that these small molecules mediated interactions between macromolecules in the crystals, provided lattice interactions and were visible in difference electron-density maps. In all of these cases, we found the small molecules to be well ordered in the crystals and held firmly in place by hydrogen bonds and electrostatic interactions. Here, we describe a crystalline complex in which a ligand is partially disordered but nonetheless is involved in interactions between protein molecules in the crystal lattice.

Proteinase K (endopeptidase K; EC 3.4.21.64) is a subtilisin-like serine protease with the classic catalytic triad of serine (224), histidine (69) and aspartic acid (39) at its active site. It is a fungal protease isolated from *Engyodontium album* (formerly *Tritirachium album*) and is so named because it is one of the few proteases that will degrade keratin. The enzyme is stable over a broad range of pH, temperature and salt concentration and is even active in the presence of strong detergents such as SDS. It shows fairly low specificity for its polypeptide targets. It comprises 279 amino acids and has a molecular weight of 28 900 Da. The structure was first determined crystallographically by Betzel *et al.* (1988), but there are currently 27 entries for the protein in the PDB (Berman *et al.*, 2000), including a number of complexes with peptides and protease inhibitors. Digalacturonic acid (DGA), the complex of which with proteinase K we describe here, is not an inhibitor or a known ligand of proteinase K; indeed, it does not associate with the protein using any active-site residues.

2. Materials and methods

The crystal screening experiments that yielded the complex crystals have been described in detail elsewhere (Larson *et al.*, 2008). The crystals were grown by vapor diffusion at room temperature in 96-well Intelli-Plates (Hampton Research, Aliso Viejo, California, USA) with 90 µl reservoirs of 25% PEG 3350 in water. The protein droplets were 2 µl in volume and consisted of equal parts of a 10–15 mg ml^{-1} stock protein solution and a 'cocktail' of small molecules

at concentrations of about 1%(*w*/*v*). The protein and cocktail solutions were buffered with 0.1 *M* HEPES and adjusted to pH 7.0. The concentrations of the small molecules in the final crystallization drop were approximately 5–10 m*M*. The small molecules in the cocktail were digalacturonic acid, 5,5-diphenylhydantoin and inosine-5′-monophosphate (IMP), the chemical structures of which, as well as that of HEPES, are shown in Fig. 1. Crystals usually appeared and grew to full size within one to two weeks.

The crystals were captured in loops (Hampton Research, Aliso Viejo, California, USA), flash-frozen in liquid nitrogen and mounted on a Rigaku RU200 generator producing Cu $K\alpha$ radiation and equipped with Osmic mirrors. X-ray diffraction data were collected at 100 K to 1.32 Å resolution, which was the limit of the detector. The data were integrated, processed, scaled and merged using the program *d*TREK* (Pflugrath, 1999). Although the highest resolution shell appears weak with $\langle I/\sigma(I)\rangle$ = 1.3 and only 46% of the data, we included this data in order to have as much data as possible for anisotropic thermal factor refinement. The space group, unit-cell parameters and relevant data-collection and processing statistics are presented in Table 1.

The initial model was taken from PDB entry 3prk (Wolf *et al.*, 1991; water and ligand removed), which had the same space group and a unit cell similar to that of our crystal. This model was initially refined as a rigid body with *CNS* (Brünger *et al.*, 1998) followed by simulated-annealing and temperature-factor refinement. An $F_o - F_c$ difference map calculated at R = 0.26 revealed density on a crystallographic twofold axis into which the digalacturonic acid was built and added to the model. Subsequently, the protein was rebuilt and water and disordered amino acids were added to the model periodically during the refinement and rebuilding cycles using *Coot* (Emsley & Cowtan, 2004). Since most proteinase K structures have a calcium-ion-binding site, inspection of maps at this site revealed only weak difference density. A full-occupancy calcium ion was added to the model which was adjusted to a partially occupied site towards the end of refinement. Furthermore, inspection of the water structure that had been built up during the course of refinement suggested the presence of an additional calcium ion (which had been modeled as a water molecule) and a partially occupied HEPES site. The water at the new calcium site was changed to a calcium ion and the HEPES molecule was built into difference density. Partial occupancies were determined by refining multiple models with variable occupancy ratios and comparing the R_{free} and B factors of equivalent atoms of the disordered groups, with the final occupancy determined by the ratio that gave similar B factors for equivalent atoms and the lowest R_{free} values. The final model was refined using *REFMAC*5 (Murshudov *et al.*, 1997) with H atoms included and anisotropic thermal parameters for nearly all non-H atoms. References to B factors throughout this text refers to equivalent isotropic thermal factors (B_{eq}) obtained as the trace of the anisotropic thermal parameters (see footnote to Table 1). The R and R_{free} for the incomplete model before anisotropic refinement were 0.238 and 0.274, respectively. The final refinement statistics are shown in Table 1.

$F_o - F_c$ difference Fourier maps were computed with *CNS* or from MTZ files from the *REFMAC*5 refinements and were displayed with the programs *O* (Jones & Kjeldgaard, 1994), *Coot* or *PyMOL* (DeLano, 2002). The latter program was used to produce figures; the program *LIGPLOT* (Wallace *et al.*, 1995) was used to aid in identifying intermolecular interactions.

3. Results

The final model contains 279 residues, two putative calcium ions, a partially occupied HEPES molecule, a twofold disordered single ring of a digalacturonic acid molecule and 379 full or partially occupied water molecules. There were 17 residues that were modeled with

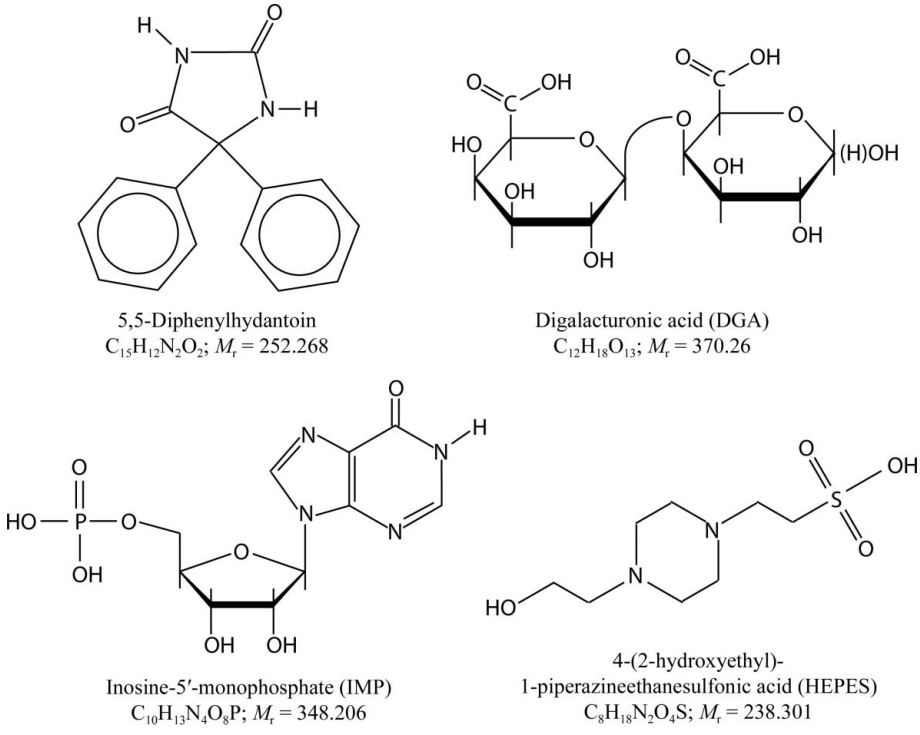


Figure 1
Schematic structures of the three components of the small-molecule 'cocktail' used in this study and the HEPES that served as the crystallization buffer and was subsequently located in a cleft of the proteinase K structure.

multiple conformations. These were Thr4, Asn5, Ser15, Gln31, Ser63, Lys87, Gln89, Ser143, Met154, Asn162, Arg185, Ser197, Ser216, Ser247, Arg250, Ser262 and Asn276. No residues were in disallowed regions of the Ramachandran plot (*PROCHECK*; Laskowski *et al.*, 1993). The overall r.m.s. deviations for bond lengths and angles were 0.011 Å and 1.27°, respectively. The final R and $R_{\rm free}$ were 0.158 and 0.207, respectively, using data to 1.32 Å.

As noted above, inspection of an $F_o - F_c$ difference electron-density map calculated after refinement of the model containing only the protein revealed ring-shaped density that was centered on a crystallographic twofold axis. This density is shown in Fig. 2. Since none of the small molecules in the cocktail possess twofold symmetry, this immediately suggested the presence of disorder in the ligand. Neither 5,5-diphenylhydantoin nor IMP, by virtue of their sizes and shapes, were compatible with the density, even assuming multiple orientations. However, one ring of a digalacturonic acid molecule could be easily placed in this density such that its twofold-related position would fill the entire density, as seen in Fig. 2. DGA lies on the surface of two twofold-related proteinase K molecules. Its $\langle B \rangle$ was 24.7 Å^2. Ser150 of one molecule interacts with O2′ of DGA (d = 2.96 Å) and Ser150 of the other proteinase K molecule interacts with O62′ of DGA (d = 2.71 Å). Ser143 OG is disordered, its position being determined by the positioning of the carboxylate group of DGA. When the carboxylate group is near, conformer *A* of Ser143 OG is only 2.38 Å away from O62′ of DGA. This may be too close and therefore the OG atom swings away from the carboxylate as in conformer *B* (Fig. 2). When DGA atom O2′ is near, conformer *A* of Ser143 has no potential steric clashes with DGA and in fact can possibly hydrogen bond to Ser150 of the twofold-related proteinase K. The occupancy of the DGA is fixed by the twofold crystallographic disorder at 50% and the occupancies of the Ser143 conformers were fixed at 50% by their correlation to DGA. The water network between DGA and proteinase K is through the carboxylate of DGA and the carboxylate of Glu174 and conformer *B* of Ser143. Analysis of the other proteinase K structures revealed that in seven structures there was a disordered Ser143, all of which only had water in the vicinity.

While inspecting the water structure around Arg185, it was observed that it inadequately satisfied the density. Removal of the water and inspection of a difference map suggested the presence of a sulfate-like molecule interacting with Arg185. Of the known components of the crystallization mixture, IMP and HEPES possess such a moiety. However, the entire difference density could only be fitted with a HEPES molecule, with the sulfonic acid moiety hydrogen bonded to Arg185 through two O atoms (O1S and O3S) and the hydroxyl group at the opposite end hydrogen bonded to the hydroxyl group of Tyr82. This is illustrated in Fig. 3(*a*). In addition, atom O2S of the sulfonic acid moiety is hydrogen bonded to the peptide N atom of Trp8. However, the *B* factor of O2S was considerably lower than the *B* factors of O1S and O3S of the sulfonic acid group. This, coupled with the residual density observed in subsequent difference maps, suggested that the HEPES site was only partially occupied. Therefore, ten equivalent models differing only in the occupancy of the HEPES molecule (the occupancies ranged from 0.1 to 1.0 in increments of 0.1) were refined. The occupancy that produced the lowest $R_{\rm free}$ was 60%. The $\langle B \rangle$ for this ligand was 33.9 Å^2. As noted above, there was considerable residual difference density around the HEPES molecule which we interpreted to be an overlying partially occupied solvent network, as illustrated in Fig. 3(*b*). Except for water 755, which shares its position with atom O2S of the sulfonic acid moiety of HEPES, we have only modeled the solvent surrounding the HEPES site, not within the HEPES site. Arg185 was modeled with a disordered side chain prior to inclusion of the HEPES. The sulfonic acid interactions may be more favorable with the *B* conformation of Arg185 (Fig. 3*a*), but because there is good interaction either way no correlation between the occupancies of Arg185 and HEPES was made. The Arg185 side-chain disorder is 55:45.

In addition to the aforementioned ligands, the model includes two calcium ions, although one appears to be only partially occupied. Calcium ion Ca300 is consistent with the calcium present in 23 previously reported proteinase K structures; this fact prompted us to inspect this site. Although density for this site appeared in difference maps (see Fig. 4), when a Ca ion placed in this density was refined at full occupancy it coalesced with an adjacent water molecule (608) and the *B* factor increased to nearly 200 Å^2. Refinement of equivalent models with occupancies for Ca300 ranging from 0 to 1 in increments of 0.1 suggested the occupancy to be as low as 0.1 based on the best combination of $R_{\rm free}$ and coordination distances. Ca300 has a distorted octahedral coordination involving interactions with the carbonyl O atoms of Pro175 and Val177, the carboxyl O atoms of Asp200 and waters 507, 594, 608 and 693. The distances range from 2.29 to 2.63 Å. Water 693, at a distance of 2.63 Å from Ca300, may not in fact contribute to the coordination. However, it should be noted that the calcium ions in 21 of the 23 structures have a very similar coordination sphere. Using the CBVS method (Müller *et al.*, 2003) for identifying metal-ion-binding sites, which is based on comparison of

Table 1
Data-processing, refinement and model statistics.

Values in parentheses are for the highest resolution shell.

Crystal data	
Space group	$P4_32_12$
Z	8
Unit-cell parameters (Å)	$a = b = 67.72$, $c = 101.89$
Data-processing statistics	
Resolution (Å)	32.14–1.32 (1.37–1.32)
No. of unique reflections	53100 (2566)
Redundancy	5.33 (1.50)
$\langle I/\sigma(I) \rangle$	11.5 (1.3)
Completeness	94.0 (46.2)
$R_{\rm merge}$†	0.066 (0.482)
Structure-refinement statistics	
Resolution (Å)	30.36–1.32 (1.35–1.32)
No. of reflections (F_o = 0 removed)	49454 (1186)
$R/R_{\rm free}$‡ (all data)	0.158/0.207 (0.455/0.476)
Reflections in test set (%)	5.10 (5.65)
No. of refined parameters	22698
No. of reflections	46808
No. of restraints	38966
Data-to-parameter ratio	2.06
Data restraints-to-parameter ratio	3.78
Model statistics	
Non-H atoms	
Protein atoms (full/partial)	1956/157
Ligand atoms (full/partial)	1/28
Water atoms (full/partial)	273/106
Geometry: r.m.s. deviations from ideal values	
Bonds (Å)	0.011
Angles (°)	1.265
Planes (Å)	0.01
Chiral centers (Å^3)	0.078
Average isotropic $B_{\rm eq}$ factors§ (Å^2)	
Overall	15.88
Protein (full/partial)	14.2/17.4
Ligands (all)	29.8
Water (full/partial)	32.3/21.5
Ramachandran plot	
Most favored region	213 [90.6%]
Allowed region	22 [9.4%]
Generously allowed and disallowed regions	0 [0.0%]

† $R_{\rm merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith used observation for unique hkl and $\langle I(hkl) \rangle$ is the mean intensity for unique hkl. ‡ $R = \sum_h |F_o - F_c| / \sum_h F_o$, where F_o and F_c are the observed and calculated structure-factor amplitudes, respectively. § $B_{\rm eq} = (8\pi^2 \sum_i U_{ii})/3$, for i = 1, 2, 3.

valence bonds of any metal-binding site with that of calcium, suggests that this is indeed a calcium-ion-binding site. In this method, the CBVS value for calcium would be 2.0, that for sodium 1.57 and that for potassium 0.64. The CBVS value is 2.15 for seven calcium ligands and 1.98 without water 693 in the coordination sphere. Because of the partial occupancy of this site, the water structure surrounding it has been modeled with disorder for waters 507 and 608 (see Fig. 4), the direction along which the Ca300 migrates when refined at full occupancy. The *B* factors of Ca300 and the coordinating waters 507(*A*), 594, 608(*A*) and 693 are 6.8, 7.6, 5.5, 39.8 and 50.0 Å^2, respectively. The *B* factors of the coordinating atoms of the protein, Pro175 O, Val177 O and Asp200 OD2, are 13.3, 15.7 and 12.6 Å^2, respectively. The noncoordinating disordered waters 507(*B*) and 608(*B*) have *B* factors of 19.1 and 26.5 Å^2, respectively. These *B* values suggest that the occupancy of Ca300 may be low but is considerably less than 100%. It should be noted that occupancies greater than 0.1 produced negative peaks in difference density maps at the Ca300 site.

The other calcium site, Ca301, involving Tyr195 and Asn163 has not been reported previously, although there are recent high-resolution structures (better than 1.08 Å) of proteinase K in the PDB

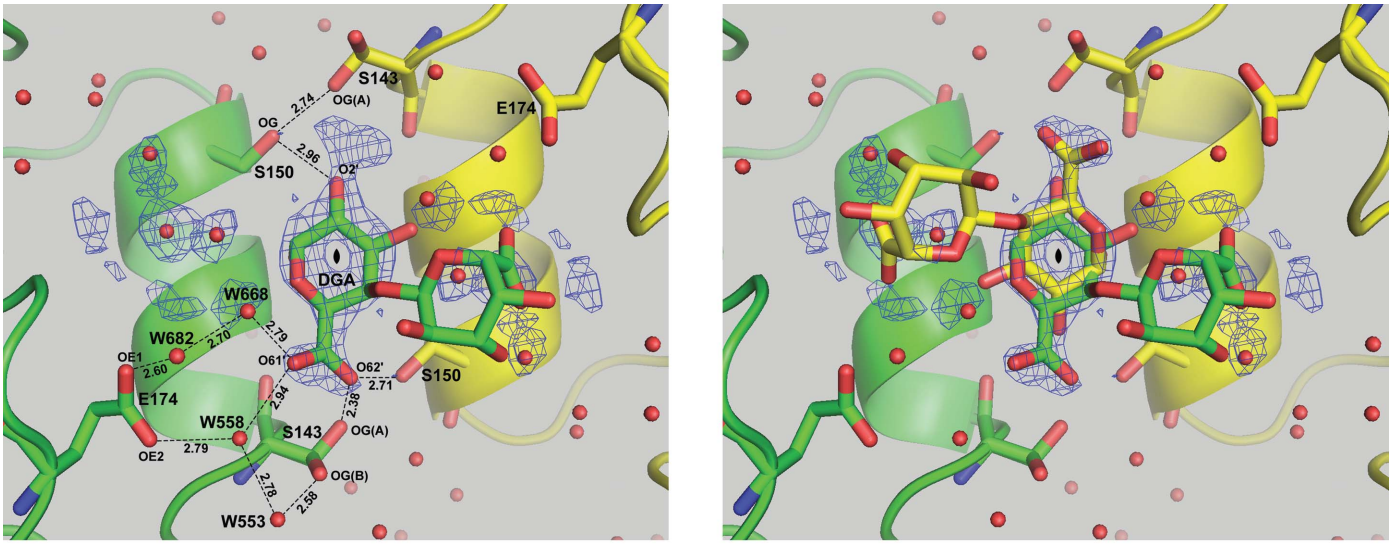


Figure 2
View approximately down the crystallographic twofold axis. The primary final model is in green and the twofold-related structure is in yellow. Although the second ring of the digalacturonic acid molecule is shown in the figure, it is placed in a stereochemically reasonable but arbitrary position since there was no difference density for this ring. The $F_o - F_c$ difference density shown here, which has been isolated around the final refined model of DGA, came from the map into which DGA was first built. The map was calculated at 1.5 Å resolution, is contoured at 1.5σ and is based on the preliminary refinement of proteinase K to an *R* factor of 0.26 without any small molecules in the model. In (*a*) the hydrogen-bonded interactions, both direct and water-mediated, between DGA and proteinase K are shown with distances. The twofold axis is indicated at the center of the ring of density. Only residues that are involved in these interactions are shown, *i.e.* Ser143, Ser150 and, indirectly, Glu174. (*b*) shows the twofold-related DGA and how the two orientations relate to each other about the twofold axis.

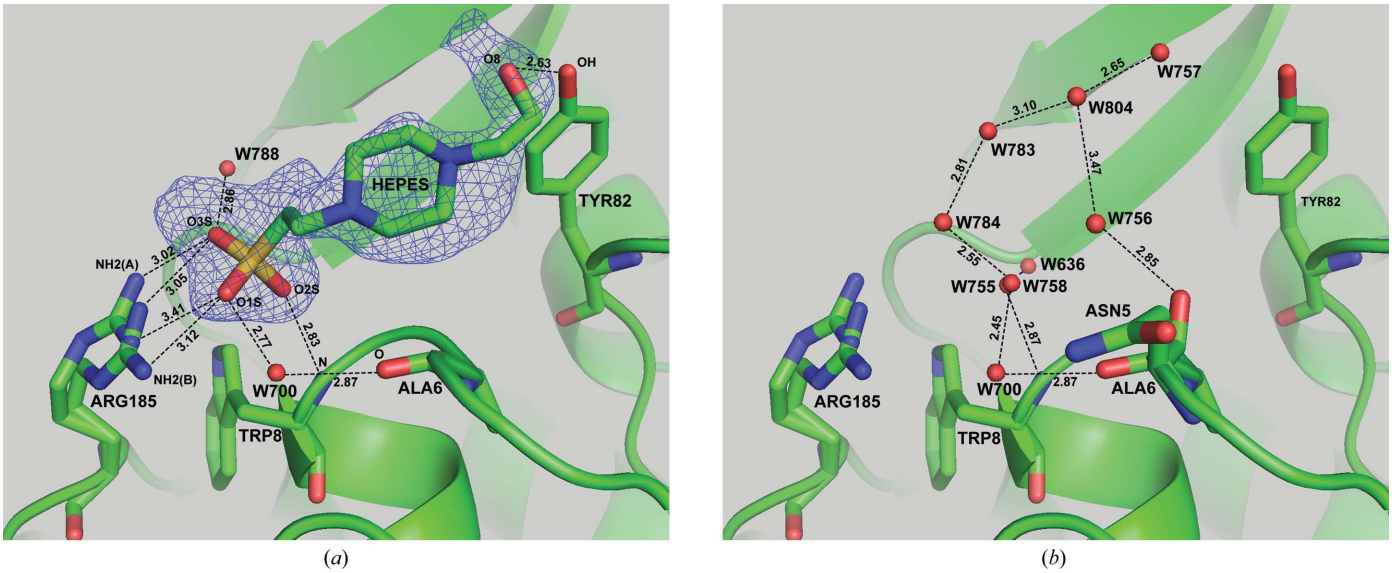


Figure 3
View in (*a*) of the HEPES molecule in a cleft on the surface of proteinase K. The sulfonic acid moiety of the HEPES molecule is hydrogen bonded to Arg185 (which is disordered) and the peptide N atom of Trp8. The hydroxyl group at the opposite end of the HEPES molecule is hydrogen bonded to the hydroxyl group of Tyr82. The HEPES molecule has an occupancy of 60% and shares the site with a solvent network shown in (*b*). The contour level of the HEPES-omitted $F_o - F_c$ difference density is 1.5σ. Possible hydrogen-bonded interactions with distances are shown as dashed lines.

that contain an octahedrally coordinated water molecule involving these two residues [PDB codes 1ic6 (Betzel *et al.*, 2001), 1p7v (S. Bilgrami, P. Kaur, V. Chandra, S. Banumathi, M. Perbandt, C. Betzel & T. P. Singh, unpublished work) and 2pwa (R. Jain, N. Singh, M. Perbandt, C. Betzel, S. Sharma, P. Kaur, A. Srinivasan & T. P. Singh, unpublished work)]. This calcium is coordinated by the hydroxyl O atom of the tyrosine and the carbonyl O atom of the asparagine, with waters 569, 638, 650 and 654 completing the octahedral coordination at distances ranging from 2.29 to 2.43 Å. The CBVS value for this site is 2.02, a very strong suggestion that it is a calcium-binding site. The *B* factors of Ca301 and its coordinating atoms Asn163 O, Tyr195 OH and waters 569, 638, 650 and 654 are 33.1, 14.3, 13.0, 22.8, 22.1, 27.3 and 25.4 Å^2, respectively. The *B* factor of 33.1 Å^2 for Ca301 is higher than those of the surrounding atoms, but not significantly higher. This calcium site is illustrated in Fig. 5.

All 17 discretely disordered residues, except Met154, are on the surface of the protein. Only Ser143 and Arg185 make interactions with ligands. Asn5 is in close proximity to the HEPES molecule but does not appear to make any direct interactions with it. Eight disordered side chains make no direct interactions with other protein atoms, while seven others have interactions through at least one of the conformers. Finally, the *B* conformers of Gln31 and Lys87 appear to be correlated with each other through a hydrogen bond. Of the 17 disordered residues, only three have not previously been observed to be disordered.

Finally, it should be noted that weak difference density which was not modeled was observed in a cavity between two strands, Leu133-Gly134-Gly135 on one side and Ala159-Gly160 on the other, and in close proximity to the active-site residues. This density is illustrated in Fig. 6(*a*). The density is fairly planar and could possibly be the inosine ring of a molecule of IMP, another component of the small-molecule cocktail. After real-space fitting of an inosine ring into this density using *Coot*, the average distance from the inosine plane to the nearly planar group of five backbone atoms of residues Leu133 (CA, C and O) and Gly134 (N and CA) is 3.4 Å and the interaction appears to be aromatic–π bond stacking. The minimum distance to the inosine plane for any atom of the Ala159-Gly160 segment is 3.37 Å for Gly160 CA. There does not appear to be any direct hydrogen bonding to the protein. How the inosine ring of IMP might fit into this density is illustrated in Fig. 6(*b*). Clearly, if this is IMP it is only partially occupied and the ribose and phosphate moieties are disordered.

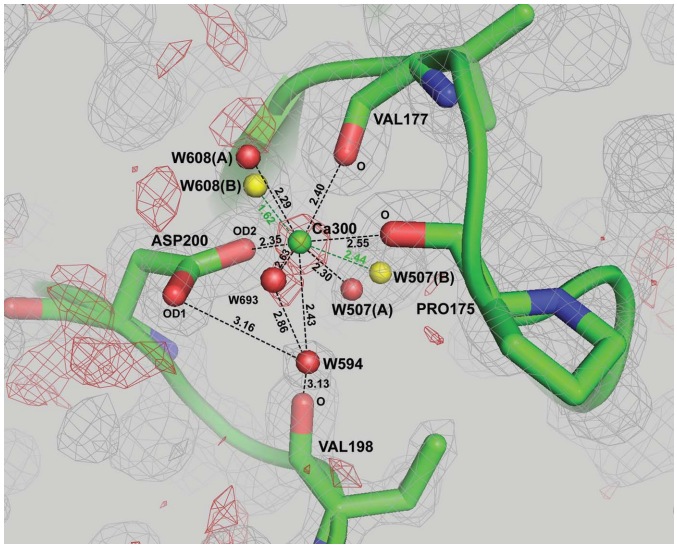


Figure 4
The putative calcium-binding site Ca300. The estimated occupancy is 10%. The removal of atom Ca300 from the model produced $F_o - F_c$ difference density encompassing the atom, shown here in red and contoured at 2.5σ (the peak maximum is 5σ). The gray density is the $2F_o - F_c$ density contoured at 1.0σ. Ca300 is either six-coordinate or seven-coordinate depending on whether water 693, at a distance of 2.63 Å from Ca300, is considered to be part of the coordination.

4. Discussion

Prior to the deposition of this structure in the PDB, 27 proteinase K structures had been deposited, 26 of which were reported as possessing space-group symmetry $P4_32_12$. The lone variant was PDB entry 1bjr (Singh *et al.*, 1998): its space group was $P2_1$ and it was crystallized from a Tris–HCl-buffered ethanol solution. In the former group, the unit cells fall into two subgroups depending on the length of the *c* axis. The average length of the *a* axis for all 27 $P4_32_12$ structures (including that reported here) is 68.1 (3) Å. One subgroup of this set has an average *c* axis of 107.9 (7) Å, while the other subgroup, which includes our structure, has an average of 101.9 (4) Å. The resolutions for all structures range from 0.83 to 2.90 Å. The range of average *B* values over all protein atoms of all 28 proteinase K structures is 5.7–18.0 Å^2 and the comparable range of average *B* values for side-chain atoms is 6.5–20.9 Å^2. These values are a tribute to the tremendous secondary and tertiary stability of proteinase K. Comparison of our protein structure with those previously deposited show a range in the r.m.s. deviations of the backbone atoms (N, C$^\alpha$ and C) of 0.12–0.42 Å, with the largest C$^\alpha$ deviation being 2.38 Å for C$^\alpha$ of Gln278 from the $P2_1$ structure; for the $P4_32_12$ structures the largest C$^\alpha$ deviation is 1.77 Å for Ala279, the C-terminal residue. In fact, the largest C$^\alpha$ deviation occurs at this residue in 18 of the 27 structures, demonstrating the slight increase in flexibility at the C-terminus. In six of the 27 structures, the largest C$^\alpha$ deviation occurs at Asn120 with a range of 1.0–1.3 Å. This residue is in an exterior loop.

13 of the 27 deposited structures were crystallized from solutions containing $CaCl_2$ and $NaNO_3$ as the precipitant with a buffer of Tris–HCl in most cases. Only three of the 27 deposited structures utilized

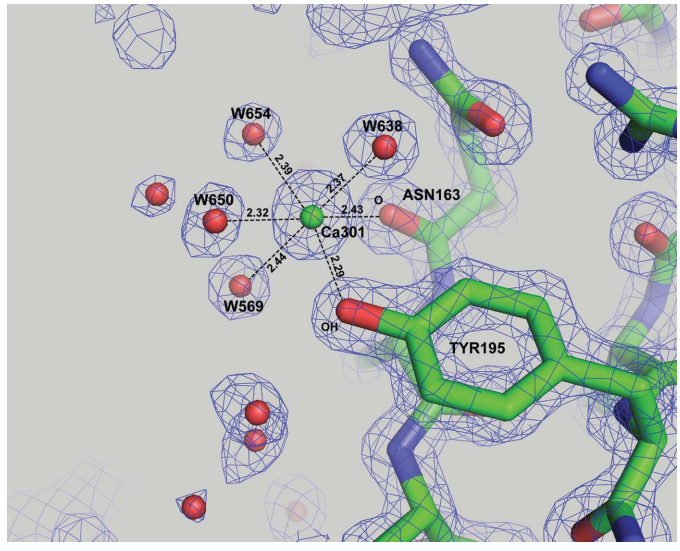
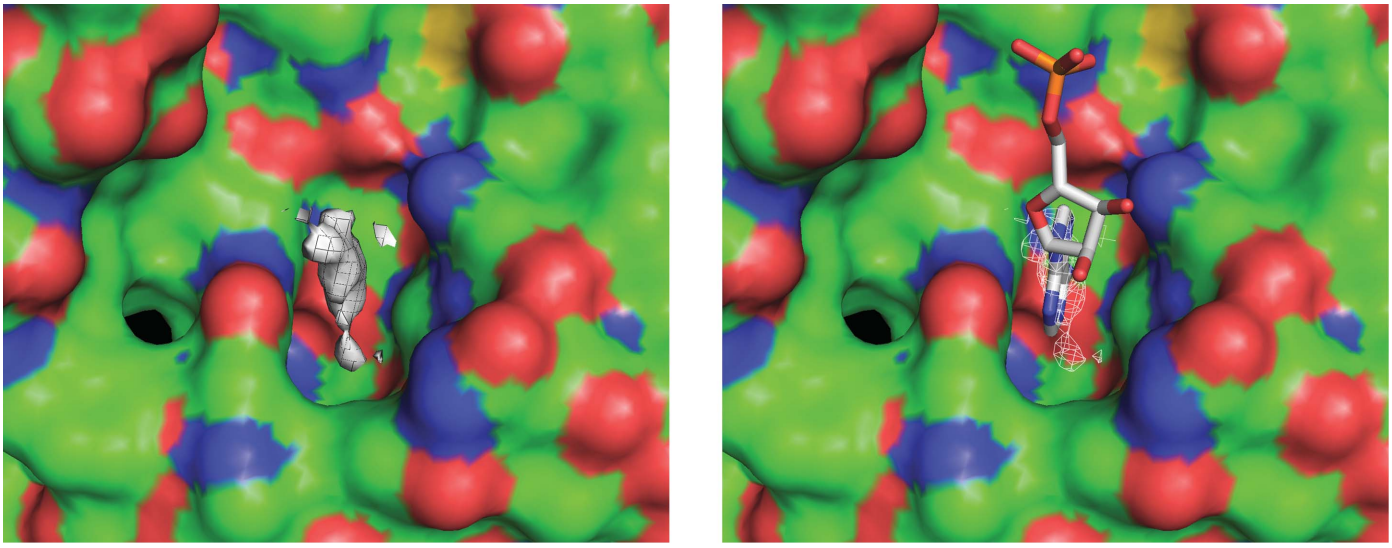


Figure 5
The octahedral coordination of the putative calcium-binding site Ca301 is illustrated. The site involves the carbonyl O atom of Asn163, the hydroxyl O atom of Tyr195 and four water molecules with calcium–oxygen distances in the range 2.29–2.43 Å. The contour level of the $2F_o - F_c$ density is 2.0σ.



(*a*) (*b*)

Figure 6
(*a*) Weak unidentified difference density is shown in white inside a cavity of proteinase K near the active-site residues. This density could possibly result from the inosine ring of a partially occupied inosine-5′-monophosphate molecule as shown in (*b*). The difference density is contoured at 2.0σ.

PEG 4000 as the precipitant and an acetate buffer; none of these reported the addition of $CaCl_2$. Two of the three have a calcium at the Ca300 position and the same two have a water at the Ca301 position without the octahedral environment found in structures 1ic6, 1p7v, 2pwa and 1p7w (S. Bilgrami, M. Perbandt, V. Chandra, S. Banumathi, P. Kaur, C. Betzel & T. P. Singh, unpublished work). The structure 2g4v (Mueller-Dieckmann *et al.*, 2007) is the only structure in which HEPES was used as the buffer, but unlike our structure there was no reported HEPES bound to the protein. The precipitant in this case was sodium/potassium tartrate. Structure 2g4v, resulting from an investigation into the identification of ions by anomalous scattering using low-energy radiation, was reported to have a disordered Ca^{2+} ion at/near the position of Ca300 and a partially occupied K^+ ion at the position of Ca301. However, the coordination spheres (distances and number of coordinating atoms) of the minor Ca^{2+} position or the K^+ ion do not convincingly support these assignments. In addition, there was one structure (2pkc; Müller *et al.*, 1994) in which the Ca300 position was occupied by a hexacoordinate Na^+ ion and another (1pfg; Saxena *et al.*, 1996) in which the site is modeled as a pentacoordinate water molecule, while two structures [1pek (Betzel *et al.*, 1993) and 1pj8 (Saxena *et al.*, 1996)] have a full coordination sphere without any ion at the site. Only the lowest resolution (2.9 Å) model, 2dp4 (A. K. Singh, N. Singh, S. Sharma, A. Bhushan & T. P. Singh, unpublished work), has no ion or any water structure at this site.

There have been three other Ca^{2+}-ion-binding sites reported in the deposited structures. In seven structures, there is a site utilizing the carbonyl O atom of Thr16 and the carboxyl O atom OD2 of Asp260. All of these models list the occupancy as 50% or less and only two have a complete octahedral coordination sphere. However, at least ten other proteinase K structures have a water at this site with a full or partial octahedral coordination sphere. Four of those with full coordination have CBVS values in the range 1.90–2.08, which suggests the presence of calcium. This calcium-binding site and the CA300 site are discussed in great detail by Müller *et al.* (1994). We observe no density in difference density maps at this position.

In 1bjr, there is a Ca^{2+} ion with 75% occupancy involved with O atoms of Arg12, Ser15, Asn257 and Ala273. However, the Ca—O distances are 2.77–3.28 Å and the positively charged guanadinium group of Arg12 is only 3.55 Å from this putative calcium position, suggesting that water is the more likely occupant. Furthermore, the coordination does not seem to be appropriate for calcium. Finally, structures 1p7v and 1p7w have a fully occupied Ca^{2+} ion bound to the hydroxyl O atom of Thr17. We have a water molecule at each of these last two positions.

The sulfonic acid moiety of the HEPES molecule also has a precedent in proteinase K structures. The structure 2v8b, in which sodium selenate ($NaSeO_4$) was used to introduce anomalous scatterers into the structure, has SeO_4^{2-} at the sulfonic acid position. Like the HEPES, the reported occupancy for this ion was 60%.

As noted above, the unidentified density lies in a cavity near the active site. Many of the structures of proteinase K that have been deposited were determined as investigations of the active site or of proteinase K inhibitors bound at the active site. Residues 131–135 are involved in forming β-strand structures with amino-acid-containing inhibitors. When superimposed on our model, many structures have small molecules of polypeptides that lie in this density or near this cavity. Specifically, this is seen for the phenyl rings of polypeptide inhibitors of structures 1p2w, 1p7v, 1pek and 1pfg, and for the alanine boronic acid of structure 2pwa.

As in previous investigations (Larson *et al.*, 2007, 2008) small-molecule components of the crystallization 'cocktails' were found as integral components of the crystals and again were involved in macromolecular interfaces in the lattice. Thus, they provided additional lattice contacts and possibly stabilizing interactions. The fact that proteinase K crystallizes predominantly in the observed space group without the aid of a ligand at the twofold interface suggests that the interactions with DGA are adventitious, resulting primarily from the well matched dimensions of the DGA molecule to the separation of the twofold-related Ser150 residues seen in Fig. 2.

This work was supported by NIH Grant GM074899 for the establishment of the Center for High Throughput Structural Biology.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Betzel, C., Gourinath, S., Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S. & Singh, T. P. (2001). *Biochemistry*, **40**, 3080–3088.
Betzel, Ch., Pal, G. P. & Saenger, W. (1988). *Acta Cryst.* B**44**, 163–172.
Betzel, C., Singh, T. P., Visanji, M., Peters, K., Fittkau, S., Saenger, W. & Wilson, K. S. (1993). *J. Biol. Chem.* **268**, 15854–15858.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
DeLano, W. L. (2002). *The PyMOL Molecular Visualization System.* DeLano Scientific, San Carlos, California, USA. http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Jones, T. A. & Kjeldgaard, M. (1994). *O – The Manual*, v.5.10. Uppsala University Press.
Larson, S. B., Day, J. S., Cudney, R. & McPherson, A. (2007). *Acta Cryst.* D**63**, 310–318.
Larson, S. B., Day, J. S., Nguyen, C., Cudney, R. & McPherson, A. (2008). *Cryst. Growth Des.* **8**, 3038–3052.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
McPherson, A. & Cudney, B. (2006). *J. Struct. Biol.* **156**, 387–406.
Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Wilmanns, M., Singh, R. K., Tucker, P. A. & Weiss, M. S. (2007). *Acta Cryst.* D**63**, 366–380.
Müller, A., Hinrichs, W., Wolf, W. M. & Saenger, W. (1994). *J. Biol. Chem.* **269**, 23108–23111.
Müller, P., Köpke, S. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 32–37.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Saxena, A. K., Singh, T. P., Peters, K., Fittkau, S. & Betzel, C. (1996). *Protein Sci.* **5**, 2453–2458.
Saxena, A. K., Singh, T. P., Peters, K., Fittkau, S., Visanji, M., Wilson, K. S. & Betzel, C. (1996). *Proteins*, **25**, 195–201.
Singh, T. P., Sharma, S., Karthikeyan, S., Betzel, C. & Bhatia, K. L. (1998). *Proteins*, **33**, 30–38.
Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng.* **8**, 127–134.
Wolf, W. M., Bajorath, J., Müller, A., Raghunathan, S., Singh, T. P., Hinrichs, W. & Saenger, W. (1991). *J. Biol. Chem.* **266**, 17695–17699.