Cysteine biosynthetic genes are up-regulated in the persistent phase of Mycobacterium tuberculosis and the corresponding enzymes are therefore of interest as potential targets for novel antibacterial agents. cysK1 is one of these genes and has been annotated as coding for an O-acetylserine sulfhydrylase. Recombinant CysK1 is a PLP dependent enzyme that catalyzes the conversion of O-acetylserine to cysteine. The crystal structure of the enzyme was determined to 1.8 Å resolution. CysK1 belongs to the family of fold type II PLP enzymes and is similar in structure to other O-acetylserine sulfhydrylases. We were able to trap the α-aminoacrylate reaction intermediate and determine its structure by cryocrystallography. Formation of the aminoacrylate complex is accompanied by a domain rotation resulting in active site closure. The aminoacrylate moiety is bound in the active site via the covalent linkage to the PLP cofactor and by hydrogen bonds of its carboxyl group to several enzyme residues. The catalytic lysine residue is positioned such that it can protonate the Cα carbon atom of the aminoacrylate only from the si-face resulting in the formation of L-cysteine.

O-acetylserine sulfhydrylases. The structure of the enzyme-peptide complex provides a framework for the design of strong binding inhibitors.

Gene expression and proteome analysis of various models of dormancy in Mycobacterium tuberculosis have shown that the genes involved in sulphate transport (1), sulphate reduction (PAPS pathway) and cysteine biosynthesis (2-5) are up-regulated. In the persistent phase the pathogen is surviving within macrophages (6,7), where it is exposed to oxidative stress and Reactive Nitrogen Intermediates (RNI) as a cellular response to pathogen invasion, leading to oxidation and S-nitrosylation of cysteine residues (8). Bacteria in the group designated as Actinomycetales, including mycobacteria, produce mycothiol (1-d-myo-inosityl-2-(N-acetyl-cysteinyl) amino-2-deoxy-α-D-glucopyranoside) as their principal low-molecular-mass thiol, which plays the role of glutathione and contains a cysteine moiety (9). Thereby the first defence line in oxidative stress is directly linked to the availability of cysteine. Inhibition of cysteine biosynthesis therefore appears to be an attractive route to novel antibacterials that may be active also against persistent M. tuberculosis.

In plants, some archaea and most eubacteria the de-novo biosynthesis of L-cysteine starts from serine and involves two enzymes. Serine acetyl transferase (SAT, EC 2.3.1.30), also denoted CysE, catalyzes the formation of O-acetylserine (OAS) from acetyl-CoA and serine. This
metabolite is then converted to cysteine through elimination of acetate and addition of hydrogen sulphide by the enzyme O-acetylserine sulphydrylase (OASS, EC 2.5.1.47) (for recent reviews see 10, 11). The sulphur source is supplied by the sulphate reduction (APS/PAPS) pathway. In most bacteria pyridoxal phosphate (PLP) dependent OASS is found as two isoenzymes, OASS-A (also denoted CysK) and OASS-B (denoted CysM). These isoenzymes show 25-45% identity in amino acid sequence, but exhibit characteristic differences in their substrate specificity with respect to the sulphur donor. CysK uses hydrogen sulphide as sulphur source, while the characterized CysM proteins tend to accept thiosulphate and larger substrates as sulphur source (12,13). The genome of Mycobacterium tuberculosis H37Rv encodes three genes which were annotated as OASS (http://genolist.pasteur.fr/TubercuList/). CysK1 (Rv2334) is homologous to OASS-A, and CysM (Rv1336) homologous to OASS-B. The third isoenzyme, encoded by Rv0848, shows similarities to the other two enzymes to the same extent, and is therefore designated as either CysK2 or CysM2.

The crystal structure analysis of OASS from Salmonella typhimurium (14) Haemophilus influenzae (15), Escherichia coli (16) and Arabidopsis thaliana (17) revealed that it belongs to the fold-type II class of vitamin B₆ dependent enzymes (18, 19). The enzymatic mechanism of OASS follows a ping-pong mechanism typical for PLP dependent enzymes (10,11,20). The reaction can be divided into two halves (scheme 1). In the resting state the cofactor forms an internal Schiff base with the invariant catalytic lysine residue (compound I in scheme 1). The incoming substrate then forms an external Schiff base with PLP (II), followed by a β elimination in which acetate is released and a proton is abstracted from the α position, most likely by the lysine side chain (21). A similar role of the invariant lysine residue has been proven in the homologous tryptophan synthase β subunit (22). The product of the first half reaction is the α-aminoacrylate intermediate (23), covalently linked to PLP (III). The second half reaction starts with a nucleophilic attack of the sulphide (or HS-) on the β-carbon of the aminoacrylate intermediate and the α carbon is re-protonated (20,24) resulting in cysteine bound as external Schiff-base (IV). Finally the product is released via regeneration of the internal aldimine. Formation of the external aldimin appears to be coupled to a large conformational change which leads to the closure of the active site as suggested by the structure of the Salmononella typhimurium CysK K41A mutant in complex with methionine, a substrate analogue (25). However, in the corresponding structure of OASS from Arabidopsis thaliana, the enzyme remained in the open conformation also after formation of the covalent PLP-methionine adduct (17). The latter is assumed to mimic the external aldimine (II).

In this work we describe the cloning, purification, characterization and the 1.8 Å X-ray structure of CysK1 from Mycobacterium tuberculosis. We were also able to trap the α-aminoacrylate in the crystal and report the structure of this reaction intermediate to 2.2 Å resolution, thus providing novel structural insights in the reaction cycle of CysK1. We also determined the crystal structure of CysK1 from M. tuberculosis bound to an inhibitory four-residue peptide derived from the C-terminus of M. tuberculosis CysE (SAT, Rv2335). The structure of this inhibited form of CysK1 may provide the basis for the design of strong binding inhibitors of this enzyme.

**Experimental procedures**

*Gene cloning and protein production*

The DNA sequence coding for CysK1 from M. tuberculosis (Rv2334) was amplified by PCR and cloned using upstream NdeI and downstream HindIII restriction sites. Subsequently, the sequence was verified and the fragment was cloned into the expression vector pET28a (Novagen) resulting in a cleavable six-histidine tag at the N-terminus of the polypeptide chain. In this construct, thrombin cleavage results in a recombinant protein with three additional amino acids (Gly-Ser-His) at the N-terminus.

*Escherichia coli* BL21(DE3) carrying the expression construct pET-His6CysK1 was cultivated in 1.5 L of LB medium supplemented with kanamycine (30 µg/ml) at 21°C. At an OD600 value of 0.5-0.6 gene expression was induced by the addition of 0.1 mM IPTG. After approximately 24 hours the cells were harvested and then resuspended in a buffer consisting of 10 mM Tris-HCl pH 8.0, 300 mM NaCl, and 10 mM imidazole. Cells were disrupted by
freeze/thaw cycles, lysozyme and DNaseI treatment and sonication. The clarified lysates were loaded on a Ni-NTA column (Qiagen) with a column volume of 1.8 ml. After extensive washing with 10 mM imidazole, remaining proteins were eluted with an imidazole step gradient. Pure His6-CysK1 eluted in fractions containing 100-200 mM imidazole. His6-CysK1 containing fractions were pooled, loaded on a desalting PD10 column (GE-Healthcare) and eluted in a buffer of 25 mM Tris-HCl pH 8.0, 150 mM NaCl. The N-terminal His6-tag was removed by thrombin cleavage. Residual non-cleaved His-CysK1 was removed by passage through a Ni-NTA column. The flow-through containing the cleaved CysK1 protein was further purified using an ion-exchange step on a Q-sepharose (Amersham biosciences) column, and NaCl gradient elution. The CysK1 containing fractions (200-250 mM NaCl) were pooled and loaded on a Superdex200 column equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl. This procedure resulted in homogeneous CysK1 as verified by SDS chromatography. CysK1 protein was concentrated to 10 mg/ml using an Amicon centrifugation device with a 3 kDa MW cut-off. Aliquots of the protein preparations were flash-frozen in liquid nitrogen and stored at -80°C until further use.

**Spectrophotometry**

UV-visible absorbance spectra of CysK1 were recorded at a concentration of 3.3 mg/ml using a Jasco-V65 spectrophotometer. Spectra of the covalent aminoacrylate complex were obtained after addition of 300 μM O-acetylserine to the enzyme solution.

**Enzyme activity**

Recombinant CysK1 was assayed for O-acetylserine sulphydrylase activity spectrophotometrically at 560 nm by monitoring the formation of cysteine using the acid-ninhydrine method (26). The reactions were carried out in 1500 μl of 100 mM MOPS buffer at pH7.0. O-acetyl-L-serine and Na-sulphide (dissolved in 1mM NaOH) were added to final concentrations of 10 mM and 0.25 mM, respectively, and the reactions were started by the addition of the enzyme (0.025 μg) to the assay mixtures. Aliquots of 250 μl were taken at various time points and the reaction was stopped by the addition of trichloro-acetic acid (final concentration 16.6%). Following centrifugation the amount of product was determined spectrophotometrically using a standard curve for cysteine in the range 0.04-0.8 mM. Specific activities were derived from the linear part of the time curves, and each measurement was carried out in triplicate at 30 °C and 37 °C, respectively.

**Peptide binding assay and inhibition studies**

The chemically synthesized oligopeptide DFSI was obtained from the GenScript Corporation, Scotch Plains NJ, USA. The binding-interaction was probed in a thermost fluor experiment (27) in the buffer consisting of 25mM Tris-HCl pH 8.0, 150mM NaCl in the presence of the dye Sypro Orange (Sigma). CysK1 at a concentration of 0.4 μg/μl (12 pmol/μl) in a total volume of 25 μl was mixed with the DFSI peptide at concentrations of 0.0016 mM, 0.016 mM, 0.16 mM and 1.6 mM. Fluorescence at the wavelength 575 nm was recorded in the temperature interval 20-90 °C.

The initial rate of cysteine formation by CysK1 was determined at 22 °C in the absence and presence of various DFSI peptide concentrations, and varying the OAS concentration in the range 1.0, 2.0, 5.0, 8.0, and 10 mM) at a constant sodium sulphide concentration of 1.0 mM in 100 mM MOPS buffer, pH 7.0. At various time intervals, aliquots of the reaction mixture were removed and assayed for L-cysteine formation as described above. The mode of inhibition was determined from double reciprocal plots and the Ki value for the peptide was derived using the formalism described in (28).

**Crystallization and data collection**

Crystallisation screening was carried out using the vapour diffusion method and a Phoenix crystallisation robot. After extensive screening and optimization two conditions were established that resulted in single crystals of CysK1. Both contained unusually high MPD (2-methyl-2,4-pentane-diol) concentrations. Crystal form I was obtained by mixing 3 μl of protein solution (10 mg/ml) with 2 μl of reservoir solution (Bis-Tris-propane pH 7.5, 60-65% MPD) in sitting drops. The rod-like crystals diffracted however poorly (approx 6 Å) and were not characterized further. Crystal form II was obtained by mixing 2 μl of protein solution (10 mg/ml) with 2 μl of reservoir solution (Na-
HEPES pH 7.2-8.2, 80% MPD) in either sitting or hanging drops. The diamond-shaped single crystals reproducibly diffracted to 1.8-2.0 Å resolution.

The CysK1 α-aminoacrylate complex was prepared by a combination of soaking and freeze trapping. Sitting drops containing CysK1 crystals obtained at pH 8.2 were layered cautiously with a droplet of mother liquor containing the substrate O-acetylserine at concentrations of 1 mM and 2 mM, respectively. At various time intervals (1mM OAS concentration after 10, 15, 20, 30, 35 to 30 minutes; at 2mM OAS concentration after 10 and 20 minutes), crystals were removed, flash frozen and stored in liquid nitrogen until X-ray analysis. Due to the high MPD concentration in the mother liquor no cryoprotectant was necessary.

Co-crystallization of CysK1 with the DFSI-peptide was achieved using the crystallisation conditions established for the holo-enzyme. CysK1 at a concentration of 10 mg/ml was mixed with the peptide at various peptide concentrations (0.1, 1.0, 4.0, 5.0, 10.0 mM) and pre-incubated at 22° for 30-60 minutes before setting up the crystallization drops using conditions described above.

Diffraction data for crystals of the holo-enzyme was collected at beamline ID14:3 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Diffraction data for the peptide complex was collected at the beamline ID14:1 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Diffraction datasets for the CysK1-α-aminoacrylate complexes were collected at beamline 9/11 at MAXLab (Lund, Sweden). All data sets were collected in a nitrogen gas stream at 110 K.

The X-ray data were processed and scaled with the programs MOSFLM and SCALA from the CCP4 suite (29). The native crystals belong to the tetragonal space group P4_2_2 with cell dimensions a = b = 71.0 Å, and c = 179.6 Å. Soaking of the crystals with OAS resulted in a minor change of cell parameters to a = b = 71.7 Å and c = 181.0 Å. The cocrystallization of CysK1 with the DFSI peptide resulted in crystals with cell dimensions a = b = 72.45 and c = 178.93. The statistics of the data sets are given in Table I.

**Molecular replacement and crystallographic refinement**

The structure was determined by molecular replacement using the program MOLREP (30). The coordinates for OASS from *Arabidopsis thaliana* (17) (PDB accession code 1Z7W) with the cofactor PLP and the bound sulphate ion omitted were used as the search model. The best solution had a score of 0.435 and an R-factor of 48.3%, with a single monomer in the asymmetric unit. In order to monitor the behaviour of the refinement process, 5% of the X-ray data were removed for the calculation of R-free. Initial cycles of restrained refinement using Refmac5 (31) resulted in a drop of the R-factor by 12.8%. The correctness of the molecular replacement solution was confirmed by electron density for the cofactor, including the covalent linkage to the active site lysine residue appearing at the expected positions.

Manual rebuilding of the model was carried out with the program Coot (32), based on sigma-A weighted 2Fo-Fc and Fo-Fc electron density maps (33). Manual adjustment of the model was interspersed with rounds of refinement by Refmac5 (31). Water molecules were added based on peak heights, shape of the electron density, temperature factor and capability to form hydrogen bonds to surrounding protein residues and/or other water molecules. The final model contains amino acid residues 1-300 from CysK1, one PLP molecule, four MPD molecules, one chloride ion and 205 water molecules. The ten last residues at the carboxy-terminus did not have defined electron density.

Due to the change in cell dimensions upon soaking with OAS, molecular replacement was used to solve the structure of the CysK1 α-aminoacrylate complex. An initial CysK1 model was used as search model. The electron density maps calculated from the dataset collected from a crystal soaked in 1mM OAS for 30 min clearly indicated formation of the α-aminoacrylate intermediate. Refinement of the complex was carried out using a similar protocol as outlined above. At the final stage TLS refinement (34) was used with 20 TLS groups as defined by the TLSMD server (35) which resulted in a 2% drop in the free R-factor. The final model contains residues 1-306 from CysK1, the amino-acrylate PLP covalent complex, one MPD and 47 crystallographic water molecules.
The structure of the CysK1–peptide complex was solved by molecular replacement using the initially refined CysK1 structure as search model. The electron density calculated after the initial refinement rounds indicated the presence of the peptide bound at the enzyme active site. Refinement of the complex was carried out using a similar protocol as outlined above. The final model comprises CysK1 residues 1-300, the four residue long peptide, one MPD molecule and 163 water molecules.

The protein models were analyzed with PROCHECK (36) in order to monitor the stereochemistry. Details of the refinement and protein models are given in Table I. Figures were made using the program Pymol (http://www.pymol.org). The sequence alignment was made by ClustalW available at www.ebi.ac.uk (37).

Results and discussion

Characterization of CysK1 from Mycobacterium tuberculosis

CysK1 was produced with a hexahistidine tag at the N-terminus. The tag was removed by thrombin leaving three extra residues Gly-Ser-His on the N-terminus of the purified protein. Recombinant CysK1 has a yellow colour and shows the typical absorption spectrum of PLP dependent enzymes (38). The absorbance peak at 410 nm is characteristic of the internal Schiff-base formed between the cofactor and the active site lysine residue (Fig. 1).

Gel filtration chromatography suggested that CysK1 is a dimer in solution and the enzyme migrated in a native gel as a single well defined band (data not shown). Purified recombinant CysK1 is catalytically active as an O-acetylserine sulfhydrylase. The specific activity was determined with O-acetyl-L-serine and Na-sulphide as substrates. The turnover number was 211 sec\(^{-1}\) at 30 °C and 352 sec\(^{-1}\) at 37 °C. These values compare well to those determined for the related enzyme from \textit{S. typhimurium} which showed 280 sec\(^{-1}\) at 25 °C as turnover rate for the reaction (39).

Overall Structure of \textit{M. tuberculosis} CysK1

The asymmetric unit contains a single copy of the CysK1 subunit, corresponding to a solvent content of 65%. The subunit consists of two domains each having a α/β fold (Fig. 2a and supplementary information, Fig. S1) typical of the fold-type II PLP enzymes (18,19). The N-terminal domain (residues 43–145) is built up by a central four-stranded parallel β-sheet which is flanked by three α helices on one side and one α helix on the other side. The C-terminal domain consists of residues 11–42 and 151–304 that fold into a six-stranded mixed β-sheet sandwiched by four α-helices. In the crystal the enzyme is a dimer, with the molecular axis coinciding with the 2-fold crystallographic axis. The buried surface at the dimer interface is 2130 Å\(^2\), which corresponds to 15.7% of the total surface of one monomer.

CysK1 is related in fold to other PLP enzymes of the fold type II class, i.e. the β-subunit of tryptophan synthase (40), threonine deaminase (41), and OASS (14-17). The closest structural relatives of \textit{M. tuberculosis} CysK1 are OASS from \textit{S. typhimurium} (14) and \textit{A. thaliana} (17) with r.m.s.d. values after structural superposition of 0.81 Å (294 equivalent C\(_\alpha\) atoms) and 0.7 Å (298 equivalent C\(_\alpha\) atoms), respectively.

Pyridoxal phosphate binding site

The PLP binding site is located in a cleft formed between the C-terminal ends of the β-sheets of the two domains within one subunit (Fig. 2a). The cofactor is bound via a covalent linkage to the side chain of K44 (Fig. 2b). The phosphate group forms hydrogen bonds to the highly conserved glycine- and threonine-rich loop (178-GTGGT-182) at the N-terminus of helix 6 and three water molecules (Fig. 2c & d). The aromatic ring of PLP is positioned at a cross-over connection in the β-β sheet of the C-terminal domain. The N-1 atom of the pyridine ring is hydrogen bonded to the side chain of S266, located in another highly conserved region (264-GISSGAA-270) of the polypeptide chain. The 3’ hydroxyl group of PLP forms a hydrogen bond to the side chain of N74 from the highly conserved sequence stretch 69-EPTSGNTG-76, and to one water molecule. Related OASS enzymes (14-17) show basically identical interaction networks with the PLP cofactor.

The active site in the CysK1 holoenzyme

The active site of the CysK1 holoenzyme contains several ligands from the crystallisation solution. A MPD molecule is bound in the active site pocket and, with the exception of a hydrogen bond to carbonyl oxygen of G222, mostly forms...
hydrophobic interactions with residues lining the pocket. The electron density map further contained a strong spherical electron density close to the internal aldimine linkage and the conserved loop (71-TSGNT-75) at the N-terminus of helix α2, which was modelled as a chloride ion, mostly likely originating from the crystallisation buffer (figure 2 c & d). A similar anion binding site has also been found in other O-acetylserine sulphydrylases (14, 17).

Structure of the CysK1 – α-aminoacrylate complex
The formation of the aminoacrylate Schiff-base intermediate, the product of the first half reaction, was observed in solution by spectrophotometry. The UV-Vis spectrum of CysK1 in the presence of 300 μM OAS (Fig. 1) showed the absorbance maxima at 321 and 469 nm, which are characteristic for the formation of the α-aminoacrylate intermediate (23,38). Subsequent addition of 0.25 mM Na-sulphide resulted in the fast disappearance of the band at 469 nm, indicating decomposition of the intermediate. The approximately 60 nm shift of the absorbance maximum in the spectrum corresponds to a visible colour change, which was also observed in the crystals during soaking experiments with OAS. Diffraction data collected from such a crystal after arresting the reaction by flash-freezing were used to solve the structure of CysK1 with the trapped α-aminoacrylate intermediate. Most notably, parts of the N-terminal domain had moved significantly upon formation of the reaction intermediate and had to be rebuilt completely. The largest conformational changes (maximal atomic displacements of 6 Å) involve the loop region consisting of the conserved sequence stretch 71-TSGNT-75, the loop between strand β5 and helix α3 (residues 94-101) and between strand β6 and helix α4 (residues 117-124) that fold over the active site (figure 3a). The changes in cell dimensions in the crystals containing the reaction intermediate most likely reflect the conformational changes of this part of the polypeptide chain. The domain rotation (Fig. 3a) results in the closure of the active site, which becomes less accessible and thus protects the reaction intermediate from solvent.

The electron density map clearly showed that the internal Schiff base linkage between PLP and the catalytic lysine residue was broken, and that the external Schiff base with the α-aminoacrylate had formed in the crystal (Fig. 3b). We did not observe any electron density beyond the Cβ carbon atom of the external Schiff base, indicating that the acetate moiety had been expelled from the substrate, OAS. The carbon atoms of the aminoacrylate moiety and the nitrogen involved in the Schiff base linkage are planar, as expected for the aminoacrylate intermediate. One of the oxygen atoms of the carboxylate group is within hydrogen bonding distance to the side chains of Q144 and T71, and the main chain nitrogen atom of S72. The second oxygen atom is tethered to the backbone amide groups of N74 and T75 (Fig. 3c & d). Most of these residues are part of the anion binding site in the holoenzyme, but as a result of the loop movement are closer to the active site and interact with the carboxyl group of the aminoacrylate.

The formation of the reaction intermediate results in a number of subtle changes in the active site. The pyridine ring of PLP is tilted by approximately 19° along an axis passing through the 2'-5' carbon atoms. A small shift in the position of the phosphate group was found (0.5 Å), which leads to slightly different interactions when compared to the structure of the holoenzyme. One of the phosphate oxygen atoms forms a hydrogen bond with the ε−amino group of the catalytic residue K44 and it is likely that the protonated state of the K44, crucial in the second half reaction, is ensured by this interaction. Another phosphate oxygen atom acquires a water molecule as a new hydrogen bonding partner. In total, the number of hydrogen bonds of the phosphate group increases from seven to nine and is thus firmly anchored in the active site.

In the structure of the CysK1- aminoacrylate complex, the Cβ carbon atom points into the active site pocket and is the only atom of the intermediate accessible for a nucleophilic attack by hydrogen sulphide through a narrow tunnel extending from the enzyme surface (Fig. 4). The ε-amino group of Lys44 is within 4.3 Å distance of the Cα atom of the aminoacrylate, but can easily adopt a position suitable for the proposed role in protonation of the Cα carbon atom (24,
20) by a simple rotation along the Cδ-Cε carbon bond. K44 is positioned such that protonation of the Cα carbon can only occur from the si face of the intermediate. The channel from the surface leading to the Cβ carbon atom (Fig. 4) directs the nucleophilic attack of the sulphide towards the re face of the intermediate, leading to an anti-addition resulting in the formation of L-cysteine.

In previous attempts to obtain structures of O-acetylserine sulphydrylases with bound substrate and intermediates, active site mutants where the catalytic lysine residue had been replaced by alanine were employed. The structure analysis of such mutants from S. typhimurium (25) and A. thaliana (17) revealed that an external aldimine had formed via a reaction with a methionine residue during protein production (21). This structure represented a mimic of the first intermediate, the species preceding the formation of the α-aminoacylrate in the reaction cycle of OASS (compound II in scheme 1). While the methionine complex of OASS from A. thaliana (crystallized in 4.5 M Na-formate) did not show any major structural rearrangements upon ligand binding, the S. typhimurium (crystallized at pH 7.0 TRIS buffer) enzyme showed a similar conformational change as seen in CysK1 from M. tuberculosis. The external methionine Schiff base adduct in S. typhimurium OASS and the aminoacylrate Schiff base in M. tuberculosis CysK1 overlap well upon superposition, in particular the Cα, Cβ carbon atoms and the carboxyl group (supplementary information, Fig. S2a). Consistent with the key role of T71, N74, S72, and Q144 in recognizing and binding the carboxyl group of the aminoacylrate intermediate as suggested by the structure analysis, replacement of these amino acids in OASS from Arabidopsis thaliana leads to mutants with a 10^3 - 10^5-fold lower catalytic efficiency than wild-type enzyme (17).

It is of interest to note that in tryptophan synthase from S. typhimurium, the only other member of the fold type II PLP enzymes with a known structure of an aminoacylrate external Schiff base (42), the interactions of the intermediate with the enzyme are rather different from that in CysK1 and are not conserved (supplementary information, Fig. S2b). Interactions are made to the side chain of T110 and backbone nitrogens from A112, G113 and Q114. These residues are structurally equivalent to the substrate binding loop (residues 71-75) of OASS enzymes, but are not conserved. A notable difference in the recognition and binding of the aminoacylrate moiety in CysK1 and tryptophan synthase is the absence of Q144 in the latter. The opening of the hydrophobic channel that connects the two active sites in the heterodimeric tryptophan synthase is located at approximately this position and allows access of indole, formed in the α subunit, to the PLP cofactor.

M. tuberculosis CysK1 interacts with a SAT derived C-terminal peptide

OASS associates with SAT to form a bi-enzyme complex, in which the OASS activity is downregulated (43,44). Crystallographic studies of OASS from Haemophilus influenzae and A. thaliana (15, 45) have shown that an essential part of the interactions between the two enzymes most likely involves binding of the C-terminal four residues of SAT to the active site of OASS. We therefore addressed the question whether CysK1 from M. tuberculosis is able to form similar interactions with SAT, and whether binding of a peptide derived from the C-terminus of M. tuberculosis SAT (CysE) would lead to inhibition of CysK1. Addition of a peptide comprising the last four residues (DFSI) of CysE to CysK1 gave a shift in the thermal denaturation curve of the enzyme by 5°C. Inhibition studies showed the competitive nature of the inhibition with a K_i value of 5.0 μM (Fig. 5). These observation indicated formation of an enzyme-peptide complex, and the DFSI peptide was therefore used for co-crystallisation.

The electron density maps calculated from X-ray data collected on crystals of CysK1 obtained in the presence of 1.0 and 4.0 mM DFSI peptide clearly showed that the peptide indeed was bound to CysK1 (Fig. 6a & b). The peptide is bound in the active site cleft between the two domains, and extends from the enzyme surface into the interior of the active site. It blocks the active site cleft completely, and also acts like a wedge preventing complete domain rotation seen in the enzyme-intermediate complex. CysK1 is thus trapped in an open conformation similar to the holo-enzyme.

The peptide interacts with CysK1 through direct hydrogen bonds as well as hydrophobic interactions (figure 6a). Two main chain atoms
of the peptide are engaged in hydrogen bonds. The carbonyl oxygen of S3* (the asterisk denotes residues from the DFSI peptide) interacts with main chain nitrogen atom from G222, and the amide of S3* forms an indirect hydrogen bond to the carbonyl oxygen atom of I223 via a water molecule. Residue D4* is firmly held in place through interactions with the side chain of K215 and the main chain amide of M122. The side chain of F2* binds to a hydrophobic pocket lined by M122, F145, A225 and F227. Finally, I4* is buried in the interior of the active site and packs against F145, G178, G222 and A225. The side chain of F2* binds to a hydrophobic pocket lined by M122, F145, A225 and F227. Finally, I4* is buried in the interior of the active site and packs against F145, G178, G222 and A225. The carboxylate group of I4* occupies the position of the carboxylate of the aminoacrylate moiety in the CysK1 – intermediate complex. Hydrogen bonds are formed to main chain nitrogen atoms from residues of the conserved sequence stretch 71-TSGNT-75, the serine-loop. Additional hydrogen bonds are made to the side chain of Q144.

A comparison of the CysK1-DFSI complex with similar peptide complexes of OASS from H. influenzae and A. thaliana (15,45) reveals that peptide recognition is mediated via side chain interactions in a sequence specific manner. The only enzyme-peptide interaction common to the three enzymes is the binding mode of the conserved C- terminal isoleucine residue via specific hydrogen bonds to its carboxyl group and hydrophobic interactions with the side chain. The residues of the remaining peptide residues are either not conserved or in cases they are, their interactions with the enzyme are different. For instance, S3* corresponds to asparagine in H. influenzae or valine in A. thaliana and the hydrogen bond formed between the side chains of S3* and S72 is specific for the M. tuberculosis Cysk1-peptide complex. The N-terminal peptide residue D1*, although conserved in A. thaliana, has a different conformation and interacts with different enzyme residues.

In conclusion, the strong inhibition and the unique interactions of the peptide with CysK1 suggest this complex to be an attractive scaffold for the rational design of CysK1 inhibitors specific for M. tuberculosis. Furthermore, we have shown that CysK1 is able to interact with the C-terminal peptide from SAT in a similar manner as seen in OASSs, where formation of a bi-enzyme complex has been shown. This suggests that the M. tuberculosis enzyme may form such a complex in vivo and be subject to a similar mechanism of regulation.

Acknowledgements
We gratefully acknowledge access to synchrotron radiation at the ESRF, Grenoble, France and MAX laboratory, Lund University, Sweden. R.S. thanks the David and Astrid Hageléns Foundation for support. This work was supported by the European Commission; contract LSHP-CT-2005-018729.

Footnotes:
1. Abbreviations: OAS, O-acetyl-L-serine; OASS, O-acetylserine sulfhydrylase; r.m.s.d., root mean square deviation; SAT, serine acetyl transferase; MPD, 2-methyl-2,4-pentane-diol; 2. The atomic coordinates for the crystal structures reported are available through the NCBI Protein Data Bank under accession codes 2Q3B (CysK1 holoenzyme), 2Q3D (CysK1–aminoacrylate complex) and 2Q3C (CysK1-peptide complex).

References
### Table I: Data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CysK1 - native</th>
<th>CysK1-Aminoacrylate complex</th>
<th>CysK –DFSI complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beam line</td>
<td>ID14–3 (ESRF)</td>
<td>9/11 (MAX-lab)</td>
<td>ID14–1 (ESRF)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.931</td>
<td>1.000</td>
<td>0.934</td>
</tr>
<tr>
<td>Resolution (Highest resolution shell)</td>
<td>37.9 - 1.80 (1.90 – 1.80)</td>
<td>71.80 - 2.20 (2.32-2.20)</td>
<td>59.66 – 2.10 (2.21 – 2.10)</td>
</tr>
<tr>
<td>No. of observed reflections</td>
<td>310962 (45355)</td>
<td>131874 (18830)</td>
<td>134681 (19632)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>43621 (6244)</td>
<td>24607 (3550)</td>
<td>26571 (3941)</td>
</tr>
<tr>
<td>I/σI</td>
<td>16.4 (3.1)</td>
<td>14.2 (2.4)</td>
<td>11.5 (3.3)</td>
</tr>
<tr>
<td>Completeness</td>
<td>100.0 (100.0)</td>
<td>98.9 (99.8)</td>
<td>93.2 (96.2)</td>
</tr>
<tr>
<td>Rsym</td>
<td>0.094 (0.542)</td>
<td>0.091 (0.499)</td>
<td>0.089 (0.352)</td>
</tr>
<tr>
<td>R-factor from Wilson plot (Å2)</td>
<td>21.2</td>
<td>43.0</td>
<td>31.0</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution interval</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rfactor (%)</td>
<td>17.4 (21.7)</td>
<td>19.4 (30.2)</td>
<td>18.7 (26.7)</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>19.5 (24.0)</td>
<td>24.9 (31.8)</td>
<td>20.7 (29.6)</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>2277</td>
<td>2297</td>
<td>2240</td>
</tr>
<tr>
<td>No. of cofactor / ligand atoms</td>
<td>15</td>
<td>21</td>
<td>49</td>
</tr>
<tr>
<td>N of solvent atoms (MPD, Water, chloride)</td>
<td>236</td>
<td>51</td>
<td>171</td>
</tr>
<tr>
<td>Overall B-factor (Å²)</td>
<td>24.6</td>
<td>56.1</td>
<td>36.5</td>
</tr>
<tr>
<td>Average B-factor protein atoms (Å²)</td>
<td>23.1</td>
<td>56.3</td>
<td>36.0</td>
</tr>
<tr>
<td>Average B-factor ligand atoms (Å²)</td>
<td>16.8</td>
<td>42.5</td>
<td>33.0</td>
</tr>
<tr>
<td>Average B-factor solvent atoms (Å²)</td>
<td>38.6</td>
<td>52.9</td>
<td>42.9</td>
</tr>
<tr>
<td>r.m.s. deviation bonds (Å)</td>
<td>0.01</td>
<td>0.012</td>
<td>0.009</td>
</tr>
<tr>
<td>r.m.s. deviation angles (degrees)</td>
<td>1.28</td>
<td>1.42</td>
<td>1.20</td>
</tr>
<tr>
<td><strong>Ramachandran plot</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of residues in most favorable regions (excl. Gly and Pro)</td>
<td>92.4</td>
<td>91.4</td>
<td>90.5</td>
</tr>
<tr>
<td>Percentage of residues in additional allowed regions (excl. Gly and Pro)</td>
<td>7.6</td>
<td>8.6</td>
<td>9.5</td>
</tr>
</tbody>
</table>
Scheme and figure legends

**Scheme 1.** Schematic representation of the reaction catalyzed by CysK1.

**Figure 1:** UV-Vis spectra of recombinant holo- CysK1 and the complex with the α-aminoacrylate reaction intermediate bound at the active site as external Schiff-base. The spectra were recorded in 25 mM Tris-HCl buffer, pH 8.0, with an enzyme concentration of 3.3 mg/ml. The spectrum of the covalent aminoacrylate complex was obtained after addition of 300 μM O-acetylserine to the enzyme solution.

**Figure 2.** Structure of holo-CysK1. A, Stereoview of the overall structure of CysK1. The N-terminal fragment that forms part of the C-terminal domain, is shown in green, the N-terminal domain in blue, and the C terminal domain in light-brown colours. The cofactor PLP forming the internal Schiff-base with Lysine-44 is shown as stick model. B, The internal aldimin at the active site of CysK1. The PLP cofactor forming the internal Schiff base with Lysine-44 is shown as stick model. The 2Fo-Fc electron density map is contoured at 1.8σ. C, Schematic representation of the coordination of the cofactor and the anion binding site in holo-CysK1. Hydrogen bonds are indicated by dashed lines. D, stereo view of the active site in CysK1. The amino acid sidechains, the bound MPD molecule, and PLP are shown as stick models, and water molecules are shown as red spheres. The green sphere represents the bound chloride ion.

**Figure 3.** Structure of the complex of CysK1 with α-amino acrylate. A, superposition of holo-CysK1 (blue) and the CysK1-aminoacrylate complex (red) illustrating the conformational changes of the N-terminal domain leading to the closure of the active site. The covalent intermediate forming the external Schiff-base with PLP is shown as a stick model (yellow). B, Part of the 2Fo-Fc electron density map, contoured at 1.2σ, showing formation of the α-aminoacrylate intermediate. C, cartoon showing the interactions of the reaction intermediate with enzyme residues in the active site. Hydrogen bonds are indicated by dashed lines. D, stereo view of the active site of the CysK1 – α-aminoacrylate intermediate complex. The red spheres indicate the positions of bound water molecules.

**Figure 4.** Space-filling model of CysK1, highlighting the narrow tunnel that leads to the active site. The only solvent accessible atom is the Cβ carbon atom of the α-aminoacrylate –intermediate, shown in yellow.

**Figure 5.** Inhibition of CysK1 by the DFSI peptide. Double reciprocal plots of the initial rate of the CysK1 activity in the presence of different fixed concentrations of the DFSI peptide illustrating the competitive nature of inhibition. The linear curves were fitted to the data points (average of measurements from three independent reactions) at DFSI concentrations 0.4 μM (●), 4 μM (Δ), 10 μM (○), 15 μM (♦), 20 μM (▲), 40 μM (▲), and without inhibitor (○). In the inset, the values for the slopes of the double reciprocal plots are plotted against the peptide concentrations.

**Figure 6.** Structure of the CysK1-peptide complex. A, Stereo view of the active site of CysK1 with the bound DFSI peptide, shown in green. CysK1 amino acid residues are shown in yellow, water molecules are shown as red spheres. B, Part of the 2Fo-Fc electron density map at the peptide binding site in CysK1, contoured at 1.4σ.
Scheme 1

Figure 1
Figure 2
Figure 3
Figure 6