



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 6, 2023 – 06:27 am GMT

PDB ID : 2BS9
Title : Native crystal structure of a GH39 beta-xylosidase XynB1 from *Geobacillus stearothermophilus*
Authors : Czjzek, M.; Bravman, T.; Henrissat, B.; Shoham, Y.
Deposited on : 2005-05-19
Resolution : 2.20 Å (reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.36
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.36

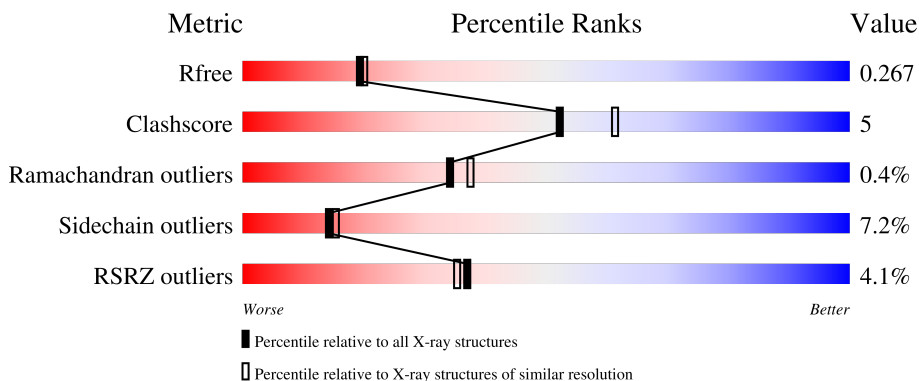
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION


The reported resolution of this entry is 2.20 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.



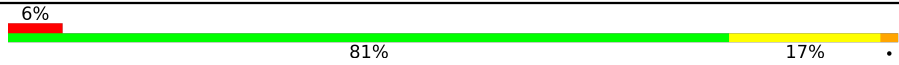


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. 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 $\leq 5\%$. 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	503	 4% 81% 16% •
1	B	503	 5% 83% 14% •
1	C	503	 5% 82% 16% •
1	D	503	 4% 79% 18% •
1	E	503	 2% 84% 12% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	503	 6% 81% 17% •
1	G	503	 4% 83% 14% •
1	H	503	 3% 83% 15% •

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 34099 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 BETA-XYLOSIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	501	4091	2643	693	744	11	0	0	0
1	B	501	4091	2643	693	744	11	0	0	0
1	C	501	4088	2642	693	742	11	0	0	0
1	D	501	4091	2643	693	744	11	0	0	0
1	E	501	4091	2643	693	744	11	0	0	0
1	F	501	4091	2643	693	744	11	0	0	0
1	G	501	4091	2643	693	744	11	0	0	0
1	H	501	4091	2643	693	744	11	0	0	0

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	GLY	LYS	conflict	UNP Q9ZFM2
A	?	-	LEU	deletion	UNP Q9ZFM2
A	?	-	GLU	deletion	UNP Q9ZFM2
A	406	GLU	PHE	conflict	UNP Q9ZFM2
A	445	ARG	PRO	conflict	UNP Q9ZFM2
A	446	GLN	-	insertion	UNP Q9ZFM2
A	447	VAL	SER	conflict	UNP Q9ZFM2
B	2	GLY	LYS	conflict	UNP Q9ZFM2
B	?	-	LEU	deletion	UNP Q9ZFM2
B	?	-	GLU	deletion	UNP Q9ZFM2
B	406	GLU	PHE	conflict	UNP Q9ZFM2
B	445	ARG	PRO	conflict	UNP Q9ZFM2
B	446	GLN	-	insertion	UNP Q9ZFM2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	447	VAL	SER	conflict	UNP Q9ZFM2
C	2	GLY	LYS	conflict	UNP Q9ZFM2
C	?	-	LEU	deletion	UNP Q9ZFM2
C	?	-	GLU	deletion	UNP Q9ZFM2
C	406	GLU	PHE	conflict	UNP Q9ZFM2
C	445	ARG	PRO	conflict	UNP Q9ZFM2
C	446	GLN	-	insertion	UNP Q9ZFM2
C	447	VAL	SER	conflict	UNP Q9ZFM2
D	2	GLY	LYS	conflict	UNP Q9ZFM2
D	?	-	LEU	deletion	UNP Q9ZFM2
D	?	-	GLU	deletion	UNP Q9ZFM2
D	406	GLU	PHE	conflict	UNP Q9ZFM2
D	445	ARG	PRO	conflict	UNP Q9ZFM2
D	446	GLN	-	insertion	UNP Q9ZFM2
D	447	VAL	SER	conflict	UNP Q9ZFM2
E	2	GLY	LYS	conflict	UNP Q9ZFM2
E	?	-	LEU	deletion	UNP Q9ZFM2
E	?	-	GLU	deletion	UNP Q9ZFM2
E	406	GLU	PHE	conflict	UNP Q9ZFM2
E	445	ARG	PRO	conflict	UNP Q9ZFM2
E	446	GLN	-	insertion	UNP Q9ZFM2
E	447	VAL	SER	conflict	UNP Q9ZFM2
F	2	GLY	LYS	conflict	UNP Q9ZFM2
F	?	-	LEU	deletion	UNP Q9ZFM2
F	?	-	GLU	deletion	UNP Q9ZFM2
F	406	GLU	PHE	conflict	UNP Q9ZFM2
F	445	ARG	PRO	conflict	UNP Q9ZFM2
F	446	GLN	-	insertion	UNP Q9ZFM2
F	447	VAL	SER	conflict	UNP Q9ZFM2
G	2	GLY	LYS	conflict	UNP Q9ZFM2
G	?	-	LEU	deletion	UNP Q9ZFM2
G	?	-	GLU	deletion	UNP Q9ZFM2
G	406	GLU	PHE	conflict	UNP Q9ZFM2
G	445	ARG	PRO	conflict	UNP Q9ZFM2
G	446	GLN	-	insertion	UNP Q9ZFM2
G	447	VAL	SER	conflict	UNP Q9ZFM2
H	2	GLY	LYS	conflict	UNP Q9ZFM2
H	?	-	LEU	deletion	UNP Q9ZFM2
H	?	-	GLU	deletion	UNP Q9ZFM2
H	406	GLU	PHE	conflict	UNP Q9ZFM2
H	445	ARG	PRO	conflict	UNP Q9ZFM2
H	446	GLN	-	insertion	UNP Q9ZFM2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	447	VAL	SER	conflict	UNP Q9ZFM2

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Ca 1 1	0	0
2	C	1	Total Ca 1 1	0	0
2	D	1	Total Ca 1 1	0	0
2	E	1	Total Ca 1 1	0	0
2	F	1	Total Ca 1 1	0	0
2	G	1	Total Ca 1 1	0	0
2	H	1	Total Ca 1 1	0	0

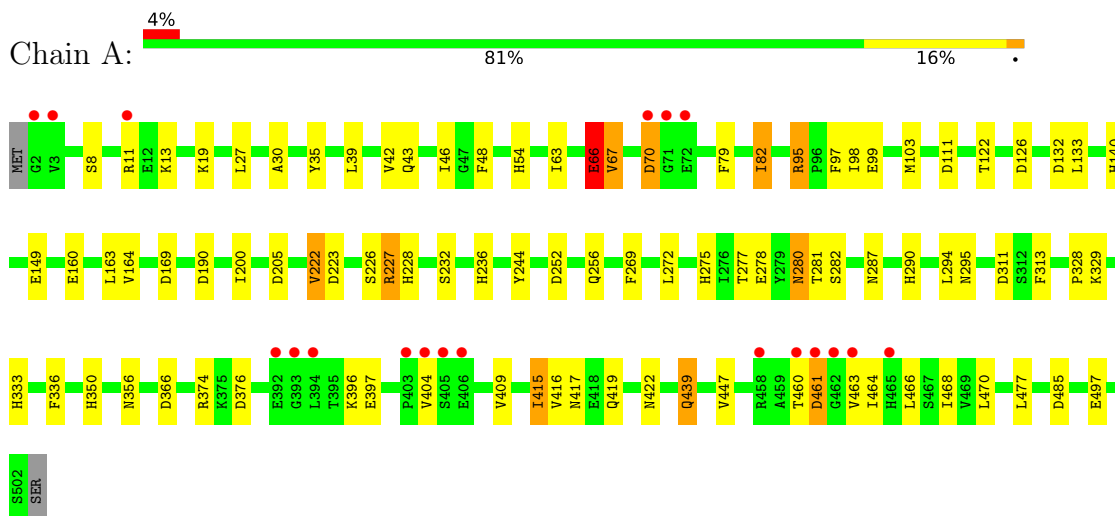
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	170	Total O 170 170	0	0
3	B	161	Total O 161 161	0	0
3	C	177	Total O 177 177	0	0
3	D	172	Total O 172 172	0	0
3	E	180	Total O 180 180	0	0
3	F	140	Total O 140 140	0	0
3	G	179	Total O 179 179	0	0
3	H	188	Total O 188 188	0	0

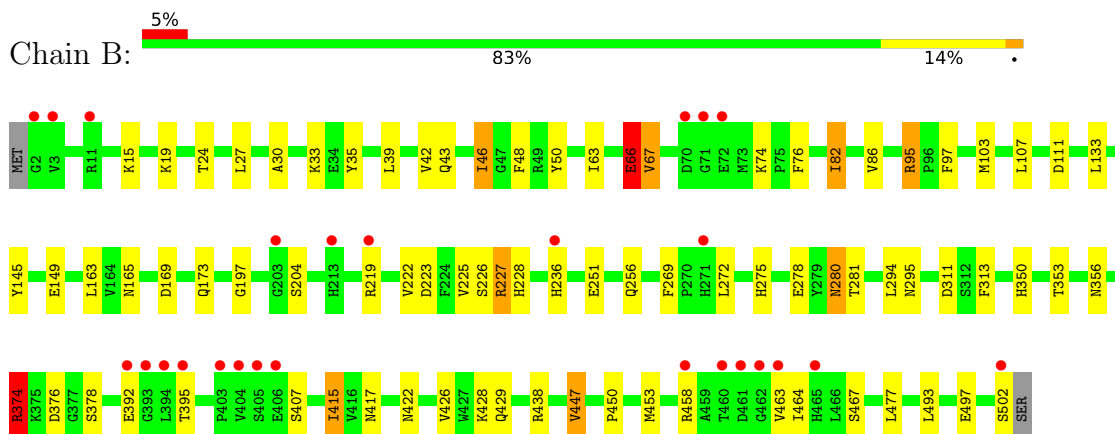
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

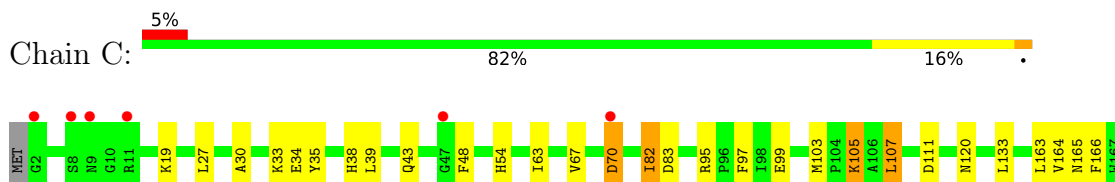
• Molecule 1: BETA-XYLOSIDASE

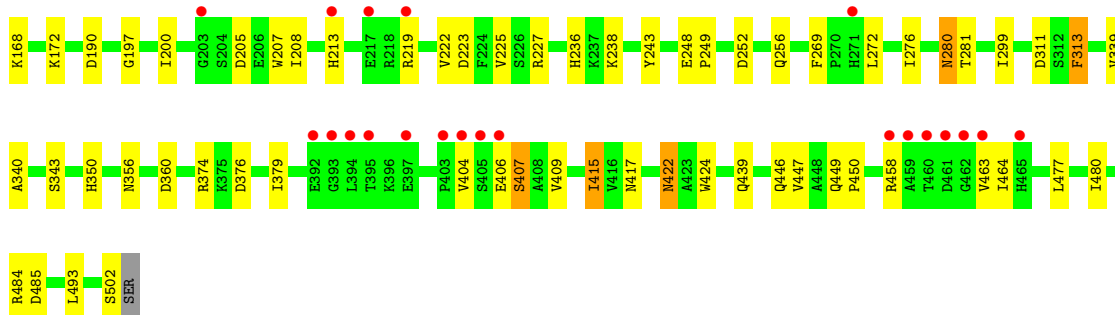


• Molecule 1: BETA-XYLOSIDASE

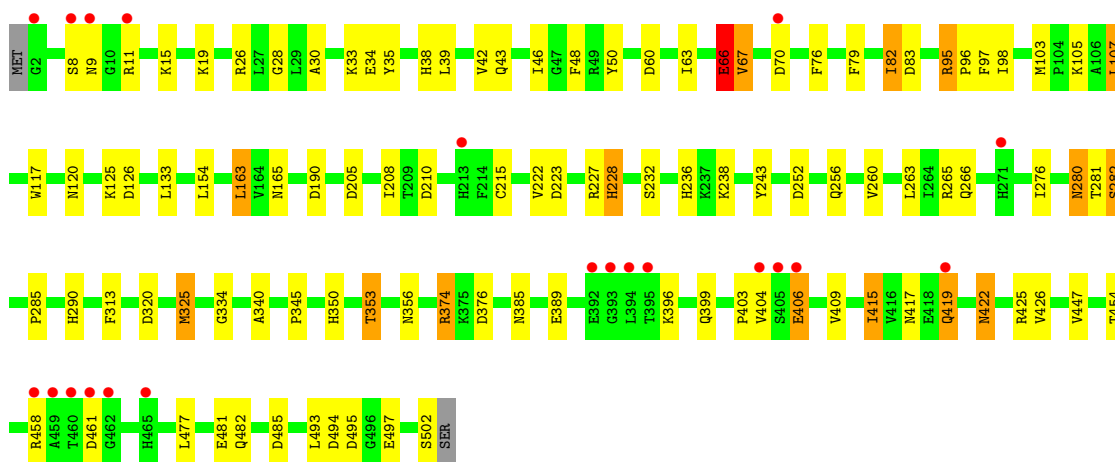
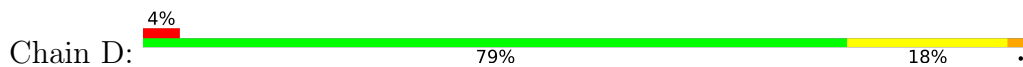


• Molecule 1: BETA-XYLOSIDASE

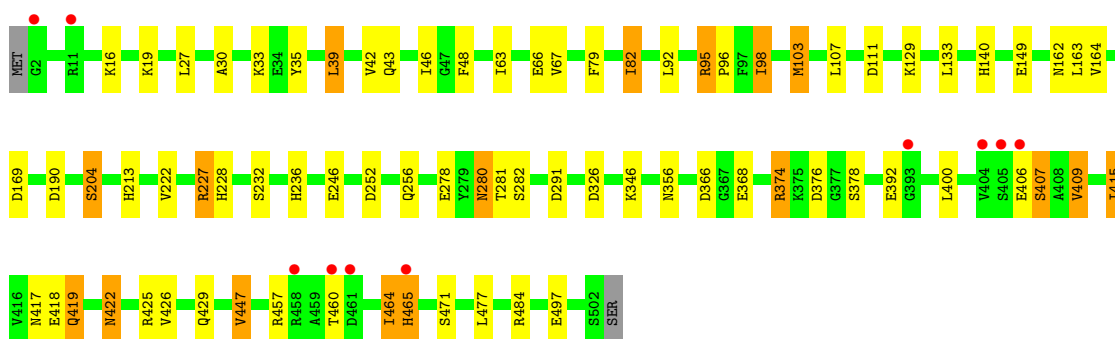
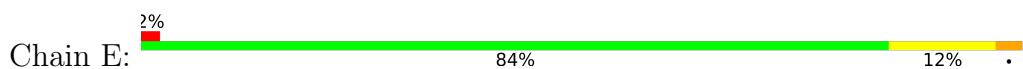




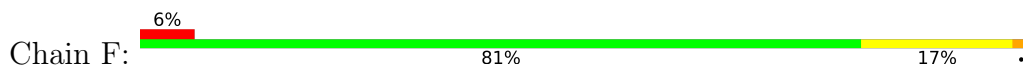
● Molecule 1: BETA-XYLOSIDASE

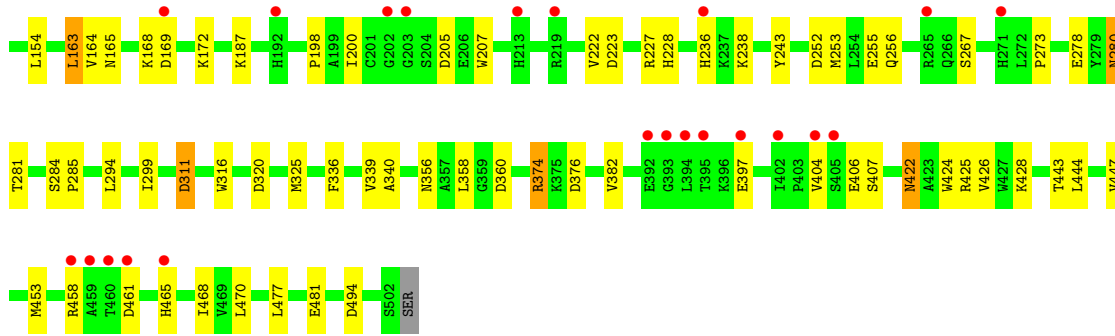


● Molecule 1: BETA-XYLOSIDASE

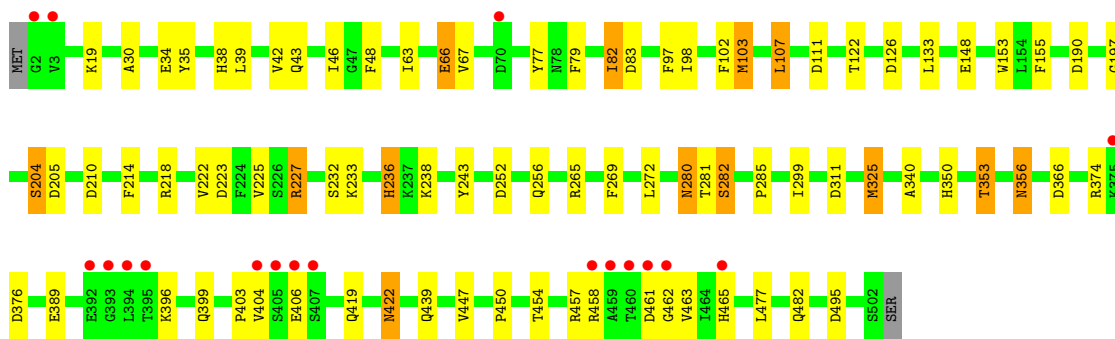
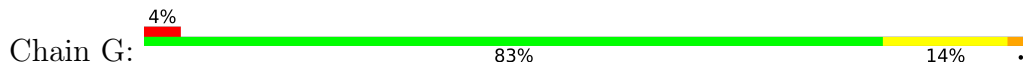


● Molecule 1: BETA-XYLOSIDASE

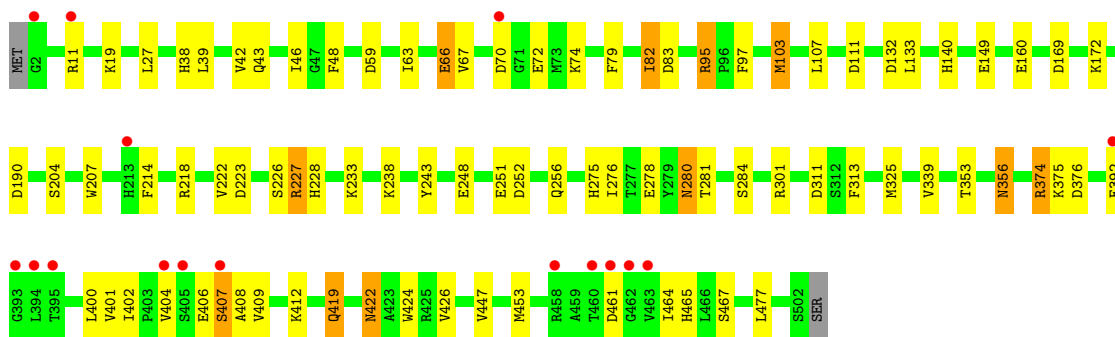
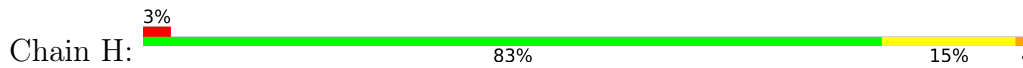




• Molecule 1: BETA-XYLOSIDASE



• Molecule 1: BETA-XYLOSIDASE



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	92.67Å 165.74Å 311.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.20 24.96 – 2.10	Depositor EDS
% Data completeness (in resolution range)	98.0 (50.00-2.20) 98.1 (24.96-2.10)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.32 (at 2.10Å)	Xtrriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.209 , 0.259 0.223 , 0.267	Depositor DCC
R_{free} test set	13754 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	13.6	Xtrriage
Anisotropy	0.613	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 48.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	34099	wwPDB-VP
Average B, all atoms (Å ²)	18.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 23.07 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 5.0212e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CA

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	0.61	0/4206	0.82	16/5710 (0.3%)
1	B	0.59	0/4206	0.80	9/5710 (0.2%)
1	C	0.58	0/4203	0.80	10/5706 (0.2%)
1	D	0.60	0/4206	0.81	16/5710 (0.3%)
1	E	0.60	1/4206 (0.0%)	0.80	6/5710 (0.1%)
1	F	0.55	0/4206	0.81	14/5710 (0.2%)
1	G	0.59	1/4206 (0.0%)	0.82	15/5710 (0.3%)
1	H	0.60	1/4206 (0.0%)	0.83	12/5710 (0.2%)
All	All	0.59	3/33645 (0.0%)	0.81	98/45676 (0.2%)

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	E	0	1
1	H	0	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	H	103	MET	SD-CE	-6.26	1.42	1.77
1	E	103	MET	SD-CE	-5.20	1.48	1.77
1	G	103	MET	SD-CE	-5.16	1.49	1.77

The worst 5 of 98 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	376	ASP	CB-CG-OD2	7.59	125.13	118.30
1	D	190	ASP	CB-CG-OD2	7.33	124.90	118.30
1	B	376	ASP	CB-CG-OD2	7.11	124.70	118.30
1	C	70	ASP	CB-CG-OD2	7.10	124.69	118.30
1	G	111	ASP	CB-CG-OD2	7.08	124.67	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	66	GLU	Peptide
1	H	66	GLU	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	4091	0	4001	44	0
1	B	4091	0	4001	45	0
1	C	4088	0	3999	44	0
1	D	4091	0	4001	54	0
1	E	4091	0	4001	41	0
1	F	4091	0	4001	49	0
1	G	4091	0	4001	37	0
1	H	4091	0	4001	50	0
2	A	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
2	G	1	0	0	0	0
2	H	1	0	0	0	0
3	A	170	0	0	4	0
3	B	161	0	0	2	0
3	C	177	0	0	3	0
3	D	172	0	0	5	0
3	E	180	0	0	5	0
3	F	140	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	179	0	0	4	0
3	H	188	0	0	8	0
All	All	34099	0	32006	354	0

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 5.

The worst 5 of 354 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:D:227:ARG:HH12	1:D:256:GLN:HE21	1.12	0.93
1:F:426:VAL:HG21	1:F:447:VAL:HG11	1.51	0.91
1:B:227:ARG:HH12	1:B:256:GLN:HE21	1.09	0.91
1:B:42:VAL:O	1:B:46:ILE:O	1.89	0.90
1:D:280:ASN:HD22	1:D:281:THR:H	1.19	0.88

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

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	499/503 (99%)	473 (95%)	24 (5%)	2 (0%)	34 37
1	B	499/503 (99%)	476 (95%)	21 (4%)	2 (0%)	34 37
1	C	499/503 (99%)	477 (96%)	21 (4%)	1 (0%)	47 55
1	D	499/503 (99%)	474 (95%)	23 (5%)	2 (0%)	34 37
1	E	499/503 (99%)	476 (95%)	19 (4%)	4 (1%)	19 19
1	F	499/503 (99%)	476 (95%)	22 (4%)	1 (0%)	47 55
1	G	499/503 (99%)	476 (95%)	20 (4%)	3 (1%)	25 26
1	H	499/503 (99%)	473 (95%)	24 (5%)	2 (0%)	34 37

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	3992/4024 (99%)	3801 (95%)	174 (4%)	17 (0%)	34 37

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	67	VAL
1	B	407	SER
1	D	67	VAL
1	E	67	VAL
1	E	407	SER

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	440/442 (100%)	405 (92%)	35 (8%)	12 12
1	B	440/442 (100%)	408 (93%)	32 (7%)	14 15
1	C	439/442 (99%)	407 (93%)	32 (7%)	14 15
1	D	440/442 (100%)	408 (93%)	32 (7%)	14 15
1	E	440/442 (100%)	411 (93%)	29 (7%)	16 19
1	F	440/442 (100%)	407 (92%)	33 (8%)	13 14
1	G	440/442 (100%)	409 (93%)	31 (7%)	15 16
1	H	440/442 (100%)	411 (93%)	29 (7%)	16 19
All	All	3519/3536 (100%)	3266 (93%)	253 (7%)	14 15

5 of 253 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	280	ASN
1	G	458	ARG
1	E	246	GLU
1	G	447	VAL
1	H	280	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 112 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	419	GLN
1	H	429	GLN
1	E	449	GLN
1	H	422	ASN
1	G	439	GLN

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 7 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues

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	501/503 (99%)	-0.00	19 (3%) 40 38	10, 16, 39, 45	0
1	B	501/503 (99%)	0.03	26 (5%) 27 26	7, 16, 40, 46	0
1	C	501/503 (99%)	0.05	27 (5%) 25 24	8, 16, 39, 44	0
1	D	501/503 (99%)	0.07	21 (4%) 36 34	11, 17, 39, 45	0
1	E	501/503 (99%)	-0.07	10 (1%) 65 63	8, 16, 38, 44	0
1	F	501/503 (99%)	0.14	28 (5%) 24 23	9, 16, 39, 44	0
1	G	501/503 (99%)	-0.08	18 (3%) 42 41	10, 16, 38, 45	0
1	H	501/503 (99%)	-0.07	16 (3%) 47 45	8, 16, 38, 45	0
All	All	4008/4024 (99%)	0.01	165 (4%) 37 35	7, 16, 40, 46	0

The worst 5 of 165 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	2	GLY	7.6
1	F	405	SER	7.0
1	D	405	SER	6.3
1	H	2	GLY	6.3
1	H	11	ARG	5.9

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 monosaccharides in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	CA	F	1503	1/1	0.97	0.10	29,29,29,29	0
2	CA	D	1503	1/1	0.98	0.16	24,24,24,24	0
2	CA	E	1503	1/1	0.99	0.15	24,24,24,24	0
2	CA	C	1503	1/1	0.99	0.11	24,24,24,24	0
2	CA	G	1503	1/1	0.99	0.12	22,22,22,22	0
2	CA	H	1503	1/1	0.99	0.10	22,22,22,22	0
2	CA	A	1503	1/1	1.00	0.10	23,23,23,23	0

6.5 Other polymers [i](#)

There are no such residues in this entry.