Acta Crystallographica Section F Structural Biology and Crystallization Communications

ISSN 1744-3091

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

^aGenencor, 925 Page Mill Road, Palo Alto, CA 94304, USA, and ^bGenencor, 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.



© 2007 International Union of Crystallography All rights reserved

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

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 ultrahighresolution 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 polypetide 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





Figure 1

(a) Crystals of cellulomonadin. (b) $2F_{\rm o} - F_{\rm c}$ 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.

$P2_{1}2_{1}2_{1}$
a = 36.19, b = 52.49, c = 76.63
19.16-1.78 (1.86-1.78)
43192
13319
93.1 (87.1)
34.0 (11.7)
0.021 (0.082)
0.167
0.196
16.01
0.011
1.382
1361

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.

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 $R_{\rm free}$ 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

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 stabi-

Table 3

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

					Surface exposed			
PDB code	Hydrogen bonds	Hydrophobic interactions	Ion pairs	Disulfide bridges	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	

lize 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 $F_{\rm o} - F_{\rm c}$ 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.

	5	10 1	5 2	.0 2	5 30	35 40	0 45	50	55 60
1	FDVIGGNAY	TIGGRS	RCSIG	FAVN	GGFIT	AGHCGR	TGATTAN	• • • • • • P	TGTFAGS
2	VLGGGAI	YGGGSR	CSAA	FNVTK	GGARYFVT	AGHCTN	ISANWSA	ss Ggsv	<i>W</i>GVREGT
3	ANIVGGIEY	SINNAS		FSVTR	GATKGFVT	AGHCGT		G · · · GAV	VGTFAAR
4					25 SSGATYALT	AGHCTE		NSGQTSL	
5				FNVSV		AGHCTN			
6		, 34 У S S · Т G				AGHCTD			
	66	71 7	6 8	1 8	6 91	<u>9</u> 6 10	1 106	111	116 121
1	SFPGNDYAF		VNLL			GHTAAP		SGSTTGW	
2			SSPA		YNGS TQD I S	126 131 SAANAV		SGSTTKV	TSGTVTA
3						123 131 GSTEAA		SGRTTGY	
4	SFPGNDYGL		SAAD	GRVYL		GAGNAY		SGSTTGL	
5									RSGSVTG
6									
	197	132 1	11FKD	42 1/	17 152	157 16	2 167	172	177 182
1									
1		GTVRG			135 GDSGGSLLA		GVTSGGS	G N	
1 2 3	LNSSVTYPE 173 VNVTVNYGD 168 VNVTVNYGD		LIRTT MVRTT	ACSAG	GDSGGSLLA GDSGGAHFA		GVTSGGS 211 GIHSGSS GIHSGSS	G N G . <u></u>	CRTG
1 2 3	LNSSVTYPE VNVTVYGD KNVTVNYGD KNVTANYAE	GTVRG GPVYN GAVRG	LIRT MVRT LTQGN	130 VCAEP 192 ACSAG 192 ACMGR	GDSGGSLLA GDSGGSLLA GDSGGAHFA GDSGGSWI	201 GSVAL 210 SAGQAQ	G V T S G G S G V T S G G S G I H S G S S G V M S G G N 160	G	CRTG SGTA CGIPASQ
1 2 3 4	LNSSVTYPE VNV168VTYPE VNV178VNYGD KNVTANYAE 120NATVNYGG	GTVRG GTVRG GPVYN GAVRG GDIVSG	125 LIRTT MVRTT 184 LTQGN 135 LQTN	ACSAG ACSAG ACMGR	GDSGGSLLA GDSGGAHFA GDSGGSWIT GDSGGALFA	GSVAL 300 300 300 300 300 300 300 300 300 30	G V T S G G S G I H S G S G V M S G G N G L T S G G S G L T S G G S 211 G L T S G G S 211 G L T S G G S	G N G 219 VQSNGNN G N	CRTG 221 CG I PASQ CRTG CRTG
1 2 3 4 5	LNSSVTYPE 175 VNVTVNYGD KNVTANYGD 174 KNVTANYAE 120 LNATVNYGG LNATVNYGS	G T V R G G P V Y N G A V R G G D I V S G S G I V Y G	LIRTT MVRTT LTQGN LIQTN MILQTN	130 VCAEP ACSAG ACSAG ACMGR MGR VCAEP VCAEP	GDSGGSLLA GDSGGAHFA GDSGGSWIT GDSGGALFA GDSGGALFA	50 145 <u>G</u> NQAQ <u>G</u> SVAL <u>G</u> SVAL <u>G</u> SVAL <u>G</u> SVAL <u>G</u> SVAL <u>G</u> SVAL	150 150 155 GVTSGCS G1HSCSS 216 GVMSGGN 60 165 GUTSCS 165 165 GUTSCS 165 165 GUTSCS 165 165 GUTSCSS 165 165 GUTSCSS 165 165 GUTSCSSS 165 165	G N Q N V Q S N G N N G N N N	CRTG CRTG 221 CGIPASQ CGIPASQ CRTG CRTG CRTG
1 2 3 4 5 6	LNSSVTYPE VNVTVNYGD KNVTANYGD KNVTANYAE LNATVNYGG LNATVNYGS LNATVNYGG	GPVYRG GPVYN GAVRG GD IVSG SG IVYG GDVVYG	LIRT MVRTT LTQGN LIQTN MINQTN MINT	130 VCAEP ACSAG ACMGR ↓0 VCAEP VCAEP VCAEP	GDSGGSLLA GDSGGAHFA GDSGGSWIT GDSGGALFA GDSGGSLFA GDSGGPLYS	445 GSVA2 GSVA2 GSVA2 GSTA2 GSTA2 GSTA2 GTRA1	150 GVTSGC5 GIHSC5S GIHSC5S GVMSGCN GLTSGC5 GLTSGC5 GLTSGC5 GLTSGC5	G N G 210 VQSNGNN G N Q 218 N Q N N	CRTG SGTA CGIPASQ CRTG CRTG CRTG CSSG
1 2 3 4 5 6	115 LNSSVTYPE VNVTVNYGD KNVTANYGD 120 LNATVNYGG LNATVNYGG LNATVNYGG LNATVNYGG 128 128 128 129 LNATVNYGG 128 129 120 120 120 120 120 120 120 120 120 120	GPVYN GPVYN GAVRG GD I VSG SG I VYG GDVYG GDVVYG 193 12	125 LIRT MVRT LTQGN 135 LQTN MIRTN 98 21 140 140 140 140 140 140 140 14	130 VCAE P ACSAG ACMGR 40 VCAE P VCAE P VCAE P VCAE P VCAE P 20 20 20	GDSGGAHFA GDSGGSWIT GDSGGALFA GDSGGALFA GDSGGALFA GDSGGSLFA GDSGGPLYS	5 AGQAQ 5 AGQAQ 6 SVAL 6 SVAL 15 AGQAQ 6 STAL 6 STAL 12 C 13 22	150 100 155 GVTSGGS GIHSGGS 155 GIHSGGS 155 155 GUTSGGS 155 155 GUTSGS 155 155 GUTSSGS 155 155 GUTSSGSS 155	G N N G 219 V Q S N G N N G N N G N N C 233	CRTG CRTG CGIPASQ CGIPASQ CRTG CRTG CSSG 238 243
1 2 3 4 5 6	LNSSVTYPE LNSSVTYPE KNVTANYGD KNVTANYGD LNATVNYGG LNATVNYGG LNATVNYGG LNATVNYGG 188 195 195 195 195 195 195 195 195	G T V R G G T V R M G A V R M G A V R G G D V V S G G D V V Y G G D V V Y G H I L 236 A Y	L I R T T MVR T T L I Q R L I Q T N L I Q T N M I R T N 88 100 G L RM I	100 1 VCAE 102 ACSAG 102 ACMGR 102 MCAE 102 VCAE 103 TT 103	GDSGGSLLA GDSGGSWIT GDSGGALFA GDSGGALFA GDSGGSLFA GDSGGSLFA GDSGGPLYS 8 213	145 0 GNQAQ 0 GSVAL 210 SAGQAQ 0 GSTAL 0 GSTAL 0 GTRAI 218 218 22	150 100 155 GVTSGGS GIHSGGS GIHSGSS GISGS GUTSGGS 160 GLTSGGS 160 GLTSGGS 160 GLTSGGS 160 GLTSGSS 160 100 100	G N N G N G N G N G N G N G N C N C N C N N C N N N N N N N N N N N N N	CRTG 221 CGIPASQ CGIPASQ CCIPASQ CRTG CRTG CSSG 238 243
1 2 3 4 5 6 1 2 2	LNSSVTYPE KNVTANYGD KNVTANYGG LNATVNYGG LNATVNYGG LNATVNYGG GTTFFQPVN GSALHQPVT 2207	GTVRG GTVRG GPVYN GAVRG GD VVG GD VVG GD VVYG GD VVYG FILQAY EALSAY	125 LIRT MVRT LTQGN 135 135 135 135 135 135 135 135	100 VCAE ACSAG ACMGR ACMGR 100 </td <td>GDSGGSLLA GDSGGSAHFA GDSGGSWIT GDSGGALFA GDSGGSLFA GDSGGPLYS 213</td> <td>37 145 16 GNQAQ 15 12 GSVAL 210 210 SAGQAQ 15 14 GSTAL 15 15 GSTAL 15 14 GTRAI 218 22</td> <td>150 100 155 G V T S G S S G I H S G S S G I H S G S S G U M S G G N 100 100 G L T S G G S 100 100 G L T S G G S 100 100 G L T S G G S 100 100 G L T S G S 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 <</td> <td>G N N G 218 C V Q S N G N N G N N G N N G N N C 233</td> <td>CRTG 221 CGIPASQ CGIPASQ CCIFG 222 CGIPASQ CRTG 222 CSSG 238 243</td>	GDSGGSLLA GDSGGSAHFA GDSGGSWIT GDSGGALFA GDSGGSLFA GDSGGPLYS 213	37 145 16 GNQAQ 15 12 GSVAL 210 210 SAGQAQ 15 14 GSTAL 15 15 GSTAL 15 14 GTRAI 218 22	150 100 155 G V T S G S S G I H S G S S G I H S G S S G U M S G G N 100 100 G L T S G G S 100 100 G L T S G G S 100 100 G L T S G G S 100 100 G L T S G S 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 <	G N N G 218 C V Q S N G N N G N N G N N G N N C 233	CRTG 221 CGIPASQ CGIPASQ CCIFG 222 CGIPASQ CRTG 222 CSSG 238 243
1 2 3 4 5 6 1 2 3	LNSSVTYPE LNSSVTYPE KNVTANYGD KNVTANYGD LNATVNYGG LNATVNYGG LNATVNYGG LNATVNYGG GT276FQPYN GSALHQPVT RSSL75FRL	GTVRG GTVRG GPVYN GAVRG GP,5 IVSG SGIVYG SGIVYG GDVVYG FILQAY EALSAY PILSQY	L I R T T L I R T T L T QGN L T QGN L T QGN L T QTN M I R T N M I R T N B 20 G L 240 G V T V Y G L S L	VCAE VCAE ACSAG ACMGR WCAE VCAE VCAE VCAE VCAE VCAE VCAE VCAE TT TG	GDSGGSLLA GDSGGSLLA GDSGGAHFA GDSGGSWIT GDSGGALFA GDSGGSLFA GDSGGPLYS 213	201 202 202 202 203 204 204 204 205 204 205 204 205 204 205 204 205 204 205 205 205 205 205 205 205 205	2 150 107 155 GVTSG255 GIHS215 GUTSG6 GLTSG65 GLTSG65 GLTSG65 13 228	G N N G 210 V Q S NGNN G N N G N N C 218 C N N 218 N N 218 N N	CRTG CRTG CRTG CRTG CCSSG CSSG 238 243
1 2 3 4 5 6 1 2 3 4	115 LNSSVTYPE KNVT VNVT VNVT VNVT NYT LNATVNYGG LNATVNYGG LNATVNYGG LNATVNYGG GSAI HQPVT RSSLFERL 0 GTTFFQPVT 227 RSQU 128 129 120 120 120 120 120 120 120 120	GTVRG GTVRG GPVYN GAVRG GDVYSG SGIVYG GDVYG GDVYYG FIL23AY FIL23AY FIL23AY FIL23AY FIL23AY	L I R T T NV R T T L I Q R T L I I I I I I I I I I I I I I I I I I	VCAEP ACSAG ACMGR MCAEP VCAEP VCAEP VCAEP VCAEP 03 20 TT	GDSGGSLLA GDSGGSLLA GDSGGAHFA GDSGGALFA GDSGGALFA GDSGGSLFA GDSGGPLYS 8 213	201 201 201 201 201 201 201 201	2 150 107 155 GVTSGGS GIHSGSS GIHSGGN GLTSGGS GLTSGGS GLTSGGS 13 228	G N G C V Q S NGN N G N G N 218 G N 233	CRTG CRTG CRTG CCGIPASQ CCGIPASQ CRTG CRTG 238 243
1 2 3 4 5 6 1 2 3 4 5	115 LNSSVTYPE VNVT VNVT VNVT VNVT VNYG LNATVNYGG LNATVNYGG LNATVNYGG GT 25 F F QP VN GSA LHQP VT RSSLFERLQ GTTFFQPVT 232 GTTF27QPVT	GTVRG GTVRG GPVYN GAVRG GDIVSG SGLVYG GDVVYG GDVVYG FALSAY FILSQY EALSAY EALSAY	L I R T T NV R T T L T Q G N L I Q T N M I R T N G L RM I G V T V G L S L V G V S L G V S L G X S L V G X S L V G A T V 242	VCAEP ACSAG ACMGR VCAEP VCAEP VCAEP VCAEP VCAEP TT TG	GDSGGSLLA GDSGGAHFA GDSGGALFA GDSGGALFA GDSGGSLFA GDSGGPLYS 213	5 AGQAQ 5 AG 5 A	50 10 155 GVTSGGS GIHSC5S GVMSGGN 100 GLTSGGS GLTSGGS GLTSGGS 3 228	G N G C V Q S NGN N G N G N G N 218 G N 233	CRTG 221 CGIPASQ CGIPASQ CRTG CRTG 238 243

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.