VIROLOGY JOURNAL

Crystal structure of the lytic CHAP$_K$ domain of the endolysin LysK from *Staphylococcus aureus* bacteriophage K

Sanz-Gaitero *et al.*

RESEARCH   **Open Access**

Crystal structure of the lytic CHAP$_K$ domain of the endolysin LysK from *Staphylococcus aureus* bacteriophage K

Marta Sanz-Gaitero[1], Ruth Keary[2], Carmela Garcia-Doval[1,3], Aidan Coffey[2] and Mark J van Raaij[1*]

Abstract

Background: Bacteriophages encode endolysins to lyse their host cell and allow escape of their progeny. Endolysins are also active against Gram-positive bacteria when applied from the outside and are thus attractive anti-bacterial agents. LysK, an endolysin from staphylococcal phage K, contains an N-terminal cysteine-histidine dependent amido-hydrolase/peptidase domain (CHAP$_K$), a central amidase domain and a C-terminal SH3b cell wall-binding domain. CHAP$_K$ cleaves bacterial peptidoglycan between the tetra-peptide stem and the penta-glycine bridge.

Methods: The CHAP$_K$ domain of LysK was crystallized and high-resolution diffraction data was collected both from a native protein crystal and a methylmercury chloride derivatized crystal. The anomalous signal contained in the derivative data allowed the location of heavy atom sites and phase determination. The resulting structures were completed, refined and analyzed. The presence of calcium and zinc ions in the structure was confirmed by X-ray fluorescence emission spectroscopy. Zymogram analysis was performed on the enzyme and selected site-directed mutants.

Results: The structure of CHAP$_K$ revealed a papain-like topology with a hydrophobic cleft, where the catalytic triad is located. Ordered buffer molecules present in this groove may mimic the peptidoglycan substrate. When compared to previously solved CHAP domains, CHAP$_K$ contains an additional lobe in its N-terminal domain, with a structural calcium ion, coordinated by residues Asp45, Asp47, Tyr49, His51 and Asp56. The presence of a zinc ion in the active site was also apparent, coordinated by the catalytic residue Cys54 and a possible substrate analogue. Site-directed mutagenesis was used to demonstrate that residues involved in calcium binding and of the proposed active site were important for enzyme activity.

Conclusions: The high-resolution structure of the CHAP$_K$ domain of LysK was determined, suggesting the location of the active site, the substrate-binding groove and revealing the presence of a structurally important calcium ion. A zinc ion was found more loosely bound. Based on the structure, we propose a possible reaction mechanism. Future studies will be aimed at co-crystallizing CHAP$_K$ with substrate analogues and elucidating its role in the complete LysK protein. This, in turn, may lead to the design of site-directed mutants with altered activity or substrate specificity.

Keywords: Bacteriophage, Calcium, Crystallography, Endolysin, Peptidoglycan, Protease, *Staphylococcus*, Zinc

* Correspondence: mjvanraaij@cnb.csic.es
[1]Departamento de Estructura de Macromoleculas, Centro Nacional de Biotecnologia (CNB–CSIC), Calle Darwin 3, E-28049 Madrid, Spain
Full list of author information is available at the end of the article

Spanish abstract

Introducción: Los bacteriófagos codifican endolisinas para lisar sus bacterias hospedadoras y permitir la liberación de su progenie. Las endolisinas también son activas contra bacterias Gram positivas cuando se aplican desde el exterior, y por lo tanto, son consideradas agentes antibacterianos atractivos. LysK, una endolisina del fago K que infecta estafilococos, contiene un dominio N-terminal amidohidrolasa/peptidasa dependiente de cisteína e histidina (CHAP$_K$), un dominio amidasa central y un dominio C-terminal SH3b de unión a la pared bacteriana. CHAP$_K$ corta el peptidoglicano bacteriano entre el tetrapéptido y los puentes pentaglicina.

Métodos: El dominio CHAP$_K$ de LysK fue cristalizado y se obtuvieron datos de difracción a alta resolución tanto de un cristal de proteína nativo como de un cristal derivado con cloruro de metilmercurio. La señal anómala presente en los datos derivados permitió la localización de la posición de los átomos pesados y la determinación de la fase. Las estructuras resultantes se completaron, refinaron y analizaron. La presencia de iones de calcio y zinc en la estructura fue confirmada por espectroscopía de emisión de fluorescencia de rayos X. Se llevaron a cabo análisis de zimograma sobre la enzima nativa y sobre mutantes puntuales seleccionados.

Resultados: La estructura de CHAP$_K$ reveló una topología tipo papaína con un bolsillo hidrofóbico donde se localiza la tríada catalítica. Moléculas de tampón ordenadas presentes en este hueco pueden mimetizar el substrato de peptidoglicano. Cuando se compara con dominios CHAP resueltos previamente, CHAP$_K$ contiene un lóbulo adicional en su dominio N-terminal, con un ión de calcio estructural, coordinado por los residuos Asp56, Asp45, Asp47. También se observa la presencia de un ión de zinc en el centro activo, coordinado con el residuo catalítico Cys54 y un posible análogo del substrato. Se usó mutagénesis dirigida para demostrar que los residuos involucrados en la unión a calcio y los presentes en el centro activo propuesto eran importantes para la actividad enzimática.

Conclusiones: Se determinó la estructura del dominio CHAP$_K$ de LysK a alta resolución, sugiriendo la localización del centro activo, del bolsillo de unión al sustrato y revelando la presencia de un ión de calcio estructuralmente importante. Se encontró un ión de zinc unido más débilmente. Basándonos en la estructura, proponemos un posible mecanismo de reacción. Futuros estudios tendrán por objeto la cristalización de CHAP$_K$ con análogos del sustrato y la elucidación de su papel en la proteína LysK completa. Esto, a su vez, podría conducir al diseño de mutantes puntuales con una actividad o especificidad de sustrato modificada.

Background

Bacteriophage K is a virulent phage that infects a wide range of staphylococci. It belongs to the *Myoviridae* family of the *Caudovirales* order, with a genome of 148,317 bp [1-3]. To allow its progeny to escape from the host cell ("lysis from within"), it encodes the endolysin LysK, a peptidoglycan hydrolase [4]. When applied exogenously to the pathogen, LysK causes "lysis from without" or exolysis [5]. Gram-positive endolysins are highly specific [4], and no bacterial variants resistant to their phage endolysins have been found despite the use of mutagenesis strategies to promote the chance of resistance development [6]. LysK kills a wide range of staphylococci, including multi-drug-resistant *Staphylococcus aureus* (MRSA) [7].

LysK contains three domains: an N-terminal cysteine-histidine dependent amido-hydrolase/peptidase (CHAP) domain, a central amidase domain and a C-terminal SH3b cell wall-binding domain. The LysK amidase domain cleaves peptidoglycan between N-acetylmuramic acid and L-alanine of the stem peptide, while the CHAP domain hydrolyzes it between the D-alanine of the tetra-peptide stem and the first glycine of the penta-glycine cross-bridge [8]. A truncated enzyme called CHAP$_K$, containing only the first 165 amino acids of LysK corresponding to the CHAP domain, also showed exolytic activity [9]. CHAP$_K$ is able to lyse several staphyloccocal species, independently from their origin, their antibiotic resistance profile and their ability to produce exopolysaccharides (associated with biofilm formation) [10,11]. It is also effective against other related genera, such as *Micrococcus* or *Streptococcus* [7].

In order to understand the reaction mechanism and perhaps improve or alter the activity, we set out to solve the structure of CHAP$_K$. The CHAP$_K$ domain was expressed in *Escherichia coli*, purified and crystallized. Although the crystallization procedure was not very reproducible and crystals grew as inter-grown plates, a high-resolution dataset could be collected from one of them, plus a dataset from a methylmercury chloride derivative of sufficient quality for structure solution by single-wavelength anomalous dispersion [12]. This structure was refined against both the native and the derivative dataset. Here we present the high-resolution structure of the CHAP$_K$ domain solved by X-ray crystallography.

Results and discussion

Overall structure

The final models of the CHAP$_K$ enzyme contain amino acids 2–165 for each of the four protein molecules present in the crystallographic asymmetric units, with good crystallographic statistics and reasonable protein geometry (Table 1). The models also contain metal ions, waters and other solvent molecules. For the native structure, a calcium ion, a zinc ion and a 2-(N-morpholino) ethanesulfonic acid (MES) molecule have been modelled associated with each of the protein chains, as discussed below. Other ordered solvent molecules have also been modelled in the asymmetric unit and consist of one glycerol molecule, four putative sodium ions and 741 water molecules. For the derivative structure, a calcium ion and a 2-[4-(2-hydroxyethyl)piperazin-1-yl] ethane-sulfonic acid (HEPES) molecule have been modelled associated with each of the protein chains, while Cys54 is modelled as methylmercury-cysteine. In this case, ordered solvent molecules modelled in the unit cell include two glycerol molecules, ten additional putative

methylmercury ions, two putative chloride ions and 770 waters. Despite the lower nominal resolution of the native dataset when compared with the derivative (1.8 vs. 1.7 Å), the general structural analyses described below are done using the structure refined against the native dataset, as that dataset is more complete (97.2 vs. 64.7%), contains more measured reflections (62028 vs. 48498) [10], and better maps with less non-interpretable noise peaks were obtained.

The four CHAP$_K$ monomers do not form extensive inter-monomer interfaces in the crystal, suggesting that in solution the protein is monomeric. When the four crystallographically independent monomers are compared with each other, it is observed that they are very similar. While in part this is due to the use of local non-crystallographic symmetry restraints in the refinement, the fact that including these restraints significantly improved correspondence of the model to the data supports the similarity of the four crystallographically independent protein chains. Chains A and B on one hand, and chains C and D on the other, can be most reliably superposed, with root mean square differences (r.m.s.d.) between C-alpha atoms of 0.07 and 0.05 Å, respectively. The r.m.s.d. between chains A or B on one hand and chains C or D on the other are 0.23-0.26 Å. The largest structural differences are concentrated in residues 29–39 and 136–143, part of surface loops that interact with each other. These differences between the monomers are likely caused by interaction with neighbouring monomers in the crystal, i.e. different crystal contacts. The loop consisting of residues 136 to 143 is right next to a putative substrate-binding groove, so it may be somewhat more flexible to allow access of the substrate and release of the cleavage products.

The CHAP$_K$ protein consists of a single globular domain that contains two alpha-helices, two 3_{10}-helices and six beta-strands (Figure 1A and B). The amino-terminal part of the protein consists of the two alpha-helices (I and II) interconnected by a long loop. This long loop borders a groove in the protein, at the bottom of which the catalytic site is located (see below). Another loop, containing a 3_{10}-helix, connects this amino-terminal part of the protein to a six-stranded beta-sheet that forms the carboxy-terminal part. The six beta-strands are arranged in an anti-parallel beta-sheet in the topology AFBCDE (Figure 1B). The structure of CHAP$_K$ had previously been predicted by in silico modelling [13]. The six-stranded beta-sheet was predicted well, but the amino-terminal alpha-helices were incorrectly placed and the calcium-binding loop between them was not present in the model. The main chain atoms of the catalytic site residues were within 2 Å of their predicted positions.

Table 1 Refinement and validation statistics for the CHAP$_K$ structure

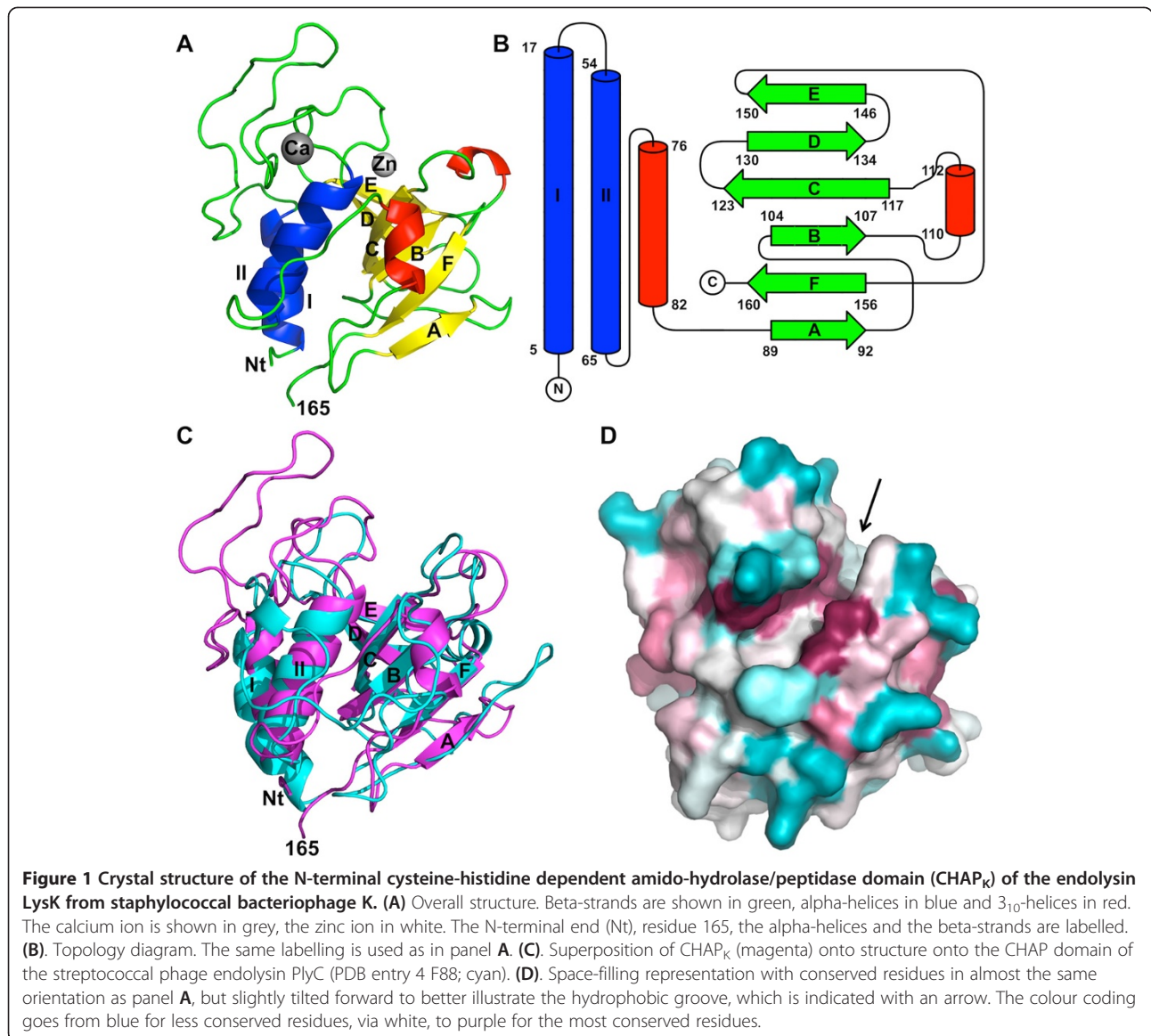|	Native	Derivative
PDB code	4CSH	4CT3
Space group	P1	P1
Cell edges (a, b, c, Å)	39.2, 61.5, 73.2	39.0, 61.5, 72.8
Cell angles (α, β, γ, °)	91.5, 98.7, 90.1	91.8, 98.7, 90.0
Resolution range used (Å)	32.9-1.79 (1.88-1.79)[a]	61.5-1.69 (1.78-1.69)
Multiplicity	2.0 (1.9)	3.4 (3.2)
Completeness	97.2 (94.3)	64.7 (10.4)
Mean <I/sigma(I)>	6.3 (2.9)	11.8 (1.8)
R$_{sym}$ (%)[b]	9.1 (25.1)	6.0 (62.0)
Number of reflections used	59686 (8628)	46067 (1349)
Number of reflections used for R-free	2338 (112)	2431 (62)
R-factor[c]	0.175 (0.233)	0.181 (0.278)
R-free	0.201 (0.282)	0.224 (0.295)
Number of atoms (protein/water/other)	5286/741/66	5259/770/104
Average B-value/Wilson B-value (Å^2)	18.4/14.5	22.2/14.9
Ramachandran statistics[d] (%)	97.7/100.0	98.2/100.0
R.m.s. deviations[e] (bonds, Å/angles, °)	0.015/1.49	0.012/1.37

[a]Values in parentheses are for the highest resolution bin, where applicable.
[b]$R_{sym}=\Sigma_h\Sigma_i|I_{hi}-<I_h>|/\Sigma_h\Sigma_i|I_{hi}|$, where I_{hi} is the intensity of the ith measurement of the same reflection and $<I_h>$ is the mean observed intensity for that reflection.
[c]$R=\Sigma||Fobs(hkl)|-|Fcalc(hkl)||/\Sigma|Fobs(hkl)|$.
[d]Determined with MOLPROBITY. The percentages are indicated of residues in favoured and allowed regions of the Ramachandran plot, respectively.
[e]Provided by REFMAC.

Figure 1 Crystal structure of the N-terminal cysteine-histidine dependent amido-hydrolase/peptidase domain (CHAP$_K$) of the endolysin LysK from staphylococcal bacteriophage K. (A) Overall structure. Beta-strands are shown in green, alpha-helices in blue and 3$_{10}$-helices in red. The calcium ion is shown in grey, the zinc ion in white. The N-terminal end (Nt), residue 165, the alpha-helices and the beta-strands are labelled. **(B)**. Topology diagram. The same labelling is used as in panel **A**. **(C)**. Superposition of CHAP$_K$ (magenta) onto structure onto the CHAP domain of the streptococcal phage endolysin PlyC (PDB entry 4 F88; cyan). **(D)**. Space-filling representation with conserved residues in almost the same orientation as panel **A**, but slightly tilted forward to better illustrate the hydrophobic groove, which is indicated with an arrow. The colour coding goes from blue for less conserved residues, via white, to purple for the most conserved residues.

When the structure is analyzed, it is clear that CHAP$_K$ belongs to the cysteine protease CA peptidase clan Pfam: CL0125; http://pfam.xfam.org/; Ref. [14], with a papain-like fold. CHAP$_K$ is a member of the CHAP family of this clan (Pfam: PF05257), as expected from sequence homology. A structural similarity search revealed that the most similar structure is the CHAP domain of the streptococcal phage endolysin PlyC (PDB entry 4F88) [15], with a root mean square difference (r.m.s.d.) of 2.5 Å when the backbone atoms of 124 residues are superposed onto CHAP$_K$ (Z-score 11.4). The next most similar structure is the C-terminal endopeptidase domain of the NlpC/P60 family cell-wall remodelling protein *Bacillus cereus* PDB code 3H41; Ref. [16], with an r.m.s.d. of 2.8 Å when the backbone atoms of 114 residues are superposed (Z-score 10.2).

When the PDB database is searched for sequence-similar structures, the first hit is the CHAP domain from *Staphylococcus saprophyticus* CHAP domain protein (PDB entry 2K3A) [17], with a sequence identity of 28% over a stretch of 94 residues. However, this structure cannot be superimposed as well as those previously mentioned (r.m.s.d. of 3.4 Å when backbone atoms of 101 residues are superposed, Z-score 6.3) and our attempts to solve the CHAP$_K$ structure by molecular replacement using this model were unsuccessful. This lower similarity may be due to the fact that this structure was determined by NMR spectroscopy rather than crystallography. Superposition of the CHAP$_K$ structure onto the CHAP domain of the streptococcal phage endolysin PlyC (PDB entry 4F88) is shown in Figure 1C. The two alpha-helices and six beta-strands

of CHAP$_K$ superpose quite well with the backbone of the homologous structures, but the loops, including the 3_{10}-helices, are very different.

The globular CHAP$_K$ protein has a relatively long and deep hydrophobic groove. When sequence conservation is mapped onto the surface, one notices that several residues lining the groove are highly conserved (Figure 1D; the sequence alignment underlying this figure is in Additional file 1: Table S1). In the native structure, a MES molecule is located in this groove (Figure 2, PDB entry 4CSH), while in the derivative structure a HEPES molecule is present (PDB entry 4CT3). These molecules may well be mimicking the natural peptidoglycan substrate of the protein. Residues in the groove that might contact the peptidoglycan substrate are: Phe36, Asp47, Tyr49, Tyr50, Gln53 and Cys54 from the loop between helices 1 and 2; Asp56 and Thr59 from helix 2; Arg71, Trp73 and Asn75 from the loop between helix 2 and beta-strand A; Trp115 and His117 from the BC-loop and Asn136 and Trp137 from the DE-loop.

Bound metal ions

While building and refining the protein model, relatively strong density peaks were observed near the terminal atoms of the side-chains of Cys54 and Asp56 in each of the four protein chains in the asymmetric unit, suggesting the presence of metal ions. X-ray fluorescence spectroscopy is a powerful method to identify trace elements in biological samples [18]. Therefore, we recorded an X-ray fluorescence spectrum from a frozen native CHAP$_K$ protein crystal, which revealed significant amounts of

Figure 2 MES buffer molecule bound to the CHAP$_K$ enzyme putative substrate binding site. The CHAP$_K$ protein is shown in transparent surface and secondary structure cartoon representation; the calcium ion is also shown.

zinc and calcium (Figure 3A). Sulphur (from methionine, cysteine residues and buffer molecules) and chlorine (from the crystallization buffer) were also detected. The presence of trace amounts of titanium and copper is likely the result of interaction of the beam with certain beamline or sample holder components not related to the sample.

The calcium ion is bound in the amino-terminal part of the protein, involving residues of the long loop connecting the first and second alpha-helices (residues 17–54) and Asp56 in the second alpha-helix. It is bound in a monodentate way to the side chain of residues Asp45 and Asp47 and in a bidentate way to both oxygen atoms of the Asp56 side chain (Figure 3B). Additional ligands are the main chain oxygen atoms of Tyr49 and His51 and an ordered water molecule. The coordination is octahedral and almost exclusively involves carbonyl oxygen atoms, as expected for calcium. Experimentally determined metal ion-oxygen distances are 2.3-2.5 Å, which is also consistent with usual calcium (II) coordination [19]. The occupancy of the calcium site appears to be complete and the refined temperature factors of the calcium ions are very near those of the coordinating atoms (the temperature factors for the calcium ions vary between 10 and 12 Å^2, while those for the coordinating ligand atoms are between 7 and 14 Å^2). The calcium ion is near the proposed catalytic site (Figure 2). We propose that the calcium ion plays a structural role, helping to maintain the structure of the amino-terminal domain and thus its catalytic residues in the correct relative orientation. The calcium ion binding loop also contains residues that may be in contact with the substrate and thus play a role in determining substrate specificity. In the derivative protein structure, the calcium is present at the same occupancy and with the same coordinating ligands.

In contrast to the tightly bound calcium ion, the zinc ions appear to be bound more loosely and the derivative structure shows they could be replaced by methylmercury ions upon soaking of the crystals with methylmercury chloride. Also, the occupancy appears to be less than unity, we estimate it to be around 0.67 based on refinement runs performed at different occupancies. Finally, the resulting electron density around the zinc ions is somewhat ambiguous and we could not model the ligands without some remaining uncertainty. The zinc ions are coordinated by the sulphydryl group of Cys54, the sulphate group of the bound MES and several water molecules (Figure 3C). It is also near the main chain oxygen atom of Gly116. The coordination distances for the zinc ion are not ideal; the zinc ion is too close to Cys54 and too far from the coordinating oxygen atoms. A report by another group showed that zinc ions inhibit the LysK enzyme, while calcium ions have no
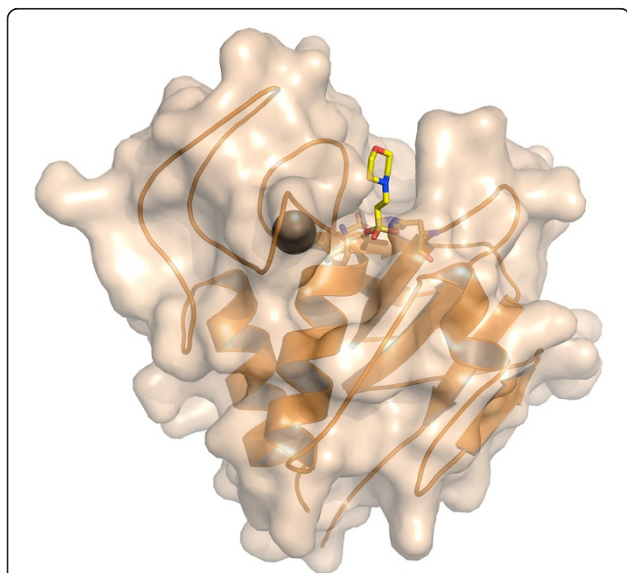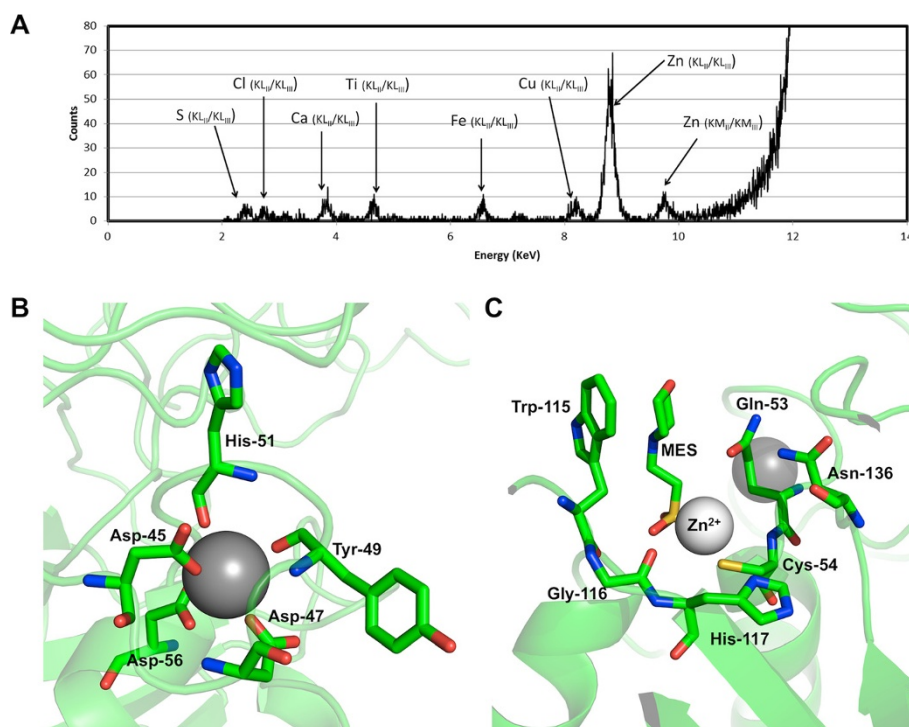
Figure 3 Presence of metal ions in the CHAP$_K$ crystal structure. A. X-ray fluorescence emission spectrum collected from a CHAP$_K$ crystal irradiated with monochromatic synchrotron radiation (12.7 KeV). **B**. Detail of the calcium ion coordination. Coordinating atoms are one Oδ atom of each of Asp45 and Asp47 residues, both Oδ atoms of Asp56, the main chain oxygen atoms of Tyr49 and His51 and an ordered water molecule (behind the calcium ion in this view). **C**. Detail of the zinc coordination. The zinc ion is sandwiched between Cys54 and the sulphate group of the MES ion, about 10 Å away from the calcium ion.

effect on activity, but significantly enhance stability of the enzyme [20]. However, in this assay, metal ions were not removed from the protein solution prior to testing their effects on the enzyme. Zinc ions may play a regulatory role, and their binding near Cys54 suggests they may regulate access of the substrate to the catalytic site.

The importance of the calcium ion in relation to the catalytic ability of CHAP$_K$ was investigated by creation of mutants containing a single amino acid change to alanine at each of the five residues involved in calcium coordination. Zymogram analysis demonstrated that mutation of residues Asp45, Asp47 and Asp56 resulted in the complete abolishment of the staphylolytic activity of the enzyme (Figure 4). This result indicates that the coordinated calcium ion is essential for the catalytic mechanism of the enzyme and complements a previous study, which showed that the chelator EDTA was able to reduce CHAP$_K$ activity by 99% [21]. While mutant His51-Ala retained staphylolytic ability, activity of the enzyme was visibly reduced in comparison with the parental CHAP$_K$. Mutation of Tyr49 to alanine did not appear to affect the staphylolytic ability of the enzyme as the clearing produced on a zymogram gel was comparable to that seen for non-mutated CHAP$_K$ (Figure 4). The fact that mutants His51-Ala and Tyr49-Ala retained

activity while the other mutants did not may be explained by the fact that main chain oxygen atoms are involved in coordination as opposed to the side chain oxygens. Therefore these residues are more amenable to substitution without eliminating catalytic activity.

Catalytic centre and proposed reaction mechanism

By comparing the CHAP$_K$ protein with other proteins with a similar function and structure (endolysins, CHAP domains and others) and by doing an alignment between them, we can deduce that the catalytic residues are highly conserved. In the CHAP domain of *Staphylococcus saprophyticus* (PDB code 2K3A), the authors describe the presence of a proteolytic triad formed by Cys57, His109 and Glu126 [17], a catalytic triad also found in other members of the CA clan. In the streptococcal phage lysin PlyC (PDB code 4 F88), the catalytic residues are Cys333 and His420 [15], while in NlpC/P60 domain of lipoprotein SPR from *E. coli* (PDB code 3H41) the catalytic residues are Cys68, His119 and His339 [22]. In CHAP$_K$ these residues correspond in the alignment to Cys54 located in the second alpha-helix, His117 in beta-strand C and Glu134 in beta-strand D, making these amino acids good candidates to form the catalytic triad of the enzyme (Figure 5). These hypothetical catalytic

Figure 4 Overexpression and activity of CHAP$_K$ mutants. A. Sodium dodecyl sulphate polyacryalamide electrophoresis gel of lysates containing over-expressed CHAP$_K$ and site-directed mutants. A control not expressing CHAP$_K$ is also included. **B**. Composite zymogram gel of CHAP$_K$, site-directed mutant CHAP$_K$ variants and negative control expression lysates.
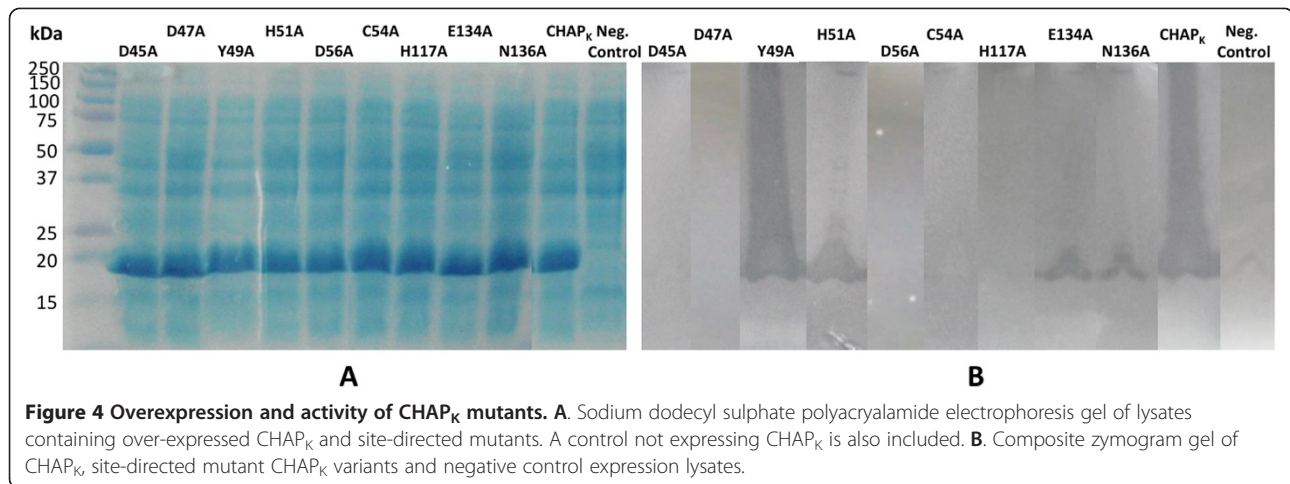
residues are close to the hydrophobic cleft, which supports the possibility that the catalytic part of the molecule is located in the hydrophobic groove. The predicted pKa of His117 is 9.3. This value contrasts with those of the rest of histidines in the protein: His51 (pKa 5.4), His91 (pKa 6.8) and His 157 (pKa 5.2). His117 may thus be protonated at physiological pH.

Mutation of the conserved Cys54 and His117 residues to alanine resulted in complete elimination of staphylolytic activity of the enzyme as demonstrated by zymographic analysis, indicating an essential role of these residues and supporting the hypothesis that they are part of the catalytic triad. Glu134 is believed to be the other residue of the catalytic triad, but is not as highly conserved as the

Figure 5 The proposed catalytic triad of the bacteriophage K endolysin CHAP domain CHAP$_K$. Cys54 (bottom), His117 (middle) and Glu134 (top) and the distances between them (in Å) are shown.

other two residues. When this residue was mutated to alanine, it was clear from zymogram results that, although the catalytic activity was not completely eliminated, it was strongly reduced. In the absence of Glu134 perhaps another residue can take over its role.

A likely mechanism of action, analogous to that of other papain proteases [23,24], is the following: Glu134 accepts a proton from the protonated imidazole group of His117. His117 subsequently accepts a proton from the hydroxyl group of Cys54 (through its N-epsilon). The deprotonated Cys54 then performs a nucleophilic attack on the peptidic bond between D-Ala and Gly in the staphylococcal peptidoglycan. As a result, a transacylation reaction between the enzyme and substrate occurs, giving rise to an acyl-enzyme intermediate. This intermediate may be hydrolyzed to release the enzyme and the cleaved peptidoglycan [25]. In the NlpC/P60 domain of lipoprotein SPR from *E. coli*, there is a tyrosine residue (Tyr56) that has been reported to be very conserved and which may modulate Cys nucleophilicity or help in substrate binding [22]. In the case of CHAP$_K$, Tyr140 is located in an equivalent position, but having a different role, since its phenol group is pointing in the opposite direction. Cysteine proteases have an oxyanion hole, which helps to stabilize the developing negative charge during the formation of the acylenzyme intermediate [26]. Asn136, which is located in close proximity to the catalytic triad, is one residue hypothesized to be involved in creating the oxyanion hole. When this residue was mutated to an alanine, the activity of the enzyme was visibly reduced, but not completely eliminated, supporting the aforementioned hypothesis.
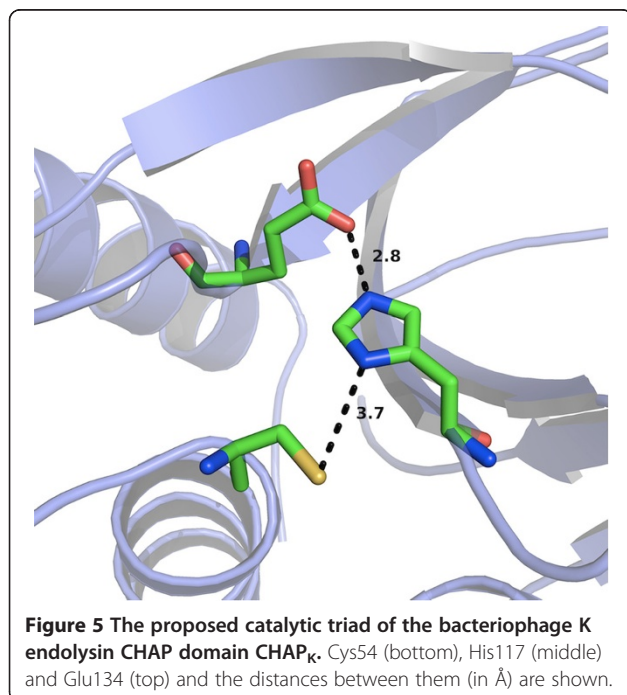
Comparison with LysGH15 CHAP domain structure

While this manuscript was under review, a paper describing the structures of the CHAP domain (PDB entry 4 OLK), amidase-2 domain (PDB entry 4OLS) and the

SH3 domain (PDB entry 2MK5) of the endolysin LysGH15 from phage GH15 was published [27]. The first two were solved by X-ray crystallography at 2.7 and 2.2 Å resolution respectively, while the latter was solved by NMR spectroscopy. Phages GH15 and K share 97% identity in 84% of their genomes (Genbank entries NC_019448 and NC_005880, respectively) [2,28]. The LysGH15 and LysK protein sequences are virtually identical, with only four amino acid differences in their 495-residue sequences. Of the differences, two are in the CHAP domain: Val26 of $CHAP_K$ is an isoleucine in $CHAP_{GH15}$ and Glu113 of $CHAP_K$ is a glutamine in $CHAP_{GH15}$. The high sequence similarity means the enzymes are almost identical and expected to share the same properties.

When the crystal structures of the CHAP domains are compared, it is notable the spacegroups and crystal packing are very different, which suggests the protein is a monomer in solution and inter-monomer interactions in the crystal are not likely to be biologically relevant. Given the almost identical sequences, it is not surprising that the monomer structures are highly similar; superposition of the two CHAP domains leads to an r.m.s.d. of 0.3 Å when 139 C-alpha atoms are superposed. The only significant difference in main-chain conformation is present in residues 109–116, which follow a different path in the two structures. This may indicate that this loop, which is directed away from the active site, is flexible and of limited importance to the structure and activity of the enzyme. The large side-chains of Tyr49, Trp73, Tyr140 and Tyr153, which are all on the surface of the protein, show different orientations.

The higher resolution of the $CHAP_K$ structure when compared to the $CHAP_{GH15}$ structure (1.8 vs. 2.7 Å) should have led to more accurate placement of side-chain atoms and solvent molecules. In both structures, a buffer molecule occupies the groove that likely accommodates the peptidoglycan substrate: a Bis-Tris molecule (2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)-1,3-propanediol) in between the two monomers of the asymmetric unit of $CHAP_{GH15}$ and a MES and HEPES molecule in the case of the native and derivative structures of $CHAP_K$, respectively. The calcium ion is in exactly the same position, as are its coordinating residues and the EF-hand-like domain in which it is incorporated. No zinc ion was observed in the $CHAP_{GH15}$ crystals.

Gu et al. also performed site-directed mutagenesis studies [27], but on the intact LysGH15 enzyme, not on the isolated $CHAP_{GH15}$ domain. As observed for $CHAP_K$, it was found that mutating the active site residue Cys54 affected bacterial lysis activity strongly. Mutating the calcium ion coordinating residues Asp45, Asp46 and Asp56 also diminished activity about ten-fold, while Tyr49 and His51 seem less important, the same as we observed.

Conclusions

We determined the structure of the $CHAP_K$ domain of LysK at 1.8 Å resolution (1 Å = 0.1 nm). The structure has the papain-type fold with a long loop between the two amino-terminal alpha-helices. The structure suggests the location of the active site near a hydrophobic groove, with Cys54, His117 and Glu134 forming the catalytic triad. The substrate most likely binds to the hydrophobic groove.

A calcium ion was found tightly bound to the protein. Its ligands are the side-chains of Asp45, Asp47 and Asp56, plus the backbone oxygens of Tyr49 and His51, all in the amino-terminal domain specific to $CHAP_K$. It likely has a structural role, stabilizing the protein fold. It may also be involved in ensuring the correct location of the peptidoglycan inside the catalytic cleft or in the stabilization of the negative charge of the tetrahedral intermediate during catalysis. A zinc ion was also found and is likely more loosely bound, as it is less buried, has less protein ligands and could be exchanged for a methylmercury ion upon derivatization. Its role, if any, may be regulatory.

Based on the structure, we propose a possible reaction mechanism, involving all three residues of the likely catalytic triad. Future studies will include co-crystallization with peptidoglycan analogues and elucidating the role of the $CHAP_K$ domain in the complete LysK protein. This may allow site-directed mutation to modulate the peptidoglycan specificity and activity of both the $CHAP_K$ and LysK enzymes.

Methods

$CHAP_K$ was expressed, purified, crystallized and crystallographic data was collected as described [9,12]. A complete native dataset was collected to 1.8 Å resolution with good statistics. A dataset to 1.7 Å resolution, but with inferior completeness, was also collected from a methylmercury chloride derivative at the Hg L-I edge [12]. However, this dataset allowed phase determination by single anomalous dispersion (SAD) and automatic model building of four crystallographically independent protein molecules in the P1 unit cell [12] (Table 1) using the ARP-WARP program [29]. The model was refined against the derivative dataset and separately against the native dataset. The models were completed and adjusted using COOT [30] and refined with REFMAC5, using local non-crystallographic symmetry restraints [31] and taking care to select the same reflections for calculation of Rfree [32]. To confirm the presence of zinc and calcium ions in the sample, an X-ray fluorescence emission spectrum was collected on a native protein crystal at ESRF beamline ID23-1 [33]. Validation was performed with MolProbity [34]. Refinement and validation statistics are shown in Table 1.

Crystal contact analysis was done with PISA [35]; other analyses were performed with the CCP4 suite [36]. Structural similarity analysis was performed with DALI [37]; for plotting a protein surface coloured according to amino acid conservation, CONSURF was used [38]. The pKa of selected residues in the protein structure was predicted with PROPKA [39]. The structural models and underlying data files have been submitted to the PDB (accession code 4CSH for the native structure and 4CT3 for the derivative). PYMOL (Schrödinger LLC, Portland OR, USA) was used for making structure figures and TOPDRAW [40] to draw the secondary structure diagram.

CHAP$_K$ mutants were created using the QuikChange II Site-Directed Mutagenesis Kit from Agilent (Santa Clara CA, USA) as per the manufacturer's instructions. Crude cell lysate was analyzed for over-expression using sodium dodecyl sulphate gel electrophoresis and for ability to lyse *Staphylococcus aureus* cells using zymographic gels as described previously [41].

Additional file

Additional file 1: Table S1. Sequence aligment underlining the colour coding of Figure 1D.

Author details
[1]Departamento de Estructura de Macromoleculas, Centro Nacional de Biotecnologia (CNB–CSIC), Calle Darwin 3, E-28049 Madrid, Spain. [2]Department of Biological Sciences, Cork Institute of Technology, Bishopstown, Cork, Ireland. [3]Current address: Department of Biochemistry, University of Zurich, Zurich, Switzerland.

References
1. Rees PJ, Fry BA: The morphology of staphylococcal bacteriophage K and DNA metabolism in infected *Staphylococcus aureus*. *J Gen Virol* 1981, **53**(Pt 2):293–307.
2. O'Flaherty S, Coffey A, Edwards R, Meaney W, Fitzgerald GF, Ross RP: Genome of staphylococcal phage K: a new lineage of *Myoviridae* infecting Gram-positive bacteria with a low G + C content. *J Bacteriol* 2004, **186**(9):2862–2871.
3. Gill JJ: Revised genome sequence of staphylococcus aureus bacteriophage K. *Genome Announcements* 2014, **2**(1):e01173.
4. Loessner MJ: Bacteriophage endolysins - current state of research and applications. *Curr Opin Microbiol* 2005, **8**(4):480–487.
5. Ralston DJ, McIvor M: Lysis-from-without of *Staphylococcus aureus* strains by combinations of specific phages and phage-induced lytic enzymes. *J Bacteriol* 1964, **88**:676–681.
6. Loeffler JM, Nelson D, Fischetti VA: Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. *Science* 2001, **294**(5549):2170–2172.
7. O'Flaherty S, Coffey A, Meaney W, Fitzgerald GF, Ross RP: The recombinant phage lysin LysK has a broad spectrum of lytic activity against clinically relevant staphylococci, including methicillin-resistant *Staphylococcus aureus*. *J Bacteriol* 2005, **187**(20):7161–7164.
8. Becker SC, Dong S, Baker JR, Foster-Frey J, Protchard DF, Donovan DM: LysK CHAP endopeptidase domain is required for lysis of live staphylococcal cells. *FEMS Microbiol Lett* 2009, **294**(1):52–60.
9. Horgan M, O'Flynn G, Garry J, Cooney J, Coffey A, Fitzgerald GF, Ross RP, McAuliffe O: Phage lysin LysK can be truncated to its CHAP domain and retain lytic activity against live antibiotic-resistant staphylococci. *Appl Environ Microbiol* 2009, **75**(3):872–874.
10. Fenton M, Casey PG, Hill C, Gahan CG, Ross RP, McAuliffe O, O'Mahony J, Maher F, Coffey A: The truncated phage lysin CHAP(k) eliminates Staphylococcus aureus in the nares of mice. *Bioeng Bugs* 2010, **1**(6):404–407.
11. Fenton M, Keary R, McAuliffe O, Ross RP, O'Mahony J, Coffey A: Bacteriophage-derived peptidase CHAP$_K$ eliminates and prevents staphylococcal biofilms. *Internat J Microbiol* 2013, **2013**:625341.
12. Sanz-Gaitero M, Keary R, Garcia-Doval C, Coffey A, van Raaij MJ: Crystallization of the CHAP domain of the endolysin from *Staphylococcus aureus* bacteriophage K. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2013, **69**(Pt 12):393–1396.
13. Fenton M, Cooney JC, Ross RP, Sleator RD, McAuliffe O, O'Mahony J, Coffey A: *In silico* modeling of the staphylococcal bacteriophage-derived peptidase CHAP$_K$. *Bacteriophage* 2011, **1**(4):198–206.
14. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD: The Pfam protein families database. *Nucl Acids Res* 2012, **40**:D290–D301.
15. McGowan S, Buckle AM, Mitchell MS, Hoopes JT, Gallagher DT, Heselpoth RD, Shen Y, Reboul CF, Law RHP, Fischetti VA, Whisstock JC, Nelson DC: X-ray crystal structure of the streptococcal specific phage lysin PlyC. *Proc Natl Acad Sci U S A* 2012, **109**(31):12752–12757.
16. Xu Q, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, Cai X, Carlton D, Chen C, Chiu HJ, Chiu M, Clayton T, Das D, Deller MC, Duan L, Ellrott K, Farr CL, Feuerhelm J, Grant JC, Grzechnik A, Han GW, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Krishna SS, Kumar A, Lam WW, Marciano D, Miller MD, *et al*: Structure of the c-D-glutamyl-L-diamino acid endopeptidase YkfC from *Bacillus cereus* in complex with L-Ala-c-D-Glu: insights into substrate recognition by NlpC/P60 cysteine peptidases. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2010, **66**(Pt 10):1354–1364.
17. Rossi P, Aramini JMR, Xiao R, Chen CX, Nwosu C, Owens LA, Maglaqui M, Nair R, Fischer M, Acton TB, Honig B, Rost B, Montelione GT: Structural elucidation of the Cys-His-Glu-Asn proteolytic relay in the secreted CHAP domain enzyme from the human pathogen staphylococcus saprophyticus. *Proteins* 2009, **74**(2):515–519.
18. Jones KW, Gordon BM, Hanson AL, Kwiatek WL, Pounds JG: X-ray fluorescence with synchrotron radiation. *Ultramicroscopy* 1988, **24**(2–3):313–328.
19. Harding MM: Small revisions to predicted distances around metal sites in proteins. *Acta Crystallogr Sect D Biol Crystallogr* 2006, **62**(Pt 6):678–682.
20. Filatova LY, Becker SC, Donovan DM, Gladilin AK, Klyachko NL: LysK, the enzyme lysing staphylococcus aureus cells: specific kinetic features and approaches towards stabilization. *Biochimie* 2010, **92**:507–513.

21. Fenton M, Ross RP, McAuliffe O, O'Mahony J, Coffey A: **Characterization of the staphylococcal bacteriophage lysin CHAP$_K$.** *J Appl Microbiol* 2011, **111**:1025–1035.
22. Aramini JM, Rossi P, Huang YJ, Zhao L, Jiang M, Maglaqui M, Xiao R, Locke J, Nair R, Rost B, Acton TB, Inouye M, Montelione GT: **Solution NMR structure of the NlpC/P60 domain of lipoprotein Spr from Escherichia coli: structural evidence for a novel cysteine peptidase catalytic triad.** *Biochemistry* 2008, **47**(37):9715–9717.
23. Shokhen M, Khazanov N, Albeck A: **The mechanism of papain inhibition by peptidyl aldehydes.** *Proteins* 2011, **79**:975–985.
24. Brömme D: **Papain-like cysteine proteases.** *Curr Protoc Protein Sci* 2001, **21**:21.2.1–21.2.14.
25. Lau EY, Bruice TC: **Consequences of breaking the Asp-His hydrogen bond of the catalytic triad: effects on the structure and dynamics of the serine esterase cutinase.** *Biophys J* 1999, **77**(1):85–98.
26. Menard R, Storer AC: **Oxyanion hole interactions in serine and cysteine proteases.** *Biol Chem Hoppe Seyler* 1992, **373**(7):393–400.
27. Gu J, Feng Y, Feng X, Sun C, Lei L, Ding W, Niu F, Jiao L, Yang M, Li Y, Liu X, Song J, Cui Z, Han D, Du C, Yang Y, Ouyang S, Liu ZJ, Han W: **Structural and biochemical characterization reveals LysGH15 as an unprecedented "EF-hand-like" calcium-binding phage lysin.** *PLOS Path* 2014, **10**(5):e1004109.
28. Gu J, Liu X, Lu R, Li Y, Song J, Lei L, Sun C, Feng X, Du C, Yu H, Yang Y, Han W: **Complete genome sequence of staphylococcus aureus bacteriophage GH15.** *J Virol* 2012, **86**(16):8914–8915.
29. Langer G, Cohen SX, Lamzin VS, Perrakis A: **Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7.** *Nat Protoc* 2008, **3**(7):1171–1179.
30. Emsley P, Cowtan K: **Coot: model-building tools for molecular graphics.** *Acta Crystallogr Sect D Biol Crystallogr* 2004, **60**(Pt 12 Pt 1):2126–2132.
31. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA: **Refmac5 for the refinement of macromolecular crystal structures.** *Acta Crystallogr Sect D Biol Crystallogr* 2011, **67**(Pt 4):355–367.
32. Brünger A: **Free R value: a novel statistical quantity for assessing the accuracy of crystal structures.** *Nature* 1992, **355**(6359):472–475.
33. Leonard GA, Solé VA, Beteva A, Gabadinho J, Guijarro M, McCarthy J, Marrocchelli D, Nurizzo D, McSweeney S, Mueller-Dieckmann S: **Online collection and analysis of X-ray fluorescence spectra on the macromolecular crystallography beamlines of the ESRF.** *J Appl Crystallogr* 2009, **42**:333–335.
34. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC: **MolProbity: all-atom structure validation for macromolecular crystallography.** *Acta Crystallogr Sect D Biol Crystallogr* 2010, **66**(Pt 1):12–21.
35. Krissinel E, Henrick K: **Inference of macromolecular assemblies from crystalline state.** *J Mol Biol* 2007, **372**(3):774–797.
36. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS: **Overview of the CCP4 suite and current developments.** *Acta Crystallogr Sect D Biol Crystallogr* 2011, **67**(Pt 4):235–242.
37. Holm L, Rosenström P: **Dali server: conservation mapping in 3D.** *Nucleic Acids Res* 2010, **38**(Web Server issue):W545–W549.
38. Armon A, Graur AD, Ben-Tal N: **ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information.** *Bioinformatics* 2003, **19**(1):163–164.
39. Rostkowski M, Olsson MHM, Søndergaard CR, Jensen JH: **Graphical analysis of pH-dependent properties of proteins predicted using PROPKA.** *BMC Struct Biol* 2011, **11**:6.
40. Bond CS: **Topdraw: a sketchpad for protein structure topology cartoons.** *Bioinformatics* 2003, **19**(2):311–312.
41. Keary R, McAuliffe O, Ross RP, Hill C, O'Mahony J, Coffey A: **Genome analysis of the staphylococcal temperate phage DW2 and functional studies on the endolysin and tail hydrolase.** *Bacteriophage* 2014, **4**(1):e28451.