BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Structure and catalytic mechanism of the evolutionarily unique bacterial chalcone isomerase

Maren Thomsen,[a] Anne Tuukkanen,[b] Jonathan Dickerhoff,[a] Gottfried J. Palm,[a] Hanna Kratzat,[a] Dmitri I. Svergun,[b] Klaus Weisz,[a] Uwe T. Bornscheuer[a] and Winfried Hinrichs[a]*

[a]Institut für Biochemie, Universität Greifswald, Felix-Hausdorff-Strasse 4, 17489 Greifswald, Germany, and [b]EMBL, c/o DESY, Notkestrasse 85, Gebäude 25A, 22603 Hamburg, Germany. *Correspondence e-mail: winfried.hinrichs@uni-greifswald.de

Received 24 December 2014
Accepted 29 January 2015

Edited by Z. Dauter, Argonne National Laboratory, USA

Keywords: chalcone isomerase; flavonoid degradation; *Eubacterium ramulus*; SAXS.

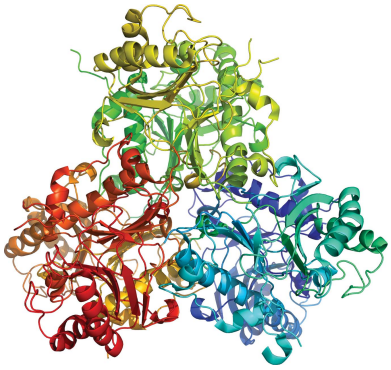
PDB references: bacterial chalcone isomerase, 3zph; 4c9s; SeMet derivative, 4c9t; naringenin complex, 4d06

Supporting information: this article has supporting information at journals.iucr.org/d

Flavonoids represent a large class of secondary metabolites produced by plants. These polyphenolic compounds are well known for their antioxidative abilities, are antimicrobial phytoalexins responsible for flower pigmentation to attract pollinators and, in addition to other properties, are also specific bacterial regulators governing the expression of *Rhizobium* genes involved in root nodulation (Firmin *et al.*, 1986). The bacterial chalcone isomerase (CHI) from *Eubacterium ramulus* catalyses the first step in a flavanone-degradation pathway by ring opening of (2*S*)-naringenin to form naringenin chalcone. The structural biology and enzymology of plant CHIs have been well documented, whereas the existence of bacterial CHIs has only recently been elucidated. This first determination of the structure of a bacterial CHI provides detailed structural insights into the key step of the flavonoid-degradation pathway. The active site could be confirmed by co-crystallization with the substrate (2*S*)-naringenin. The stereochemistry of the proposed mechanism of the isomerase reaction was verified by specific $^{1}H/^{2}H$ isotope exchange observed by ^{1}H NMR experiments and was further supported by mutagenesis studies. The active site is shielded by a flexible lid, the varying structure of which could be modelled in different states of the catalytic cycle using small-angle X-ray scattering data together with the crystallographic structures. Comparison of bacterial CHI with the plant enzyme from *Medicago sativa* reveals that they have unrelated folds, suggesting that the enzyme activity evolved convergently from different ancestor proteins. Despite the lack of any functional relationship, the tertiary structure of the bacterial CHI shows similarities to the ferredoxin-like fold of a chlorite dismutase and the stress-related protein SP1.

1. Introduction

In plants, flavonoids are known for their antioxidative abilities and, in addition to other properties, are also responsible for flower colouration. The first step in flavonoid biosynthesis is the formation of 4,2′,4′,6′-tetrahydroxychalcone from the precursors malonyl-CoA and coumaroyl-CoA catalysed by chalcone synthase (Winkel-Shirley, 2001). Subsequently, the chalcone isomerase (CHI) of plants catalyses stereospecific intramolecular cyclization by an oxa-Michael addition (Nising & Bräse, 2012) to (2*S*)-naringenin. This forms the skeleton for a variety of flavonoids and can be further modified to various bioactive compounds. The three-dimensional structure of CHI from plants (*Medicago sativa*; Jez *et al.*, 2000), their evolutionary history (Ngaki *et al.*, 2012) and the reaction mechanism have been investigated in detail (Jez *et al.*, 2002; Jez & Noel, 2002). A flavonoid-degrading gut bacterium was first observed by Schneider *et al.* (1999) and was later identified as *Eubacterium ramulus* (Braune *et al.*, 2001; Schneider & Blaut, 2000).



The enzymes involved in naringenin degradation were purified and characterized as a chalcone isomerase (Herles *et al.*, 2004), an enoate reductase (ERED; Gall *et al.*, 2014) and a phloretin hydrolase (Schoefer *et al.*, 2004). Based on these results, a degradation pathway for naringenin was postulated (Fig. 1). Whole-genome sequencing of *E. ramulus* identified the chalcone isomerase gene (GenBank accession code KF154734). A *PSI-BLAST* (Altschul *et al.*, 1997) search of sequence databases revealed no identity compared with plant CHI or other characterized polypeptide sequences (<10%). Thus, neither structural homology nor a close evolutionary relationship was expected. In contrast, sequence-database searches identified homologues of plant CHI in gammaproteobacteria and ascomycetes, although their physiological role remains unclear (Gensheimer & Mushegian, 2004). The naringenin catabolic pathway is a multistep enzymatic degradation cascade (Fig. 1), but *in vitro* the bacterial CHI catalyses more efficient ring closure of the chalcone, comparable to the flavanone-synthesis step in plants (Herles *et al.*, 2004). Thus, efficient degradation of flavanones can only occur in a steady-state equilibrium when the chalcone is subsequently eliminated. Recently, biocatalytical conversion of naringenin to phloretin was shown using recombinant CHI and ERED expressed in *E. coli* (Gall *et al.*, 2014). Thus, the bacterial CHI is of general interest for biotechnological applications in stereospecific flavonoid syntheses and flavanone conversion.

We determined the crystal structure of the bacterial CHI from *E. ramulus* at 1.8 Å resolution by single isomorphous replacement with anomalous scattering (SIRAS) *via* selenomethionine substitution. The structure of CHI in complex with the substrate (2*S*)-naringenin at 2.0 Å resolution was solved by molecular replacement. The quaternary structure and the conformational changes of a polypeptide loop that closes the active site as a lid were investigated by small-angle X-ray scattering (SAXS) in solution. The catalytic mechanism of the oxa-Michael addition was verified by $^{1}H/^{2}H$ isotope exchange of the substrate observed by ^{1}H NMR experiments and the mutagenesis of active-site residues. Altogether, these results provide functional and evolutionary insights into the bacterial chalcone isomerase.

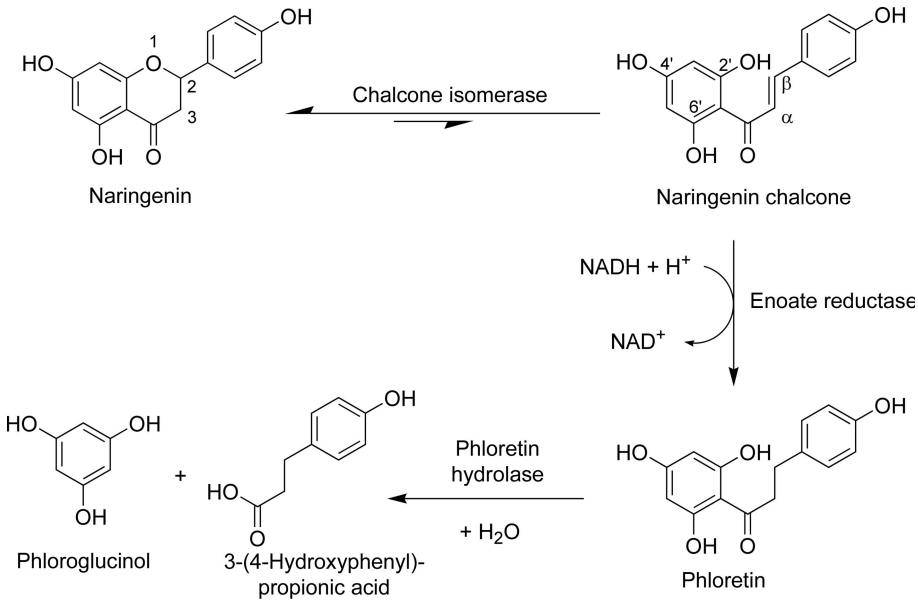


Figure 1
The degradation pathway of naringenin facilitated by *E. ramulus* postulated by Schneider & Blaut (2000) and proven by Gall *et al.* (2014).

2. Materials and methods

2.1. Protein expression

The CHI gene from *E. ramulus* was cloned into the expression vector pET-28b as described by Gall *et al.* (2014). Expression of the CHI gene in *Escherichia coli* Rosetta (DE3) cells was initiated by the inoculation of 5 ml LB_{kan} with frozen stock culture and growth overnight at 37°C. These starter cultures were used to inoculate 500 ml LB_{kan}. Induction of the protein expression with 0.1 m*M* IPTG was carried out when the cells reached an OD_{600} of 1. After 21 h of protein expression at 25°C, the cells were harvested by centrifugation (30 min, 4000*g*, 277 K). The pellets were frozen at 253 K or used directly for protein purification.

Production of selenomethionine-substituted protein was performed in *E. coli* Rosetta (DE3) cells by metabolic inhibition (Doublié, 1997). The purification protocol was the same as that for the native protein.

2.2. Protein purification

The protein-purification protocol for the recombinantly expressed protein was modified slightly compared with the isolation of the protein from the original host *E. ramulus* (Herles *et al.*, 2004). The cell pellet was resuspended in 50 m*M* sodium phosphate buffer pH 6.8 and sonicated with a Sonopuls sonicator (Bandelin, Germany) on ice. After centrifugation, the supernatant was used for protein purification. The extract was bound to a DEAE-Sepharose column (GE Healthcare) which was equilibrated with 50 m*M* sodium phosphate buffer pH 6.8 (buffer *A*). Nonbound proteins were washed out with three column volumes of buffer *A*. Bound proteins were eluted using a linear gradient of 0–0.15 *M* KCl in buffer *A* followed by two column volumes of 1 *M* KCl in buffer *A*.

Solid ammonium sulfate was added on ice to the combined fractions, resulting in a final concentration of 1.5 *M*. This solution was applied onto a butyl-Sepharose column (GE Healthcare) equilibrated with 1.5 *M* ammonium sulfate in buffer *A*. Nonbound protein was eluted with two column volumes of 1.5 *M* ammonium sulfate in buffer *A*. Elution of the bound protein was carried out with a linear gradient of

Table 1
Data-collection and processing statistics.

Values in parentheses are for the last resolution shell.

Data set	Native	Selenomethione derivative	Native in-house	Naringenin complex
X-ray source	Beamline 14.1, BESSY II	Beamline 14.1, BESSY II	MicroMax-007, Rigaku	Beamline 14.1, BESSY II
Wavelength (Å)	0.97935	0.97935	1.5418	0.97857
Temperature (K)	100	100	110	100
Space group	$I2_12_12_1$	$I2_12_12_1$	$I2_12_12_1$	$I2_12_12_1$
Unit-cell parameters (Å)				
a	171.2	172.3	177.1	181.2
b	192.2	192.6	203.1	188.0
c	204.6	203.8	206.1	196.2
Resolution (Å)	1.8 (1.91–1.80)	1.98 (2.10–1.98)	2.8 (2.95–2.80)	2.0 (2.12–2.00)
Unique reflections	309268	460230†	91147	436945†
Multiplicity	4.15 (4.08)	7.72 (7.30)	4.10 (4.00)	3.49 (3.41)
R_{merge}‡ (%)	9.9 (70.6)	20.1 (125.9)	19.2 (98.7)	12.1 (61.7)
$CC_{1/2}$	0.998 (0.748)	0.995 (0.626)	0.977 (0.794)	0.993 (0.649)
Mean $I/\sigma(I)$	12.25 (2.14)	11.41 (1.75)	5.40 (1.20)	8.94 (1.93)
Completeness (%)	99.6 (97.6)	99.7 (98.1)	99.6 (99.9)	99.5 (98.4)
Wilson *B* factor (Å^2)	28.5	29.7	41.5	27.2

† Friedel pairs not merged. ‡ $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity of multiple measurements.

Table 2
Summary of SIRAS phasing.

SHELXD	
CC(all)	0.367
CC(weak)	0.266
PATFOM	1.44
SHELXE	
Overall CC between E_{obs} and E_{calc} (%)	27.27
Mean FOM	0.788
MapCC	0.942
Pseudo-free CC	0.833

1.5–0.3 *M* ammonium sulfate followed by one column volume of 0.3 *M* ammonium sulfate and a final linear gradient of 0.3–0 *M* ammonium sulfate.

The fractions containing chalcone isomerase activity were pooled and were injected onto a Superdex 200 column (HiLoad 16/60, GE Healthcare) for gel filtration. Elution of the chalcone isomerase was performed with buffer *A* at a flow rate of 0.3 ml min^{-1}. An average of 4–6 mg protein was obtained from 250 ml cultivation volume.

2.3. Crystallization and data collection

Crystals were obtained by the hanging-drop method using equal volumes of protein and reservoir solutions followed by one round of streak-seeding with cat whiskers (Bergfors, 2003). The protein concentration used was 7.5 mg ml^{-1} and was determined spectrophotometrically at 280 nm. The first small crystals were obtained with a reservoir solution consisting of 2.0 *M* ammonium sulfate, 0.2 *M* NaCl, 0.1 *M* HEPES pH 7.5. To improve the quality of the crystals, they were seeded with a cat whisker into drops with a reservoir consisting of 1.7 *M* ammonium sulfate, 0.2 *M* NaCl, 0.1 *M* HEPES pH 7.5. (*2S*)-Naringenin-bound crystals were obtained using 2.0 *M* ammonium sulfate, 0.2 *M* NaCl, 0.1 *M* HEPES pH 7.5, 6% ethanol, 370 µ*M* naringenin chalcone as the crystallization condition. Crystallization was performed at 293 K. X-ray diffraction data were collected at 100 K on beamline 14.1 at the BESSY II synchrotron, Berlin, Germany. For all crystals a cryoprotectant solution consisting of 22% glycerol, 2.2 *M* ammonium sulfate, 0.2 *M* NaCl, 0.1 *M* HEPES pH 7.5 was used. The crystals of native CHI and the SeMet derivative were isomorphous and belonged to the orthorhombic space group $I2_12_12_1$ or $I222$. All diffraction images were processed with *XDS* (Kabsch, 2010) using the *XDSapp* graphical user interface (Krug *et al.*, 2012). In-house data were collected from a cryoprotected native crystal at 100 K using a Saturn 92 CCD detector mounted on a MicroMax-007 rotating-anode source (Rigaku MSC). The data were processed with *MOSFLM* (Leslie & Powell, 2007) and were scaled with *SCALA* (Evans, 2006). The structure-factor amplitudes were calculated using *TRUNCATE* (French & Wilson, 1978). Data-collection and processing statistics are summarized in Table 1.

2.4. Structure determination and refinement

The self-rotation function (*MOLREP*; Vagin & Teplyakov, 2010; Winn *et al.*, 2011) revealed perpendicular twofold and threefold axes, suggesting D_3 symmetry for a hexamer in the asymmetric unit (Supplementary Fig. S1). The phases for the selenomethionine-derivative data were calculated for both possible space groups with *SHELXC*/*D*/*E* (Sheldrick, 2010) using the *HKL2MAP* GUI (Pape & Schneider, 2004), but only space group $I2_12_12_1$ resulted in interpretable electron-density maps. The resolution for heavy-atom identification with *SHELXD* was limited to 2.3 Å, resulting in a CC(all/weak) of 0.367/0.266 and a PATFOM of 1.44. In the asymmetric unit, 54 selenium positions were identified for the nine selenomethionine residues found in the bacterial CHI sequence (Supplementary Fig. S1), assuming post-translational cleavage of the N-terminal methionine. *SHELXE* determined the original hand as correct, with a pseudo-free CC of 0.832 compared with 0.169 for the inverted hand, and calculated phases with combined chain tracing with data extending to 1.8 Å resolution. Owing to the high quality of the obtained electron-density maps, no further density modification was necessary or performed (Supplementary Fig. S2). Experimental phasing statistics are listed in Table 2.

The model of one monomer was built manually into the electron density using *Coot* (Emsley & Cowtan, 2004). Six noncrystallographically related monomers were placed into the asymmetric unit and refined with *REFMAC*5 including TLS segments (Murshudov *et al.*, 2011). Sixfold NCS

averaging was not applied because of the high quality of the electron-density maps and to avoid flattening of distinct differences of the monomers, for example side-chain conformations, especially in the active site.

Owing to the differences in the unit-cell parameters, the data set for the substrate complex and the data set collected in-house could only be solved by molecular replacement using one monomer of the solved structure as a model. Molecular replacement was carried out with *Phaser* (McCoy *et al.*, 2007) and refinement was carried out with *REFMAC*5 as described above. The quality of the refined protein models was validated using *MolProbity* (Chen *et al.*, 2010). Molecular graphics were prepared with *PyMOL* (DeLano, 2002). Refinement statistics are listed in Table 3.

Table 3
Refinement statistics.

Data set	Native	Selenomethionine derivative	Native in-house	Naringenin complex
Resolution (Å)	48.1–1.8	48.5–2.0	20.1–2.8	49.2–2.0
Working/test reflections	293803	222651	86192	212837
R/R_{free}† (%)	13.2/15.2	13.4/15.6	24.8/29.6	14.02/16.18
No. of protein residues	1538	1506	1564	1623
No. of water molecules	2500	2086	1061	1893
No. of chloride ions	18	15	4	17
No. of sulfate ions	1	1	0	0
No. of glycerol molecules	6	6	5	2
No. of substrate molecules	—	—	—	8
R.m.s.d. from ideality				
Bond lengths (Å)	0.0157	0.0185	0.0119	0.0190
Bond angles (°)	1.707	1.823	1.551	1.938
Average *B* factor (Å^2)	22.5	20.5	36.3	24.5
Ramachandran statistics‡ (%)				
Most favoured regions	98.2	98.2	97.1	98.3
Outliers	0	0.2	0.1	0
PDB code	4c9s	4c9t	3zph	4d06

† $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R_{free} is analogous to the *R* factor except that the summation is over 5% of reflections that were not included in refinement. ‡ Categories were defined by *MolProbity* (Chen *et al.*, 2010).

2.5. Mutant generation

The mutations for determining the catalytic residue were introduced by PCR with one primer carrying the desired mutation (for further details, see the Supporting Information) and a primer binding at the T7 terminator sequence of the pET vector. Using this PCR product, a MegaWhoP-PCR (Miyazaki, 2003) was performed. The nonmethylated DNA was digested by DpnI and then transformed into *E. coli* BL21 (DE3). The deletion of the lid and C-terminus (amino-acid residues 109–130) were performed using the FastCloning method (Li *et al.*, 2011). The nonmethylated DNA was digested by DpnI and transformed into *E. coli* BL21 (DE3). All primers used are listed in Supplementary Table S1.

2.6. Enzyme assay

The chalcone isomerase activity was measured spectrophotometrically as result of the conversion of naringenin chalcone to naringenin at 368 nm ($\varepsilon = 29\,068\ M^{-1}\ cm^{-1}$). The reaction mixture (0.5 ml, 2 mm cuvette) consisted of 50 m*M* sodium phosphate buffer pH 6.4, 1 mg BSA (bovine serum albumin), a varying naringenin chalcone concentration (dissolved in DMSO stock solution) and an appropriate amount of enzyme (Supplementary Fig. S3). The activity was not affected by DMSO concentrations of up to 5%.

2.7. Melting point

The melting points (T_m) were determined by circular-dichroism spectroscopy (CD). The purified enzymes were subjected to CD measured at 222 nm using a Jasco V-650 with a heating rate of 0.5°C min^{-1} in the range 20–80°C (Supplementary Table S2).

2.8. ^{1}H NMR experiments

For mechanistic studies, the enzyme was pre-incubated in ^2H_2O for 1 h. The enzyme assay mentioned above was performed in the deuterated solvent and the product was extracted twice with ethyl acetate. The organic solvent was evaporated with a SpeedVac and the residual product was dissolved in DMSO-d_6 for NMR studies. The ^{1}H NMR spectra (Supplementary Fig. S4) were collected with a Bruker Avance 600 MHz spectrometer equipped with an inverse $^1H/^{13}C/^{15}N/^{31}P$ quadruple-resonance cryoprobe head and *z*-field gradients. The data were processed and analyzed using the *TopSpin* 3.1 software (Bruker).

2.9. Small-angle X-ray scattering

Synchrotron-radiation X-ray scattering data for CHI with and without naringenin chalcone (the substrate concentration was varied between 2 µ*M* and 100 m*M*) as well as the CHI_Δlid construct were collected on the EMBL Hamburg P12 beamline at the PETRA III storage ring at DESY, Hamburg, Germany using a PILATUS 2M pixel detector, a sample-to-detector distance of 3.1 m and a wavelength of 1.24 Å, covering the momentum-transfer range $0.008 < s < 0.47$ Å^{-1}. For each construct, several solute concentrations in the range 0.8–8.3 mg ml^{-1} in 50 m*M* sodium phosphate buffer pH 6.8 were measured. To monitor the radiation damage, 20 successive 50 ms exposures of protein solutions were compared and no significant changes were observed. The data were normalized to the intensity of the transmitted beam and radially averaged. The scattering of the buffer without protein solute was subtracted and the difference curves were scaled for protein concentration. The low-angle data measured at lower protein concentrations were extrapolated to infinite dilution and merged with the higher concentration data to yield the final composite scattering curves. The data-processing steps were performed using *PRIMUS* (Konarev *et al.*, 2003) from the *ATSAS* software suite (Petoukhov *et al.*, 2007).

The radius of gyration R_g and the forward scattering intensity $I(0)$ were independently determined using Guinier analysis (Guinier, 1939) and the indirect Fourier transformation approach of *GNOM* (Svergun, 1992). Additionally, the maximum particle dimension D_{max} was obtained from the latter approach. The molecular masses of protein constructs (MM_{SAXS}) were calculated by comparing the extrapolated

forward scattering intensities with that of a reference sample of bovine serum albumin (BSA; MM_{ref} = 66 kDa) together with concentration information. The excluded volume of the hydrated protein V_p was obtained with *DATPOROD* (Petoukhov *et al.*, 2012) and was used to extract an independent estimate of the molecular mass (MM_{POROD}). The statistics of data collection are summarized in Supplementary Table S4.

The scattering patterns from our high-resolution structural models were calculated using *CRYSOL* (Svergun *et al.*, 1995) and were used to determine the fit of these models to the experimental scattering data. Given the atomic coordinates of a structural model, the program minimizes the discrepancy χ between the experimental and theoretical scattering intensities by adjusting the excluded volume of the particle and the contrast of the hydration layer.

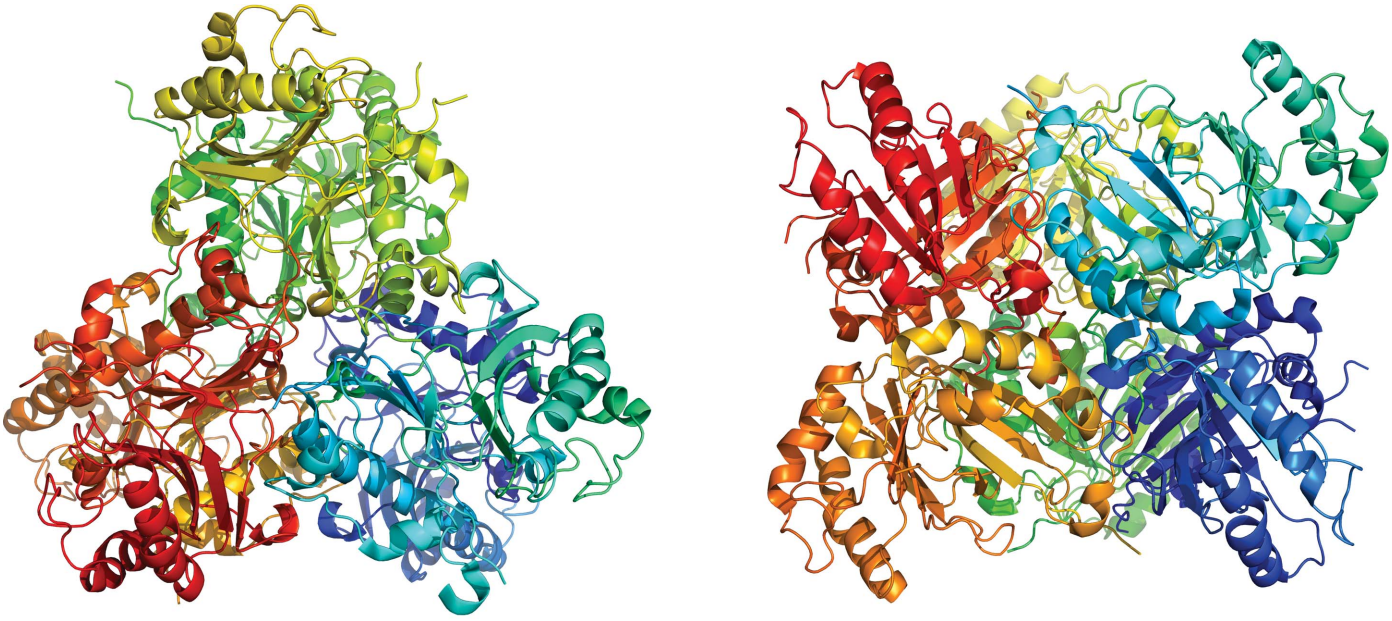


Figure 2
D_3 symmetry of the quaternary structure of the hexameric CHI. (*a*) View down the local threefold axis, (*b*) the same view rotated horizontally by 90°.

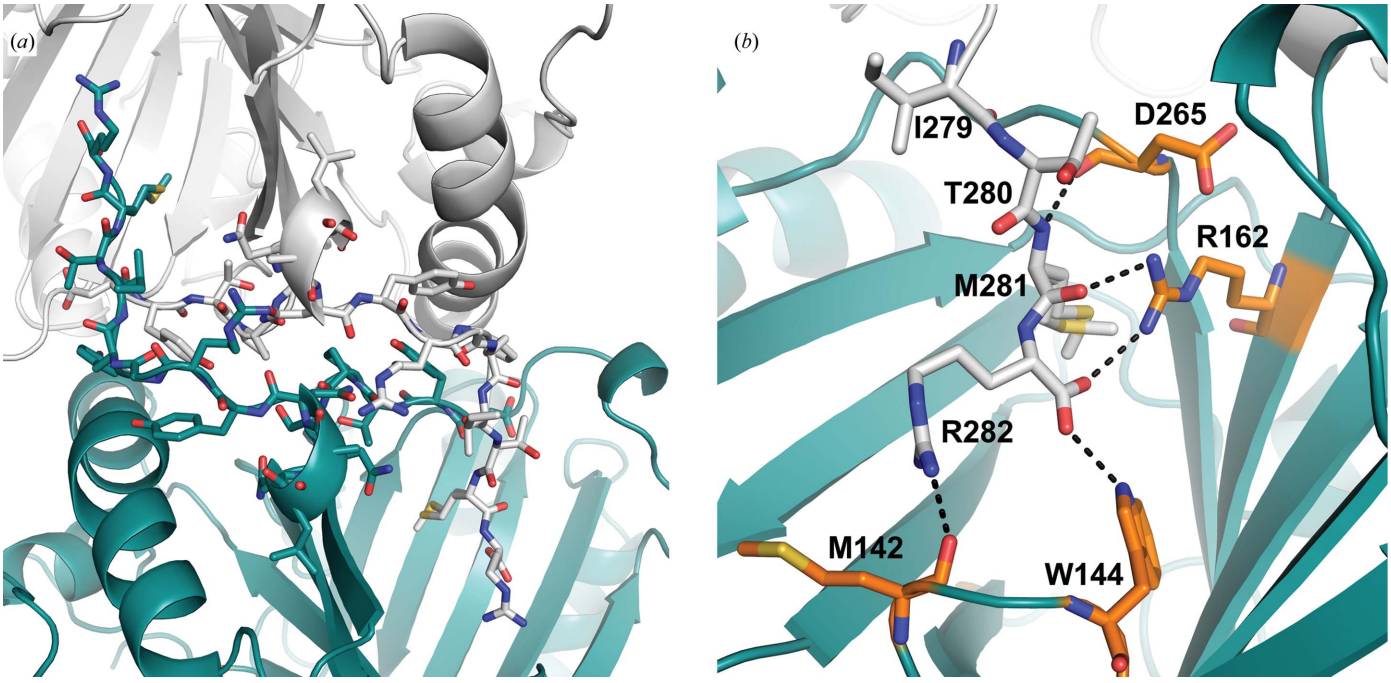


Figure 3
(*a*) Dimer formation facilitated by the C-termini. Overall view down the local twofold axis of the dimerization through reciprocal embedment of the C-terminus into the protein core. Monomer *A* is shown in cyan and monomer *B* in grey. (*b*) Close-up view of the polar interactions formed by the C-terminus. Monomer *A* is shown in cyan (yellow C atoms) and the C-terminus of monomer *B* is shown with light grey C atoms.

2.10. SAXS modelling

Flexibility analysis of the lid regions of CHI in solution was conducted using the available crystallographic models as input for the ensemble-optimization method (EOM). This approach seeks to best fit the experimental scattering profile with an ensemble of conformations (Bernadó *et al.*, 2007). The protein was modelled as a hexamer; the lid regions of each subunit (residues 107–130) were defined as flexible and their possible conformations were modelled with *RANCH* (Bernadó *et al.*, 2007), producing 10 000 random configurations, while the rest of the hexameric protein was kept fixed. A genetic algorithm was employed to find the set of conformations that best fitted the SAXS data. The structures selected from the random pool of structures were analysed with respect to the R_g distribution. The SAXS data and their interpretation are summarized in Supplementary Figs. S8–S11. The SAXS data and models of CHI with and without substrate, and CHI_Δlid have been deposited in the Small-Angle Scattering Biological Data Bank (http://www.sasbdb.org) with entries SASDAL6, SASDAM6 and SASDAN6, respectively.

3. Results

3.1. Overall structure of the bacterial CHI

We carried out X-ray crystal structure analyses of the bacterial CHI from *E. ramulus* and its (2*S*)-naringenin complex at 1.8 and 2.0 Å resolution, respectively. The bacterial CHI crystallized in space group $I2_12_12_1$, with six monomeric subunits (*A*–*F*) in the asymmetric unit constituting a quaternary structure with D_3 symmetry (Fig. 2), which we verified as the native form in solution by SAXS. Each polypeptide chain contains 282 amino-acid residues with a molecular weight of 32.4 kDa.

Detailed analyses of the protein interfaces with *PISA* (Krissinel & Henrick, 2007) revealed that the buried surface area upon dimer formation (2100 Å^2) is much higher than that upon trimer formation (910 Å^2). This observation is confirmed by the fact that the last four residues of the C-terminus (residues 279–282) are buried in the protein core of the other subunit for dimer formation and contribute 690 Å^2 to the buried surface for dimer formation (see Fig. 3). Truncation of the last five residues led to a variant (CHI_ΔC-term) that has reduced thermal stability (ΔT_m = 9°C) and almost no enzymatic activity (Supplementary Table S2 and Supplementary Fig. S3*c*).

The tertiary structure belongs to the superfamily of dimeric $\alpha+\beta$ barrels (Wilson *et al.*, 2009) and is divided into two very similar domains: the catalytic domain (residues 1–143) and the solvent-exposed domain (residues 144–266) (Fig. 4). Both domains have a bacterial ferredoxin-like fold (Abendroth *et al.*, 2004) based on two antiparallel β-sheets with four and five β-strands, respectively, each with two long α-helices on top. In the catalytic domain the antiparallel β-sheet is formed by β-strands 2↑3↓1↑4↓, and the topology-related β-strands in the solvent-exposed domain are 7↑8↓6↑9↓. Together, both domains form a β-sheet sandwich with the antiparallel β-sheets rotated with respect to each other by about 90°. The catalytic domain is strongly involved in dimer as well as trimer formation, whereas the solvent-exposed domain shows almost no interaction with other subunits, except for the small C-terminal extension (residues 267–282) responsible for dimer formation as mentioned above. A notable feature of the catalytic domain is that in most of the structures the electron density for a putative lid (amino-acid residues 107–130) is not observed (for details, see §3.3).

3.2. Identification of the active site and reaction mechanism

Owing to the fact that the bacterial CHI has no sequence homology to proteins in the PDB, there was no information about possible catalytic residues and the position of the active site. To identify the active site, co-crystallization with the substrate naringenin chalcone was performed. In the resulting

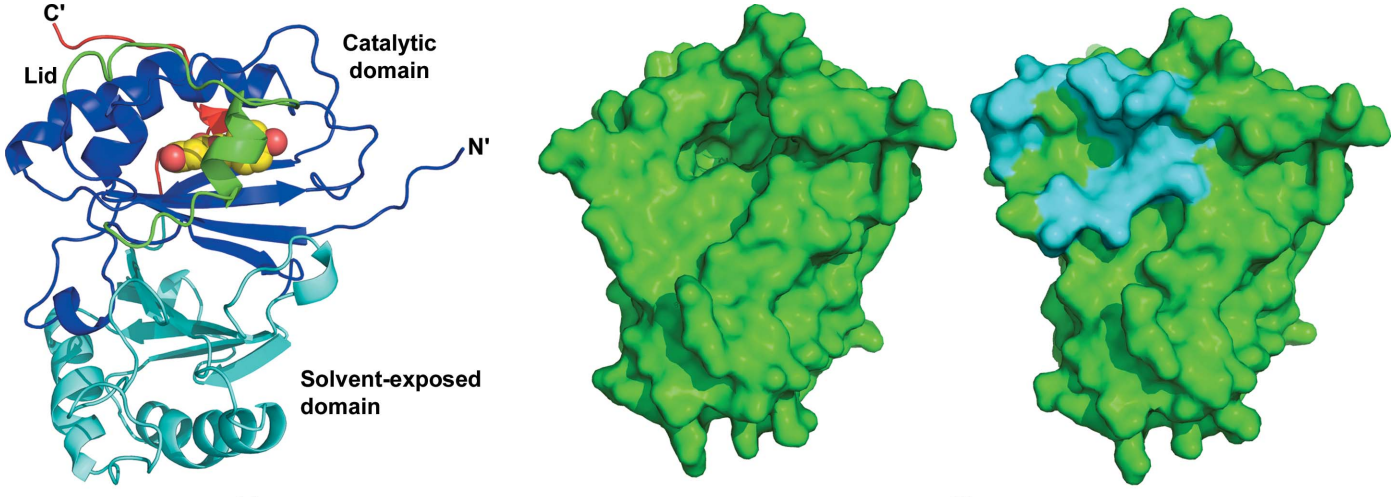


Figure 4
(*a*) The structure of a CHI monomer with colour-coded domains: the catalytic domain which hosts the active site is shown in blue, with the additional lid which covers the active site highlighted in green, the solvent-exposed domain is in cyan and the C-terminus is in red. (*b*) Comparison of the monomer surface in open conformation (left) with the closed conformation (right). The lid is highlighted in cyan and shields the active site from the solvent.

X-ray structure the hexameric quaternary structure shows three dimers (*AB*, *CD* and *EF*) in a different state with respect to the active site. In each subunit in one trimer (chains *A*, *C* and *E*) (2*S*)-naringenin could be found in the active site shielded by the lid structure (amino-acid residues 107–130), which is part of the catalytic domain. In all subunits of the other trimeric half of the quaternary structure (chains *B*, *D* and *F*) the lid structures are not in the closed conformation and are not observed in the electron-density maps. The active sites with open entry tunnels are occupied by naringenin and naringenin chalcone owing to the artificially high concentration used for co-crystallization. The electron-density maps show a complicated disorder of substrate and product molecules in the active site and the entry tunnel. This is interpreted as a superposition of snapshots of substrate-binding and product-release steps (Supplementary Fig. S5). In the closed conformation the substrate bound in a sensible way between β-strands 1 and 3 and α-helices 1, 3 and 4 (Fig. 4*a*) comparable to other examples of active sites in this superfamily, for example chlorite dismutase (de Geus *et al.*, 2009). With distances of between 3.2 and 2.5 Å from the substrate, Gln40, Tyr48, Gln69, Thr71, Asp79 and Gln101 are responsible for substrate binding. The guanidinium moiety of Arg125 positions the phenyl ring of the substrate by π-stacking (3.1–3.4 Å), but also by a hydrogen bond to the ring O1 atom of the flavanone (2.9 Å) (Fig. 5). In the substrate-free enzyme Thr71 and Gln101 are hydrogen-bonded to one glycerol molecule of the cryoprotectant, which replaces the phenolic 4′,6′-dihydroxy fragment of the chalcone (Supplementary Fig. S6). The isomerization mechanism of plant CHI relies on a catalytic water molecule and the bound substrate is solvent-accessible (Jez *et al.*, 2000, 2002; Jez & Noel, 2002). In contrast, in the active site of bacterial CHI no water molecule is observed around naringenin. The possible catalytic residues His73, His33 and Tyr48 were investigated by mutagenesis studies. The enzymatic efficiency (k_{cat}/K_m) of the wild-type bacterial CHI is unchanged in the Y48A and H73A variants (Table 4). The efficiency of the H33A and H33Q variants decreases by about three and four orders of magnitude, respectively. It is probable that the H33A variant has residual activity caused by an imidazole-replacing water molecule capable of proton transfer. This is in agreement with the almost inactive H33Q and H33E variants. The side chains of Gln33 and Glu33 have approximately the same steric requirements as His33 and may interact in a hydrogen-bonding network, but Gln33 is unable to mediate an acid–base mechanism. This is in line with

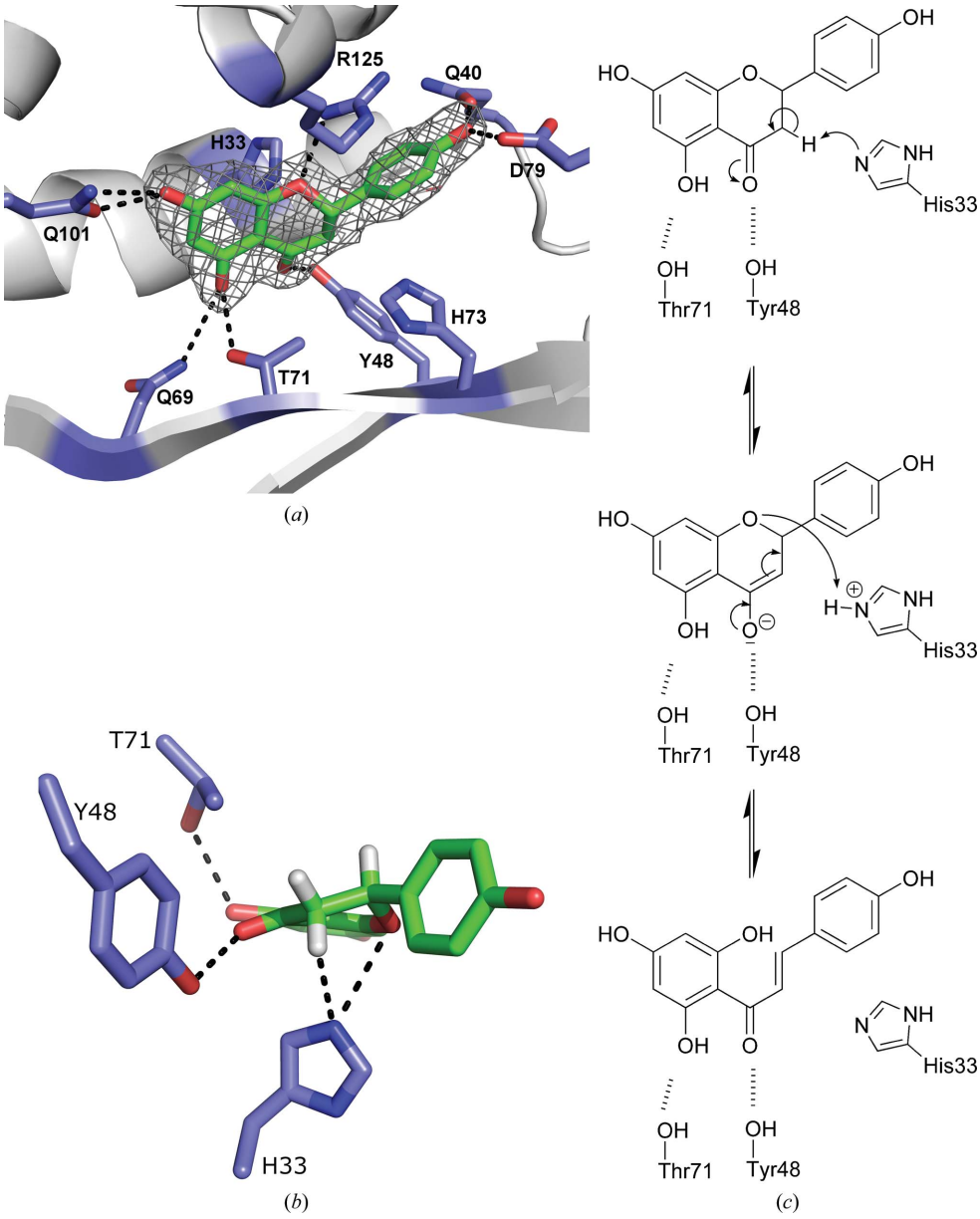


Figure 5
Active site, substrate recognition and mechanism of the reaction of CHI. (*a*) (2*S*)-Naringenin binding in the active site to amino-acid residues labelled with one-letter codes. Hydrogen bonds are shown as dashed lines. The $2F_{obs} - F_{calc}$ electron-density map for (2*S*)-naringenin is contoured at the 1σ level. (*b*) The ring-opening reaction forming the naringenin chalcone is initiated by the imidazole of His33 abstracting the proton H3b. The reversible reaction relies on the hydrogen-bonding (dashed lines) pattern of His33 with C3 and O1. (*c*) Scheme of the reaction mechanism of the reversible Michael addition catalysed by His33.

the slightly increased efficiency of the H33E variant. Based on these results, a general acid–base reaction mechanism for the degradation of naringenin to naringenin chalcone is proposed (Fig. 5*c*). The reaction is initiated by deprotonation at C3. The intermediate enolate can be stabilized by Tyr48 and Thr71, which act as an oxyanion hole. Subsequent protonation at O2′ promotes ring opening and forms the chalcone. To test this proposed reaction mechanism further, the enzyme assay was performed in ^2H_2O to investigate the formed product *via* NMR. These studies showed very distinctly that (de)protonation takes place in the *pro-S* position at C3 (Supplementary Fig. S4). This is consistent with the mutagenesis studies and the structure with bound (2*S*)-naringenin, as the axial H atom at C3 directly points towards His33 N^{ε} (Fig. 5*b*). A superposition of the substrates bound to the active sites of the plant and bacterial CHI is shown in Supplementary Fig. S7.

3.3. Lid formation

In all structures with an open conformation the CHI monomers have a solvent-exposed cleft occupied by glycerol molecules (Supplementary Fig. S6). The high-resolution data show well defined electron density for one glycerol molecule and further weak density for a second glycerol molecule, which was not observed using the in-house diffraction data. In one monomer (chain *B*) of the in-house diffraction data set the highly flexible lid structure is observed in the electron-density map shielding the active site from the solvent without bound substrate. In this case the cavity is not occupied by glycerol or specific bound water molecules. A comparison of the open and the closed conformation shows that with the exception of the movement of the loop, no conformational changes in the protein backbone are observed (Fig. 6). However, the lid closure positions the C^{α} atom of Arg125 so close to Phe135 that the phenyl group is forced to rotate by about 90°. Furthermore, upon lid closure Ile114 displaces Phe93, which in turn moves Glu91, which is then within a reasonable distance to form a salt bridge to Arg125. In addition, the distance from Glu91 $O^{\varepsilon 1/2}$ to His33 N^{ε} also decreases from 4.5 to 3.1 Å such that the newly formed hydrogen bond attracts His33. This moves His33 into a better position to be able to deprotonate C3 and protonate O1 (Fig. 6). These conformational changes derive only from lid closure and are not induced by substrate binding, as the same events can also be observed in the closed conformation without ligand (Fig. 8). An important residue for determining the lid conformation is Glu131, which shows a strong salt bridge to Lys253 (2.8 Å) in the open conformation. In the closed conformation Glu131 flips out and is exposed to the solvent.

Table 4
Kinetic parameters of CHI and its variants.

K_m and v_{max} were calculated using the Michaelis–Menten equation; k_{cat} was calculated using the determined molecular weight of 194 kDa. Data for plant CHI and the uncatalysed reaction from Jez & Noel (2002) are included for comparison.

Enzyme	K_m (μ*M*)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
CHI	35.4 ± 3.5	1142 ± 29	3.2×10^7
CHI_Δlid	199.5 ± 23.3	1177 ± 63	5.9×10^6
CHI_R125A	56.7 ± 7.7	1379 ± 64	2.4×10^7
CHI_H73A	43.9 ± 3.5	1313 ± 28	2.8×10^7
CHI_Y48F	14.3 ± 1.1	784 ± 13	5.5×10^7
CHI_H33A	20.3 ± 1.6	0.6 ± 0.01	2.9×10^4
CHI_H33E	131.9 ± 16.8	0.2 ± 0.013	1.6×10^3
CHI_H33Q	79.8 ± 8.4	0.07 ± 0.003	0.9×10^3
Plant CHI	112 ± 28	186.3 ± 23	1.66×10^6
Uncatalysed		0.78×10^{-4}	

Our SAXS data reveal significant differences between the substrate-free CHI, its naringenin complex and the lid-deletion variant CHI_Δlid. The SAXS data for the substrate-free CHI hexamer could be optimally fitted with an ensemble of hexamer structures in which some of the monomers had partially or fully closed lids (discrepancy $\chi = 1.4$; Fig. 7). This is an indication of an equilibrium between the open and closed states and of a low energy barrier for the mechanism of lid closure. The finding is further supported by the closed lid conformation observed in one subunit of the substrate-free hexamer observed in a data set at 2.8 Å resolution with similar crystallization buffer conditions (PDB entry 3zph). The overall structure of the substrate-bound CHI is more compact than that of the native protein, which is observed as differences in the model-free parameters R_g and D_{max} derived from the SAXS data. There is also a change of the minimum of the SAXS profile of the naringenin-bound complex around $s = 0.1$ Å^{-1}, indicating further a more globular structure

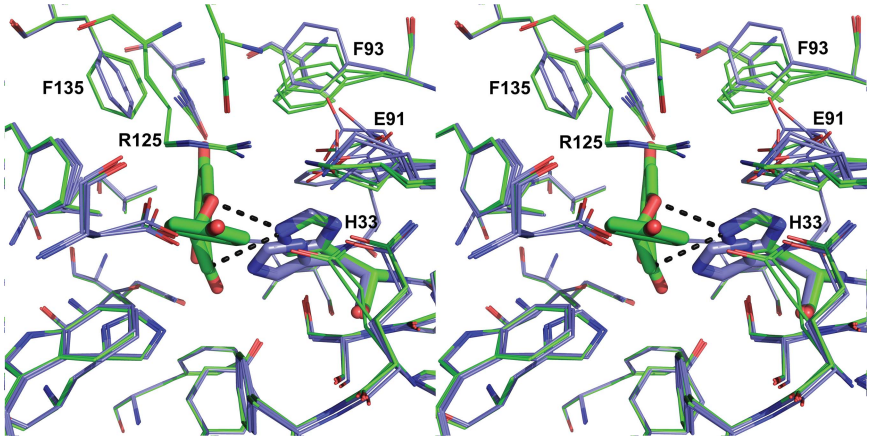


Figure 6
(2*S*)-Naringenin recognition in the active site of CHI. Stereoview of superposition of the active sites of all CHI monomers with bound substrate (green wire model) and without substrate (blue wire model). Conformational changes upon substrate binding are not observed for most of the amino-acid side chains (for example Tyr48). In the substrate-free state the imidazole of His33 (highlighted as a blue stick model) has two alternative orientations. Ligand binding [stick model of (2*S*)-naringenin] requires a distinct conformation of His33 to form the catalytically important hydrogen bonds (dashed lines) to the ring O atom and C3 of (2*S*)-naringenin, respectively (stick model). Minor changes of Gln69, Phe93 and Phe135 are caused by lid closure upon substrate binding.

compared with the substrate-free CHI. A model of the naringenin-bound hexamer constructed using the corresponding X-ray crystallographic structure of a trimer with all lids closed (PDB entry 4d06) and a symmetry operation yields an excellent fit against the substrate-bound CHI SAXS data. This suggests the closure of all lids in the substrate-bound state. The SAXS data of the lid-deletion variant CHI_Δlid are consistent with the hexameric X-ray crystallographic structure which is lacking electron density for the lids (PDB entry 4c9s).

Superposition of all monomers of CHI structures with and without bound substrate suggests a possible mechanism of stepwise substrate binding. The tertiary structure has a remarkable rigid polypeptide scaffold, with the exception of large conformational changes of the lid structure. Only the side chains of the catalytically active residues move for substrate binding and product release in the preformed active site (Fig. 6). The major advantage of lid closure is probably the 19-fold enhanced enzymatic efficiency of naringenin formation with respect to plant CHI (Jez *et al.*, 2002). Deletion of the lid structure led to a solvent-exposed active site and resulted in an active phenotype with an efficiency that was decreased about fivefold but was still fourfold higher than that of the plant CHI. The specific hydrogen bonding and π-stacking of the substrate with Arg125 is not required for efficient catalysis, because the R125A variant shows almost unchanged catalytic efficiency (Table 4). Thus, the lid structure enhances efficient catalysis by reducing the dielectric constant of the active site, but additional sequence-specific interactions are ruled out. The lack of specific strong interactions between the lid structure and the protein core is in agreement with the findings from SAXS that the substrate-free CHI data can be best modelled by an ensemble of structures with an equilibrium state of open/closed lids and a low activation-energy barrier for the lid opening.

Shielding of the active site against the solvent to prevent side reactions is obviously not the case in comparison with the solvent-exposed active site of plant CHI (Jez *et al.*, 2000). The lid closure provokes some allosteric changes of side chains, causing an in-plane shift of the imidazole centre of His33 towards the substrate by about 2 Å (Fig. 6). These conformational changes derive only from lid closure, but are not induced by substrate binding, because the same events are also observed in the closed conformation without ligand.

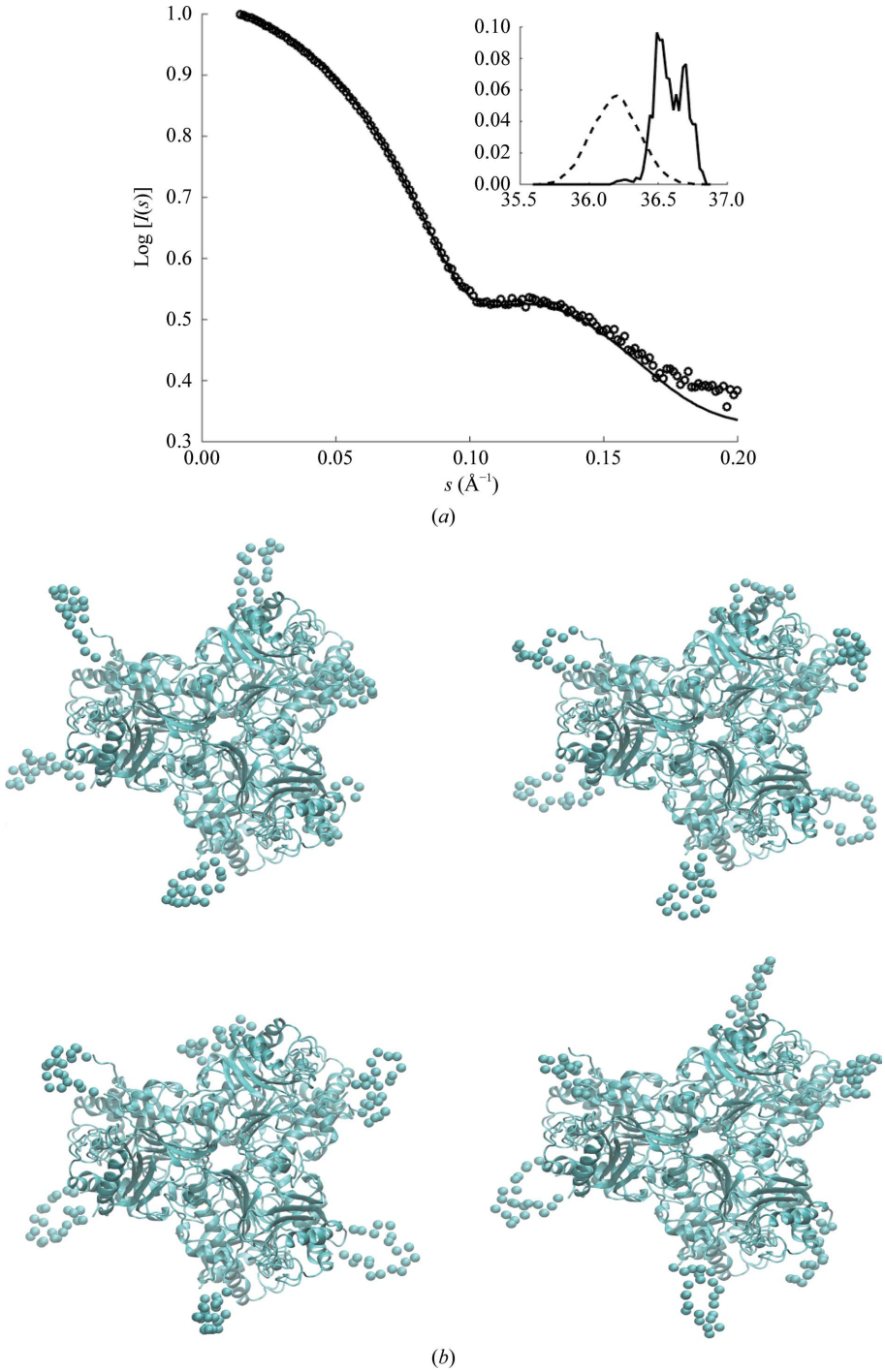


Figure 7
Ensemble-optimization method (EOM) fit of the ligand-free CHI. (*a*) The SAXS data of CHI (circles) are optimally fitted (solid line) by an ensemble of hexamers with some monomers in open lid conformations and some in closed lid conformations. The intensity is shown as a function of momentum transfer ($s = 4\pi\sin\theta/\lambda$, where λ is the X-ray wavelength and 2θ is the scattering angle). Inset: R_g distribution of the random pool (dashed line) and of the selected ensemble (solid line). (*b*) Examples of CHI conformations with the C^{α} positions of the lid residues represented as spheres.

3.4. Secondary-structure element superposition

3.4.1. Bacterial CHI *versus* plant CHI. To investigate the evolutionary relationship of CHIs from plants and bacteria with respect to tertiary structure, superposition of the secondary-structure elements was performed using *PDBeFold* (Krissinel & Henrick, 2004). Both structures are based on an antiparallel β-sheet with two helices on top. At a first glance, the solvent-exposed domains of the bacterial CHI and the plant enzyme from *M. sativa* (PDB entry 1eyq; Jez *et al.*, 2000) can be superimposed. However, β-sheet–α-helix arrangements are very common structural motifs in protein folding and the topology diagrams show rather different connections between the individual β-strands (Figs. 8*a* and 8*d*). We conclude that it is very unlikely that these enzymes have a close evolutionary history. Accordingly, the postulated ancestor protein of the plant CHI (Ngaki *et al.*, 2012), the fatty-acid-binding protein (FAP) from *Arabidopsis thaliana* (PDB entry 4doi), cannot be superimposed either.

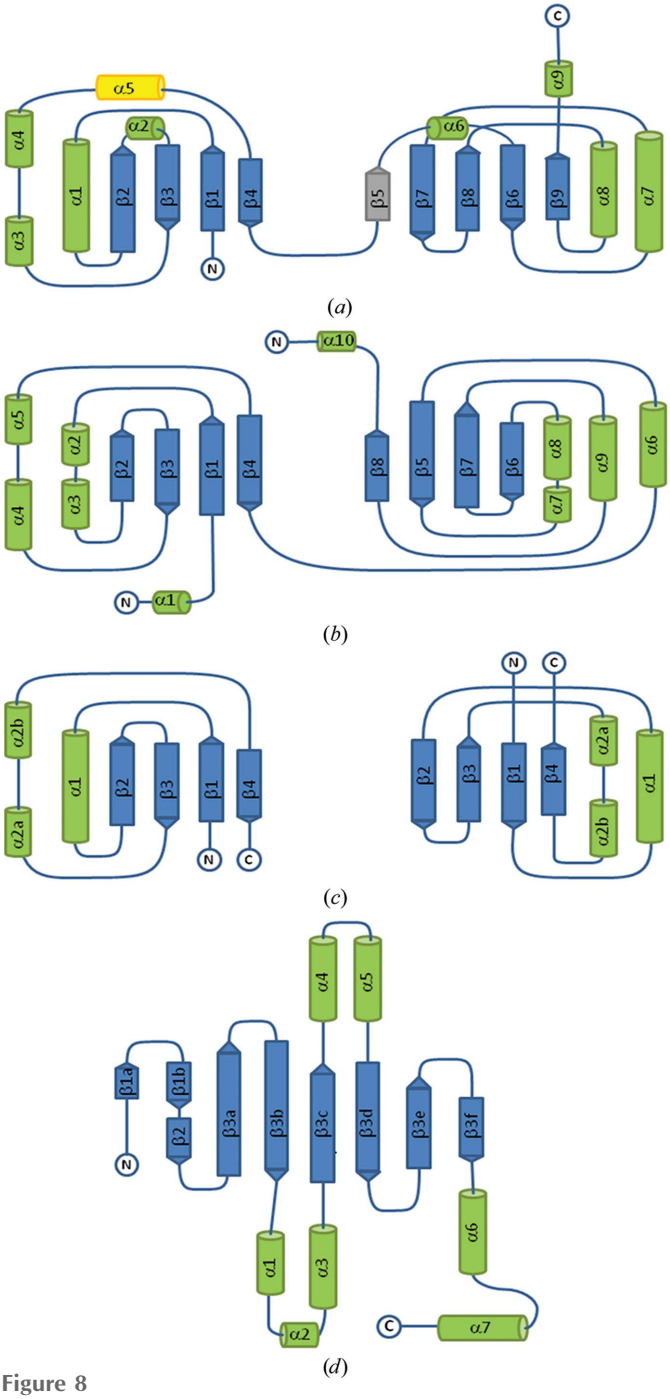


Figure 8
Topology diagrams with α-helices shown as green cylinders and β-strands as blue arrows (right). The N- and C-termini are labelled. (*a*) Bacterial CHI, (*b*) chlorite dismutase from *D. aromatica* (PDB entry 3q08) and (*c*) SP1 from *P. tremula* (PDB entry 1tr0) in a dimer orientation for comparison with CHI. (*d*) In contrast to the ferredoxin-like domains (ordered β2, β3, β1 and β4), all β-strands of plant CHI from *M. sativa* (PDB entry 1eyp) are consecutively arranged from β1a to β3a–f.

3.4.2. *DALI* database search. Structural homologues were identified through a database search with the *DALI* server (Dietmann *et al.*, 2001) based on secondary-structure elements. The best hit turned out to be a chlorite dismutase from *Dechloromonas aromatica* (PDB entry 3q08), which is a haem-dependent enzyme that is responsible for the detoxification of ClO_2^- (de Geus *et al.*, 2009). These obviously functionally unrelated enzymes exhibit very similar folds, with two ferredoxin-like domains (Fig. 8*b*). The catalytic domains superimpose with an overall root-mean-square deviation (r.m.s.d.) of 3.2 Å on C^α atoms. However, the other domains are contorted with respect to each other, although the topology diagram shows identical β-strand connections. As mentioned above, the C-terminus of a twofold-related CHI monomer penetrates into the interface of the β-sheets and causes this contortion. Interestingly, homology can only be detected for the tertiary structure. The quaternary structures of both enzymes show very different oligomerization patterns. The chlorite dismutase forms physiologically relevant pentamers and both domains are involved in oligomerization (de Geus *et al.*, 2009). In contrast, oligomerization of CHI occurs mainly *via* the catalytic domain. The N-terminal domain of the CHI harbours the active site, whereas in the chlorite dismutase the C-terminal domain hosts the active site. It is tempting to conclude that two single ferredoxin-like domains have linked up in different ways during evolution. Further database searches with only one domain identified such a possible ancestor protein. The stress-related thermostable protein SP1 from *Populus tremula* (PDB entry 1tr0) consists of 108 amino-acid residues and has a ferredoxin-like fold (Dgany *et al.*, 2004) (Fig. 8*c*). Two SP1 molecules generate a strong interaction and create a very stable dimer, with the β-sheets as the dimerization area (Dgany *et al.*, 2004). Superposition of the secondary-structure elements of SP1 with the catalytic and the solvent-exposed domains of the bacterial CHI reveals a fit on C^α atoms with overall r.m.s.d.s of 2.9 and 2.4 Å, respectively (Supplementary Table 3).

4. Conclusion

The two-domain structure of the bacterial CHI is closely related to the ferredoxin-like fold of a chlorite dismutase and the stress-related protein SP1, despite the lack of any functional relationship. The tertiary structure of bacterial CHI, with ferredoxin-like folds, is completely different to that of the

plant CHI, suggesting that the enzymes evolved convergently from different ancestor proteins. The bacterial CHI is only related to the plant CHI with respect to the products of the catalysed oxa-Michael addition. The general acid–base mechanism relies on the His33 side chain in bacterial CHI in contrast to the catalytic water molecule in plant CHI. The substrate-binding site of plant CHI is solvent-exposed during the catalytic process, but the bacterial CHI has a lid structure closing the active site upon substrate binding. The stereochemistry of the proposed isomerase reaction mechanism is supported by mutagenesis studies and specific $^{1}H/^{2}H$ exchange observed by NMR experiments. The shielding lid is unambiguously the main reason for the 19-fold higher efficiency of the bacterial CHI. SAXS elucidated the mobility of the lid and verified the quaternary structure in solution.

5. Related literature

The following references are cited in the Supporting Information for this article: Fox *et al.* (2014), Kozin & Svergun (2001) and Volkov & Svergun (2003).

Acknowledgements

Diffraction data for this study were collected on beamline 14.1 operated by the Helmholtz-Zentrum Berlin (HZB) at the BESSY II electron-storage ring, Berlin-Adlershof, Germany. The SAXS data were collected on beamline P12 operated by EMBL Hamburg at the PETRA-III ring, DESY, Hamburg, Germany. M. Backes (Symrise AG, Holzminden, Germany) synthesized the naringenin chalcone. This work was supported by the Bundesministerium für Bildung und Forschung within the Biokatalyse 2021 cluster (FKZ: 0315365 and 031A109). MT thanks the Landesgraduierten-Kolleg of Mecklenburg-Vorpommern for a PhD grant. AT was supported by the EMBL Interdisciplinary Postdoc Programme (EIPOD) under Marie Curie COFUND actions and by the BMBF research grant BioSCAT, contract No 05K12YE1.

References

Abendroth, J., Rice, A. E., McLuskey, K., Bagdasarian, M. & Hol, W. G. J. (2004). *J. Mol. Biol.* **338**, 585–596.
Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Bergfors, T. (2003). *J. Struct. Biol.* **142**, 66–76.
Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). *J. Am. Chem. Soc.* **129**, 5656–5664.
Braune, A., Gütschow, M., Engst, W. & Blaut, M. (2001). *Appl. Environ. Microbiol.* **67**, 5558–5567.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Dgany, O., Gonzalez, A., Sofer, O., Wang, W., Zolotnitsky, G., Wolf, A., Shoham, Y., Altman, A., Wolf, S. G., Shoseyov, O. & Almog, O. (2004). *J. Biol. Chem.* **279**, 51516–51523.
Dietmann, S., Park, J., Notredame, C., Heger, A., Lappe, M. & Holm, L. (2001). *Nucleic Acids Res.* **29**, 55–57.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Firmin, J. L., Wilson, K. E., Rossen, L. & Johnston, A. W. B. (1986). *Nature (London)*, **324**, 90–92.
Fox, N. K., Brenner, S. E. & Chandonia, J.-M. (2014). *Nucleic Acids Res.* **42**, D304–D309.
French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.
Gall, M. *et al.* (2014). *Angew. Chem. Int. Ed.* **53**, 1439–1442.
Gensheimer, M. & Mushegian, A. (2004). *Protein Sci.* **13**, 540–544.
Geus, D. C. de, Thomassen, E. A. J., Hagedoorn, P.-L., Pannu, N. S., van Duijn, E. & Abrahams, J. P. (2009). *J. Mol. Biol.* **387**, 192–206.
Guinier, A. (1939). *Ann. Phys. (Paris)*, **12**, 161–237.
Herles, C., Braune, A. & Blaut, M. (2004). *Arch. Microbiol.* **181**, 428–434.
Jez, J. M., Bowman, M. E., Dixon, R. A. & Noel, J. P. (2000). *Nature Struct. Mol. Biol.* **7**, 786–791.
Jez, J. M., Bowman, M. E. & Noel, J. P. (2002). *Biochemistry*, **41**, 5168–5176.
Jez, J. M. & Noel, J. P. (2002). *J. Biol. Chem.* **277**, 1361–1369.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
Kozin, M. B. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 33–41.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Krug, M., Weiss, M. S., Heinemann, U. & Mueller, U. (2012). *J. Appl. Cryst.* **45**, 568–572.
Leslie, A. G. W. & Powell, H. R. (2007). *Evolving Methods for Macromolecular Crystallography*, edited by R. J. Read & J. L. Sussman, pp. 41–51. Dordrecht: Springer.
Li, C., Wen, A., Shen, B., Lu, J., Huang, Y. & Chang, Y. (2011). *BMC Biotechnol.* **11**, 92.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Miyazaki, K. (2003). *Methods Mol. Biol.* **231**, 23–28s.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Ngaki, M. N., Louie, G. V., Philippe, R. N., Manning, G., Pojer, F., Bowman, M. E., Li, L., Larsen, E., Wurtele, E. S. & Noel, J. P. (2012). *Nature (London)*, **485**, 530–533.
Nising, C. F. & Bräse, S. (2012). *Chem. Soc. Rev.* **41**, 988–999.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *J. Appl. Cryst.* **45**, 342–350.
Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). *J. Appl. Cryst.* **40**, s223–s228.
Schneider, H. & Blaut, M. (2000). *Arch. Microbiol.* **173**, 71–75.
Schneider, H., Schwiertz, A., Collins, M. D. & Blaut, M. (1999). *Arch. Microbiol.* **171**, 81–91.
Schoefer, L., Braune, A. & Blaut, M. (2004). *Appl. Environ. Microbiol.* **70**, 6131–6137.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.
Wilson, D., Pethica, R., Zhou, Y. D., Talbot, C., Vogel, C., Madera, M., Chothia, C. & Gough, J. (2009). *Nucleic Acids Res.* **37**, D380–D386.
Winkel-Shirley, B. (2001). *Plant Physiol.* **126**, 485–493.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.