wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 03:19 PM GMT

PDB ID	:	4C2D
Title	:	Crystal structure of the protease CtpB in an active state
Authors	:	Mastny, M.; Heuck, A.; Kurzbauer, R.; Clausen, T.
Deposited on	:	2013-08-17
Resolution	:	2.70 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
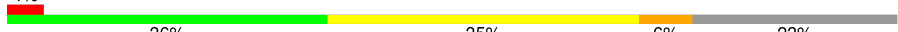*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.
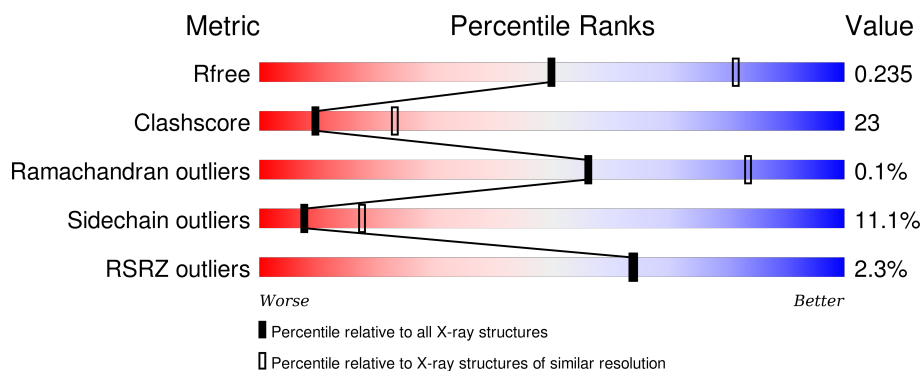
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	446	65% 29% · ·
1	B	446	% 63% 30% 5% ·
1	C	446	3% 48% 42% 7% ·
1	D	446	4% 36% 35% 6% 23%
2	E	6	83% 17%

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	6	50% 50%
2	G	6	83% 17%
2	H	6	33% 33% 33%
3	M	5	60% 40%
4	N	4	75% 25%
4	O	4	75% 25%
4	P	4	50% / 25% 50% 25%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 13353 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called CARBOXY-TERMINAL PROCESSING PROTEASE CTPB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	434	Total	C	N	O	S	0	0	1
			3382	2135	578	659	10			
1	B	434	Total	C	N	O	S	0	0	1
			3382	2135	578	659	10			
1	C	434	Total	C	N	O	S	0	0	1
			3382	2135	578	659	10			
1	D	344	Total	C	N	O	S	0	0	1
			2705	1709	456	533	7			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	43	MET	-	INITIATING METHIONINE	UNP O35002
A	481	LEU	-	EXPRESSION TAG	UNP O35002
A	482	GLU	-	EXPRESSION TAG	UNP O35002
A	483	HIS	-	EXPRESSION TAG	UNP O35002
A	484	HIS	-	EXPRESSION TAG	UNP O35002
A	485	HIS	-	EXPRESSION TAG	UNP O35002
A	486	HIS	-	EXPRESSION TAG	UNP O35002
A	487	HIS	-	EXPRESSION TAG	UNP O35002
A	488	HIS	-	EXPRESSION TAG	UNP O35002
B	43	MET	-	INITIATING METHIONINE	UNP O35002
B	481	LEU	-	EXPRESSION TAG	UNP O35002
B	482	GLU	-	EXPRESSION TAG	UNP O35002
B	483	HIS	-	EXPRESSION TAG	UNP O35002
B	484	HIS	-	EXPRESSION TAG	UNP O35002
B	485	HIS	-	EXPRESSION TAG	UNP O35002
B	486	HIS	-	EXPRESSION TAG	UNP O35002
B	487	HIS	-	EXPRESSION TAG	UNP O35002
B	488	HIS	-	EXPRESSION TAG	UNP O35002
C	43	MET	-	INITIATING METHIONINE	UNP O35002
C	481	LEU	-	EXPRESSION TAG	UNP O35002
C	482	GLU	-	EXPRESSION TAG	UNP O35002

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	483	HIS	-	EXPRESSION TAG	UNP O35002
C	484	HIS	-	EXPRESSION TAG	UNP O35002
C	485	HIS	-	EXPRESSION TAG	UNP O35002
C	486	HIS	-	EXPRESSION TAG	UNP O35002
C	487	HIS	-	EXPRESSION TAG	UNP O35002
C	488	HIS	-	EXPRESSION TAG	UNP O35002
D	43	MET	-	INITIATING METHIONINE	UNP O35002
D	481	LEU	-	EXPRESSION TAG	UNP O35002
D	482	GLU	-	EXPRESSION TAG	UNP O35002
D	483	HIS	-	EXPRESSION TAG	UNP O35002
D	484	HIS	-	EXPRESSION TAG	UNP O35002
D	485	HIS	-	EXPRESSION TAG	UNP O35002
D	486	HIS	-	EXPRESSION TAG	UNP O35002
D	487	HIS	-	EXPRESSION TAG	UNP O35002
D	488	HIS	-	EXPRESSION TAG	UNP O35002

- Molecule 2 is a protein called PEPTIDE1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	6	Total 35	C 21	N 6	O 8	0	0	0
2	F	6	Total 35	C 21	N 6	O 8	0	0	0
2	G	6	Total 35	C 21	N 6	O 8	0	0	0
2	H	4	Total 19	C 12	N 4	O 3	0	0	1

- Molecule 3 is a protein called PEPTIDE2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	M	5	Total 32	C 19	N 6	O 7	0	0	0

- Molecule 4 is a protein called PEPTIDE2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	N	4	Total 29	C 17	N 5	O 7	0	0	0
4	O	4	Total 27	C 16	N 5	O 6	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	P	3	Total	C	N	O	0	0	0
			16	9	3	4			

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	119	Total	O	0	0
			119	119		
5	B	89	Total	O	0	0
			89	89		
5	C	37	Total	O	0	0
			37	37		
5	D	23	Total	O	0	0
			23	23		
5	E	3	Total	O	0	0
			3	3		
5	M	3	Total	O	0	0
			3	3		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues 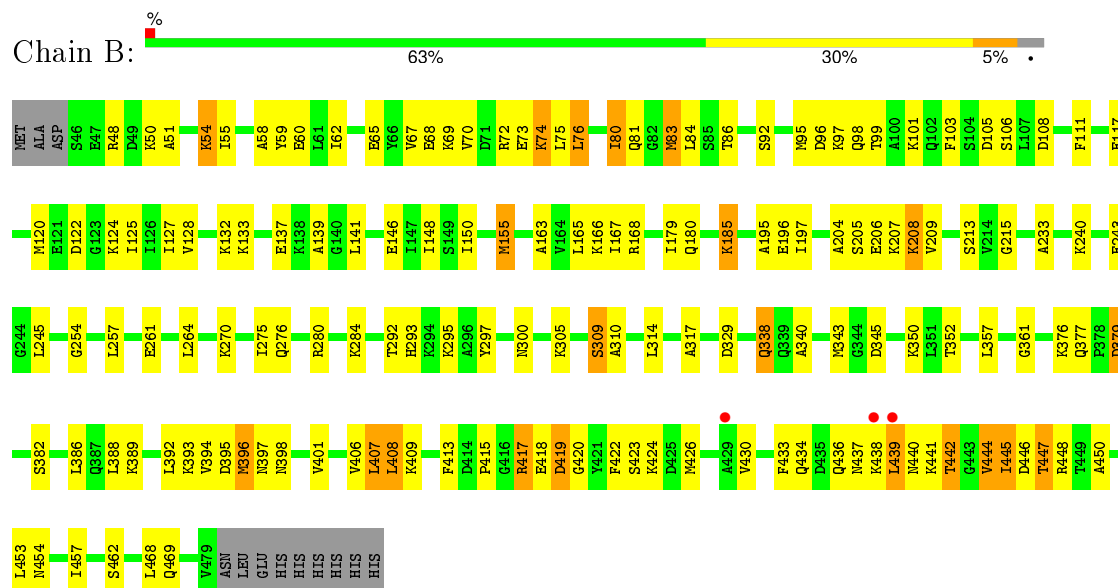are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: CARBOXY-TERMINAL PROCESSING PROTEASE CTPB
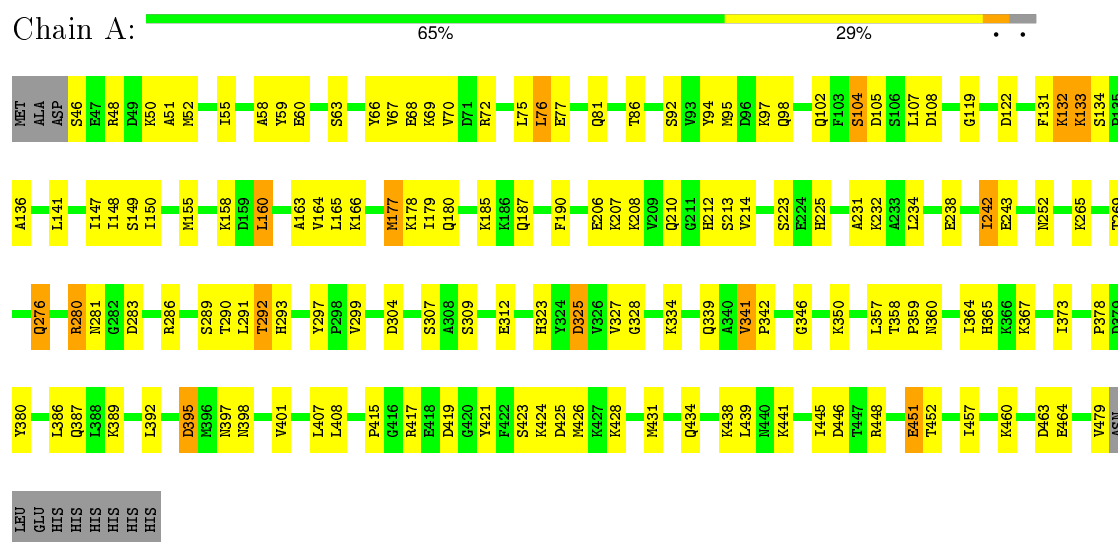
• Molecule 1: CARBOXY-TERMINAL PROCESSING PROTEASE CTPB

• Molecule 1: CARBOXY-TERMINAL PROCESSING PROTEASE CTPB

Chain C:

• Molecule 1: CARBOXY-TERMINAL PROCESSING PROTEASE CTPB

Chain D:

• Molecule 2: PEPTIDE1

Chain E:

A1 A2 S3 A6

- Molecule 2: PEPTIDE1

Chain F:

50%   50%

A1 A2 S3 A6

- Molecule 2: PEPTIDE1

Chain G:

83%   17%

A1 A2 A6

- Molecule 2: PEPTIDE1

Chain H:

33%   33%   33%

ALA ALA S3 L4 S5 A6

- Molecule 3: PEPTIDE2

Chain M:

60%   40%

A1 A2 P3 Q4 A5

- Molecule 4: PEPTIDE2

Chain N:

75%   25%

P2 Q3 T4 A5

- Molecule 4: PEPTIDE2

Chain O:

75%   25%

P2 A5

- Molecule 4: PEPTIDE2

Chain P:

50%

25%   50%   25%

PRO Q3 T4 A5
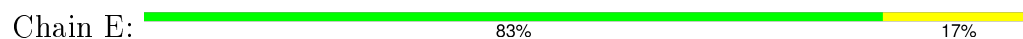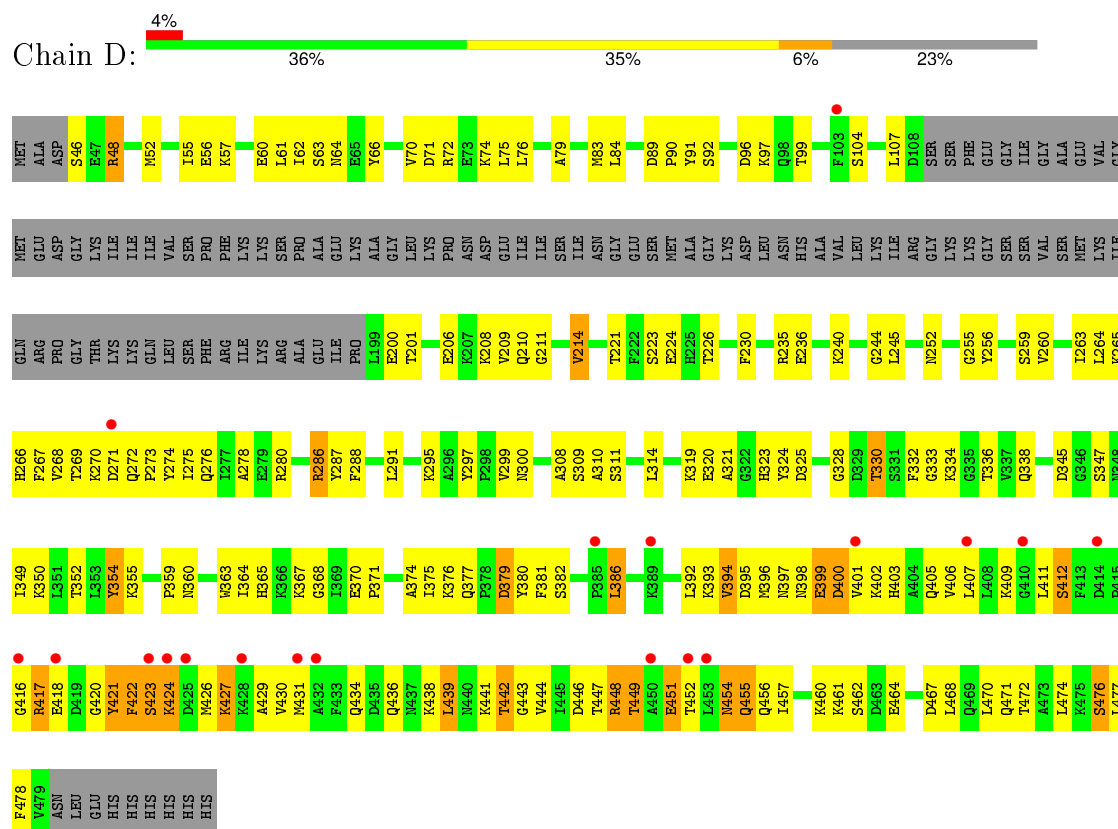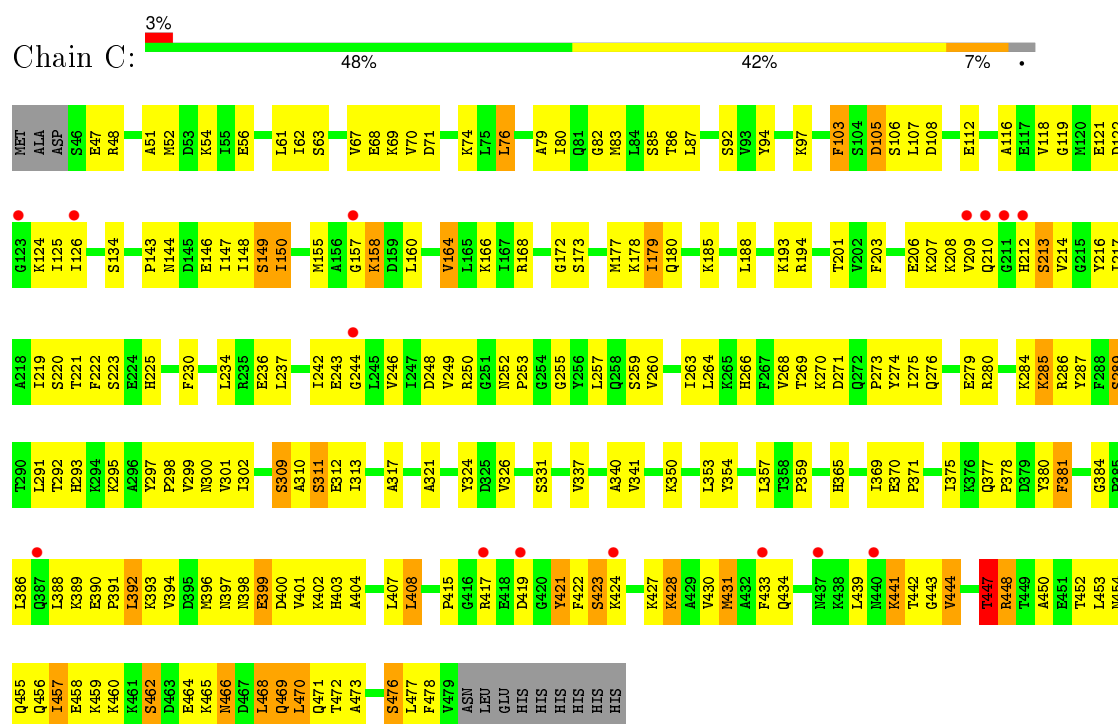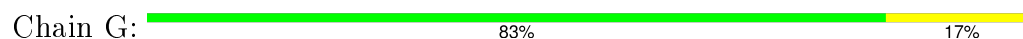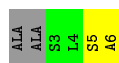
4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	117.05Å 65.38Å 169.06Å 90.00° 95.03° 90.00°	Depositor
Resolution (Å)	20.00 – 2.70 39.73 – 2.60	Depositor EDS
% Data completeness (in resolution range)	100.0 (20.00-2.70) 99.9 (39.73-2.60)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.23 (at 2.61Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.227 , 0.241 0.224 , 0.235	Depositor DCC
R_{free} test set	3536 reflections (5.28%)	DCC
Wilson B-factor (Å^2)	48.8	Xtriage
Anisotropy	0.603	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 48.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$<|L|> = 0.41, <L^2> = 0.24$	Xtriage
Outliers	1 of 79096 reflections (0.001%)	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	13353	wwPDB-VP
Average B, all atoms (Å^2)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.06% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $<|L|>$, $<L^2>$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$ >5	RMSZ	#$|Z|$ >5
1	A	0.51	0/3437	0.64	0/4615
1	B	0.50	0/3437	0.67	1/4615 (0.0%)
1	C	0.41	0/3437	0.61	3/4615 (0.1%)
1	D	0.36	0/2750	0.53	0/3699
2	E	0.58	0/34	0.68	0/45
2	F	0.74	0/34	0.68	0/45
2	G	0.39	0/34	0.71	0/45
2	H	1.14	0/18	0.83	0/24
3	M	0.49	0/32	0.48	0/42
4	N	0.58	0/29	0.50	0/37
4	O	0.45	0/27	0.45	0/34
4	P	0.26	0/15	0.52	0/18
All	All	0.45	0/13284	0.62	4/17834 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	2
1	C	0	1
1	D	0	1
All	All	0	4

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	441	LYS	CD-CE-NZ	-10.12	88.42	111.70
1	C	431	MET	CB-CG-SD	9.10	139.71	112.40
1	C	431	MET	CA-CB-CG	6.04	123.56	113.30
1	B	83	MET	CG-SD-CE	6.02	109.83	100.20

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	395	ASP	Peptide
1	B	417	ARG	Peptide
1	C	384	GLY	Peptide
1	D	436	GLN	Peptide

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3382	0	3433	117	1
1	B	3382	0	3433	133	1
1	C	3382	0	3433	217	2
1	D	2705	0	2709	168	2
2	E	35	0	38	1	0
2	F	35	0	38	6	0
2	G	35	0	38	2	0
2	H	19	0	14	2	0
3	M	32	0	32	3	0
4	N	29	0	27	1	0
4	O	27	0	22	1	0
4	P	16	0	8	9	0
5	A	119	0	0	23	0
5	B	89	0	0	14	0
5	C	37	0	0	3	0
5	D	23	0	0	2	0
5	E	3	0	0	1	0
5	M	3	0	0	0	0
All	All	13353	0	13225	593	4

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 23.

The worst 5 of 593 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:212:HIS:CE1	1:C:243:GLU:HB3	1.77	1.20
1:D:255:GLY:N	4:P:5:ALA:OXT	1.83	1.12
1:A:95:MET:SD	5:A:2017:HOH:O	2.07	1.10
1:B:393:LYS:H	1:B:396:MET:HE3	1.16	1.08
1:D:393:LYS:H	1:D:396:MET:HE3	1.19	1.07

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:462:SER:OG	1:D:210:GLN:OE1[2_556]	1.98	0.22
1:C:112:GLU:OE2	1:C:287:TYR:OH[2_646]	2.14	0.06
1:C:112:GLU:OE2	1:C:285:LYS:NZ[2_646]	2.15	0.05
1:A:438:LYS:NZ	1:D:280:ARG:O[2_656]	2.16	0.04

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	432/446 (97%)	418 (97%)	14 (3%)	0	100	100
1	B	432/446 (97%)	415 (96%)	16 (4%)	1 (0%)	52	80
1	C	432/446 (97%)	413 (96%)	18 (4%)	1 (0%)	52	80
1	D	340/446 (76%)	311 (92%)	29 (8%)	0	100	100
2	E	4/6 (67%)	4 (100%)	0	0	100	100
2	F	4/6 (67%)	4 (100%)	0	0	100	100
2	G	4/6 (67%)	4 (100%)	0	0	100	100
2	H	2/6 (33%)	2 (100%)	0	0	100	100
3	M	3/5 (60%)	3 (100%)	0	0	100	100
4	N	2/4 (50%)	2 (100%)	0	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	O	2/4 (50%)	1 (50%)	1 (50%)	0	100	100
4	P	1/4 (25%)	0	1 (100%)	0	100	100
All	All	1658/1825 (91%)	1577 (95%)	79 (5%)	2 (0%)	56	83

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	439	LEU
1	C	447	THR

5.3.2　Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	369/381 (97%)	342 (93%)	27 (7%)	17	39
1	B	369/381 (97%)	335 (91%)	34 (9%)	11	25
1	C	369/381 (97%)	319 (86%)	50 (14%)	5	11
1	D	295/381 (77%)	249 (84%)	46 (16%)	3	8
2	E	3/3 (100%)	3 (100%)	0	100	100
2	F	3/3 (100%)	3 (100%)	0	100	100
2	G	3/3 (100%)	3 (100%)	0	100	100
2	H	1/3 (33%)	1 (100%)	0	100	100
3	M	2/2 (100%)	2 (100%)	0	100	100
4	N	3/3 (100%)	3 (100%)	0	100	100
4	O	2/3 (67%)	2 (100%)	0	100	100
All	All	1419/1544 (92%)	1262 (89%)	157 (11%)	8	17

5 of 157 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	149	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	309	SER
1	D	438	LYS
1	C	158	LYS
1	C	236	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	293	HIS
1	D	397	ASN
1	C	454	ASN
1	C	212	HIS
1	C	397	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2		OWAB(Å^2)	Q<0.9
1	A	434/446 (97%)	-0.33	0 100	100	20, 45, 76, 95	0
1	B	434/446 (97%)	-0.24	3 (0%) 89	90	21, 41, 100, 132	0
1	C	434/446 (97%)	0.05	15 (3%) 48	48	40, 70, 129, 147	0
1	D	344/446 (77%)	0.08	19 (5%) 29	27	37, 75, 136, 153	0
2	E	6/6 (100%)	0.47	0 100	100	23, 39, 59, 61	0
2	F	6/6 (100%)	0.63	0 100	100	27, 34, 61, 73	0
2	G	6/6 (100%)	0.20	0 100	100	50, 57, 66, 73	0
2	H	4/6 (66%)	0.70	0 100	100	59, 61, 67, 79	0
3	M	5/5 (100%)	0.35	0 100	100	68, 71, 89, 91	0
4	N	4/4 (100%)	-0.20	0 100	100	44, 55, 65, 79	0
4	O	4/4 (100%)	-0.16	0 100	100	78, 87, 94, 99	0
4	P	3/4 (75%)	2.24	2 (66%) 0	0	108, 108, 118, 123	0
All	All	1684/1825 (92%)	-0.11	39 (2%) 64	64	20, 59, 119, 153	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	438	LYS	5.0
1	D	432	ALA	4.0
4	P	4	THR	3.9
1	B	439	LEU	3.8
1	D	414	ASP	3.4

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.