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

Structure determination and analysis of a bacterial chymotrypsin from *Cellulomonas bogoriensis*

A. Shaw,[a]* M. L. Saldajeno,[a] M. A. B. Kolkman,[b] B. E. Jones[b] and R. Bott[a]

[a]Genencor, 925 Page Mill Road, Palo Alto, CA 94304, USA, and [b]Genencor, Archimedesweg 30, 2333 CN Leiden, The Netherlands

Correspondence e-mail: andy.shaw@danisco.com

Received 30 January 2007
Accepted 22 February 2007

PDB Reference: chymotrypsin, 2ea3, r2ea3sf.

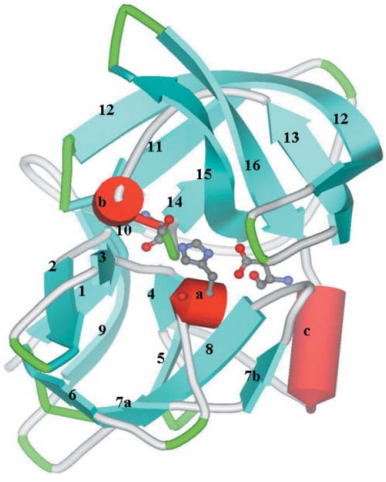
The crystal structure of a secreted chymotrypsin from the alkaliphile *Cellulomonas bogoriensis* has been determined using data to 1.78 Å resolution and refined to a crystallographic *R* factor of 0.167. The crystal structure reveals a large P1 substrate-specificity pocket, as expected for chymotrypsins. The structure is compared with close structural homologues. This comparison does not reveal clear reasons for the alkali tolerance of the enzyme, but the greater compactness of the structure and lowered hydrogen bonding may play a role.

1. Introduction

Proteases are found throughout the animal, plant and microbial kingdoms and are of immense importance in physiology, disease, medicine and biotechnology. A very large number of these are known as serine proteases. Serine proteases consist of three major families: chymotrypsin-like, subtilisins and α/β-hydrolases. Serine proteases contain an Asp-His-Ser catalytic triad (Blow *et al.*, 1969; Matthews *et al.*, 1967). This classic triad has been observed in a number of enzymes with otherwise very different structures and is the first well characterized example of convergent evolution at the molecular level; numerous variations of the classic catalytic triad have been discovered that catalyse the hydrolysis of many classes of substrate (Dodson & Wlodawer, 1998). In the serine proteases, the catalytic triad acts as a 'charge-relay' system (Blow *et al.*, 1969). Histidine removes a proton from the serine to make it a potent nucleophile. Crucial to this catalytic scheme is the formation of an unusually short 'catalytic' hydrogen bond between the histidine and aspartate. This makes the histidine a more potent base, facilitating its deprotonation of serine. This hydrogen has recently been visualized in an ultrahigh-resolution X-ray study of subtilisin (Kuhn *et al.*, 1998). The enzymes hydrolyse protein peptide bonds by attack of the nucleophilic serine of the catalytic triad, which forms an enzyme-acyl intermediate that is subsequently hydolysed by water.

A protease-secreting alkaliphile cellulomonad isolated from Lake Bogoria, Kenya has recently been characterized and typed and given the name *Cellulomonas bogoriensis* sp. nov. (Jones *et al.*, 2005). The gene and mature polypeptide of the secreted protease have been characterized (M. Kolkman, personal communication) and the polypeptide has been named cellulomonadin. The mature polypeptide is a bacterial chymotrypsin-family protease. It displays dimethylcasein activity at pH values above 5, is inhibited by PMSF, is stable at 318 K and is insensitive to chelants and oxidants and is thus of interest for biotechnological applications (Kolkman *et al.*, manuscript in preparation).

Analysis of the cellulomonadin gene reveals a precursor protein consisting of three polypeptide domains with a total of 495 amino acids. The mature catalytic domain consists of 189 amino acids (unpublished data). It has 45–48% sequence identity to the streptogrisins and 44% identity to α-lytic protease. Interestingly, the mature catalytic domain contains ten arginine residues (and no lysine residues), two glutamic acid and four aspartic acid residues, suggesting a high pI for the enzyme, which correlates with the alkaline environment in which it is found.



Here, we describe the structure of cellulomonadin. The protein has the expected fold for a bacterial chymotrypsin-like enzyme. The structure is compared with those of five close structural homologues in order to investigate potential structural factors for alkali tolerance.

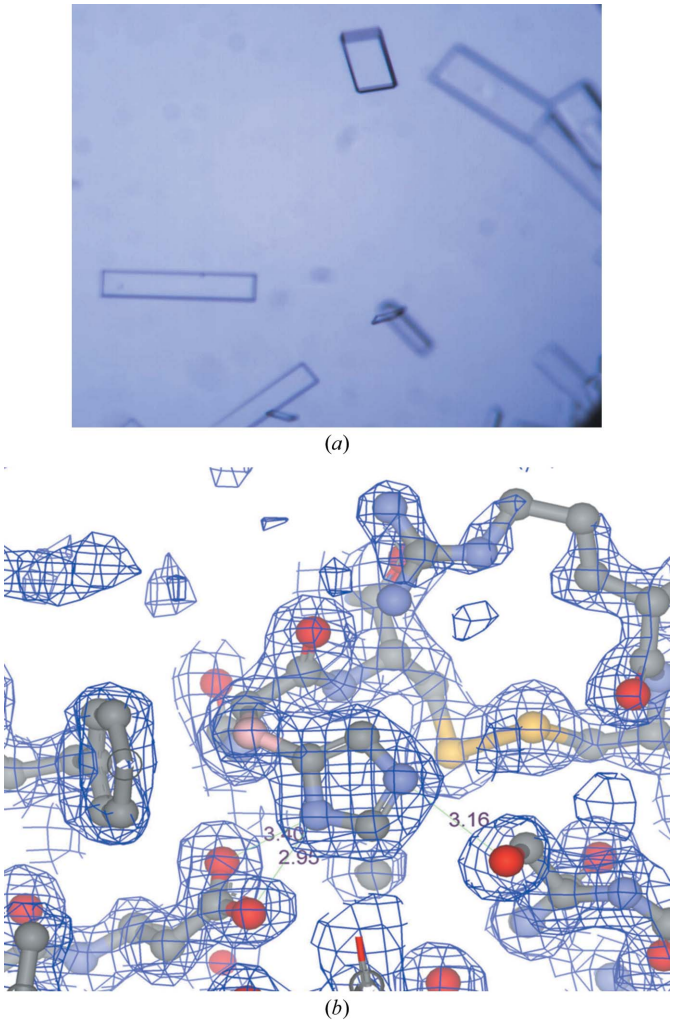
2. Methods

2.1. Crystallization of cellulomonadin

High-quality single crystals were obtained using hanging-drop vapor diffusion from 25% PEG 8000, 0.2 *M* ammonium sulfate, 15% glycerol, conditions that are cryoprotective (see Fig. 1*a*). The crystals were frozen in liquid nitrogen and kept frozen during data collection using an X-stream. Data were collected using an R-AXIS IV image plate equipped with focusing mirrors. X-ray reflection data were obtained to 1.78 Å resolution. The space group is $P2_12_12_1$, with unit-cell parameters a = 36.19, b = 52.49, c = 76.63 Å. There is one molecule per asymmetric unit. Data-collection statistics are shown in Table 1.

2.2. Molecular replacement, model building and refinement

Molecular replacement was performed with the *HT-XMR* program suite from Accelrys Inc (San Diego, CA, USA) within the DS modeling environment. Streptogrisin A (coordinates unmodified) from *Streptomcyes griseus* (PDB code 2sga) was used as the search model; *AMoRe* was used to perform the molecular-replacement calculation and rigid-body refinement was used to improve the positioning of the molecular-replacement solution. Problem areas were improved by manual rebuilding, followed by refinement with *CNX*, water placement and final *REFMAC* refinement. Fig. 1(*b*) illustrates the final $2F_o - F_c$ map.



(*a*)

(*b*)

Figure 1
(*a*) Crystals of cellulomonadin. (*b*) $2F_o - F_c$ electron density of cellulomonadin around the catalytic triad, contoured at 1σ.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outermost resolution shell.

Data-collection statistics	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 36.19, b = 52.49, c = 76.63
Resolution (Å)	19.16–1.78 (1.86–1.78)
Observed reflections	43192
Unique reflections	13319
Completeness (%)	93.1 (87.1)
Average $I/\sigma(I)$	34.0 (11.7)
R_{sym}	0.021 (0.082)
Refinement statistics	
R factor	0.167
R_{free}	0.196
Average B factor (Å^2)	16.01
R.m.s.d. bonds (Å)	0.011
R.m.s.d. angles (°)	1.382
No. of non-H atoms	1361

2.3. Structural analysis

Structural alignment and analysis were performed using *MOE* (Chemical Computing Corpotarion, Montreal, Canada; http://www.chemcomp.com).

3. Results and discussion

The final model was refined using data to 1.78 Å resolution and contains 1361 non-H atoms, including 58 water molecules and three sulfates. The model has a crystallographic R factor of 0.167 and an R_{free} of 0.196 in the resolution range 19.16–1.78 Å. Refinement and final model statistics are shown in Table 1. The Ramachandran plot shows no significant outliers.

3.1. Protein fold

Overall, the enzyme is roughly spherical with a diameter of approximately 35 Å (Fig. 2). As expected for this family of enzymes, the protein fold consists of two mostly antiparallel β-sheet subdomains, with an extensive interface between the two. The N-terminal subdomain contains 11 β-strands and one short helix between β-strands 5 and 6. The first eight β-strands form an antiparallel barrel structure packed with hydrophobic residues. β-Strands 1, 8 and 9 form an additional mixed parallel/antiparallel structure and a long loop connects this to the second subdomain. The C-terminal subdomain consists of six antiparallel β-strands, with a short helix between strands 13 and 14 and a longer helix at the C-terminus, and folds into a very compact barrel. There are three disulfide bridges: Cys17–Cys33, Cys96–Cys105 and Cys137–Cys158.

3.2. Active site

The active site is found at the extensive interface between the two β-barrel subdomains. The catalytic triad is formed by His32, Asp56 and Ser137. The hydrogen-bond distance between Asp56 OD1 and His32 ND1 is 2.9 Å and that between of His32 ND2 and Ser137 OG1

Table 2
R.m.s.d. (Å) of aligned C^{α} atoms for *C. bogoriensis* chymotrypsin structurally aligned with five closely homologous structures.

1, *C. bogoriensis* chymotrypsin (PDB code 2ea3; present work); 2, *S. griseus* glutamic acid-specific protease (1hpg); 3, *Lysobacter enzymogenes* α-lytic protease (1tal); 4, *S. fradiae* serine protease (2sfa); 5, *S. griseus* protease A (2sga); 6, *S. griseus* protease B (3sgb).

	2	3	4	5	6
1	1.42	0.98	1.32	1.27	1.25
2		1.46	1.28	1.24	1.21
3			1.37	1.37	1.35
4				0.58	0.84
5					0.89

Table 3
Summary of factors that may affect the stability of proteins in extreme environments for cellulomonadin and five close homologues.

PDB code	Hydrogen bonds	Hydrophobic interactions	Ion pairs	Disulfide bridges	Surface exposed		
					Asn	Asp/Glu	Arg/Lys
2ea3	31	52	3	3	7	2	9
1hpg	48	39	2	2	6	5	7
1tal	51	53	5	3	9	3	9
2sfa	38	50	2	2	6	5	7
2sga	41	49	1	2	10	3	6
3sgb	43	53	2	2	7	7	6

is 3.1 Å. The histidine and serine side chains lie between two short helices, one from each subdomain. Adjacent to these two residues is a large P1 substrate side-chain binding pocket formed by the second subdomain, which is characteristic of chymotrypsin-like enzymes

3.3. Substrate-binding cleft

The structure of the homologous streptogrisin B with a protein inhibitor bound is known and when structurally aligned with cellulomonadin reveals that in the latter six amino-acid residues are likely to bind across the substrate-binding cleft, with four S sites and two S′ sites. The protein self-cleaves to produce the mature protease domain. The sequence at the N-terminal pro-domain–catalytic domain junction is PRTMFD, where F is the first residue of the catalytic domain. This indicates that a residue at least as large as methionine can fit into the P1 specificity pocket.

3.4. Oxyanion hole

In addition to the conserved catalytic triad, serine proteases usually contain an oxyanion hole. These are basic groups that stabilize the developing negative charge of the tetrahedral intermediate. Comparison with the structure of streptogrisin B (Fujinaga *et al.*, 1982) leads to the prediction that residues Gly135 and Asp136 should form the oxyanion hole. The difference maps are ambiguous in this region. The $F_o - F_c$ map shows negative features at the carbonyl O atom of Gly135 and a positive peak which suggests that the peptide should be flipped. However, in the alternative conformation, the carbonyl O atom once again has negative difference density and a positive peak now appears where the O atom was previously. This suggests that the structure is flexible in this region. The model has been built with one conformation, with Gly135 forming the oxyanion hole as expected.

3.5. Comparison with close structural homologues

C. bogoriensis was discovered in Lake Bogoria, Kenya. This is a highly alkaline environment, having a pH of ∼10. Cellulomonadin was compared with five close structural homologues in order to investigate any differences between them that may explain its alkali tolerance. The r.m.s.d. for superposition of aligned C^{α} atoms is shown in Table 2 and the sequence alignment based on this structural

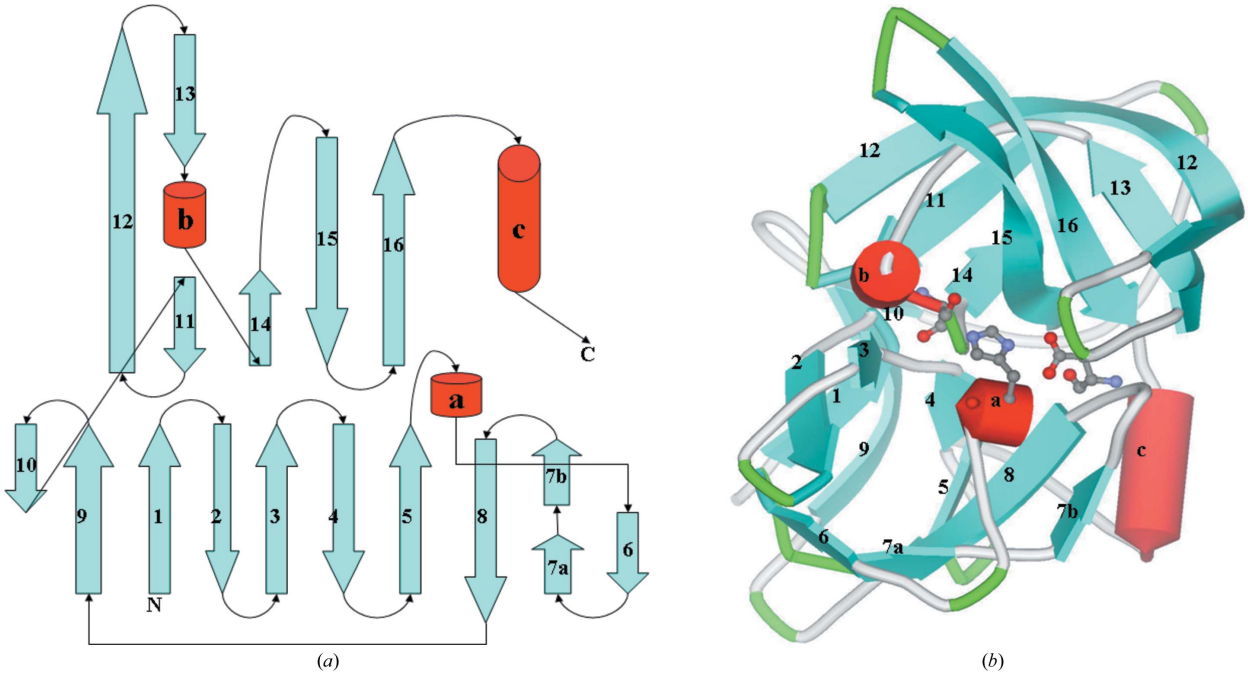


Figure 2
Structure of cellulomonadin. (*a*) Topology diagram of cellulomonadin. β-Sheets are shown as blue arrows pointing in the direction from the N-terminal to the C-terminal ends. α-Helices are shown as red cylinders. Connections between secondary-structure elements are shown as black line arrows. (*b*) Tertiary structure of cellulomonadin. The catalytic triad residues His32, Asp56 and Ser137 are shown as ball-and-stick representations.

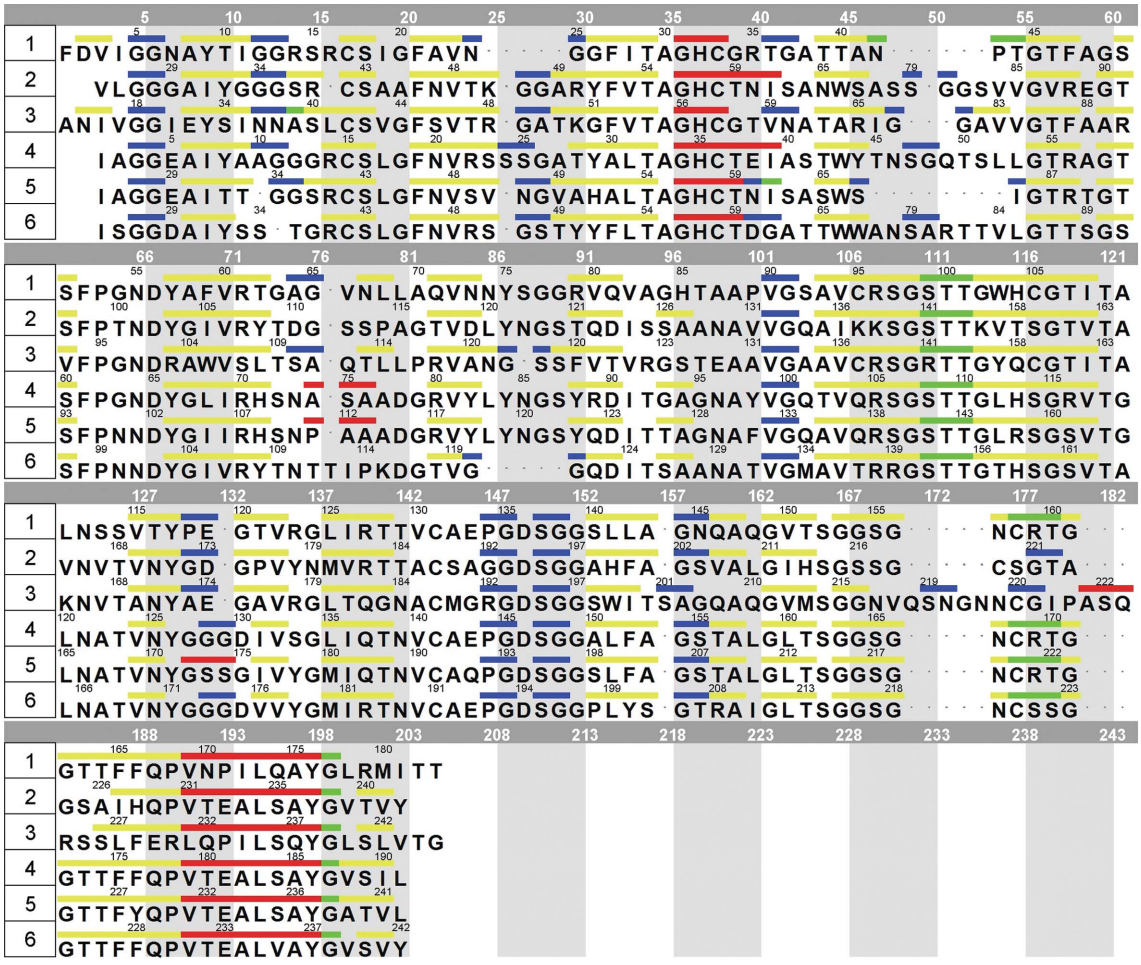


Figure 3
Structural alignment of *C. bogoriensis* chymotrypsin with closely homologous enzymes. Yellow bars indicate β-strands. Red bars indicate α-helices. Green and blue bars indicates polypeptide turns. The sequences are labeled as in Table 2.

alignment is shown in Fig. 3. The structural alignment reveals one loop (region 23–26) of cellulomonadin that is smaller than those of the reference set. There are no regions of the polypeptide where cellulomonadin has unique insertions compared with the reference set. It is possible that the greater degree of compactness of cellulomonadin assists in alkali tolerance. Compactness may be of particular importance to proteases, as they are prone to self-digestion (autolysis) and surface-exposed loops may be the regions most prone to autolysis, especially in extreme environments.

Analysis of other factors that may modulate stability are summarized in Table 3. We can see that there is no difference in the number of hydrophobic interactions, ion pairs and disulfide bridges and of surface-exposed Asn, negatively charged and positively charged residues compared with the reference set. Only the number of hydrogen bonds appears to be different, with cellulomonadin having 31 and the reference set having 38–51. It is possible that the preponderance of hydroxyl ions in the pH 10 environment may disrupt hydrogen bonds within the protein and thus the protein has a reduced number. However, there are no obvious increases in other interaction types to compensate for this.

4. Conclusions

We have reported the crystal structure of a bacterial chymotrypsin from *C. bogoriensis*, an extremophile from the highly alkaline Lake Bogoria, Kenya, at 1.78 Å. Comparison with close structural homologues does not provide any clear structural rationale for alkali tolerance. Potential factors are greater compactness, providing greater structural stability and/or lowered susceptibility to autolysis, and a reduction in the total number of hydrogen bonds.

References

Blow, D. M., Birktoft, J. J. & Hartley, B. S. (1969). *Nature (London)*, **221**, 337–340.

Dodson, G. & Wlodawer, A. (1998). *Trends Biochem. Sci.* **23**, 347–352.

Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M. Jr & James, M. N. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4868–4872.

Jones, B. E., Grant, W. D., Duckworth, A. W., Schumann, P., Weiss, N. & Stackebrandt, E. (2005). *Int. J. Syst. Evol. Microbiol.* **55**, 1711–1714.

Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.

Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. (1967). *Nature (London)*, **214**, 652–656.