wwPDB X-ray Structure Validation Summary Report (i)

Oct 13, 2024 – 01:31 pm BST

PDB ID	:	1UP4
Title	:	Structure of the 6-phospho-beta glucosidase from Thermotoga maritima at 2.85 Angstrom resolution in the monoclinic form
Authors	:	Varrot, A.; Yip, V.; Withers, S.G.; Davies, G.J.
Deposited on	:	2003-09-26
Resolution	:	2.85 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 (i) 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 (i)) were used in the production of this report:

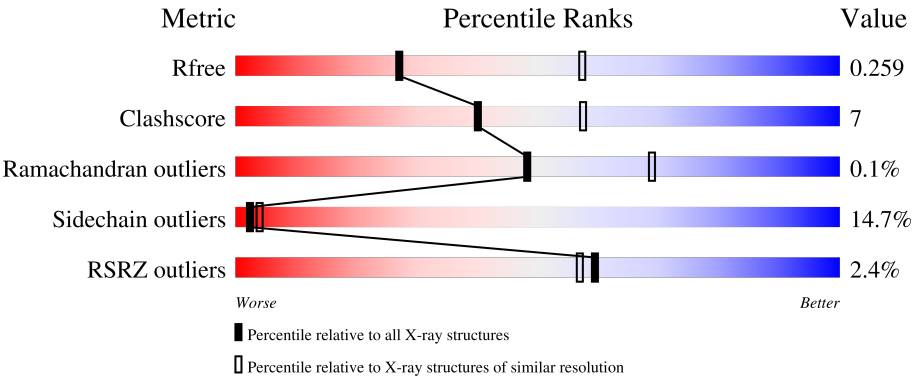
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 2.85 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.



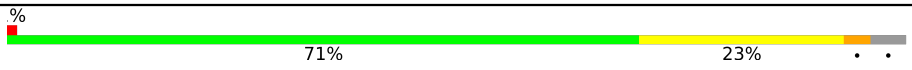
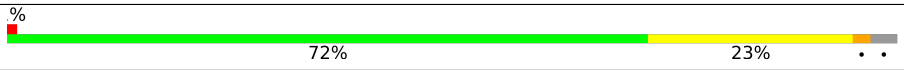
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1268 (2.88-2.84)
Clashscore	180529	1351 (2.88-2.84)
Ramachandran outliers	177936	1318 (2.88-2.84)
Sidechain outliers	177891	1319 (2.88-2.84)
RSRZ outliers	164620	1269 (2.88-2.84)

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 <=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	415	5% 69% 25% 5%
1	B	415	2% 72% 22%
1	C	415	2% 72% 23%
1	D	415	2% 70% 22%
1	E	415	3% 73% 20% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	415	
1	G	415	
1	H	415	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 26328 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 6-PHOSPHO-BETA-GLUCOSIDASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	410	Total	C	N	O	S	Se	88	0	0
			3326	2143	554	620	2	7			
1	B	403	Total	C	N	O	S	Se	67	1	0
			3274	2112	544	609	2	7			
1	C	408	Total	C	N	O	S	Se	86	0	0
			3309	2133	551	616	2	7			
1	D	400	Total	C	N	O	S	Se	90	0	0
			3251	2099	539	604	2	7			
1	E	396	Total	C	N	O	S	Se	147	0	0
			3210	2073	533	595	2	7			
1	F	399	Total	C	N	O	S	Se	88	0	0
			3242	2092	538	603	2	7			
1	G	409	Total	C	N	O	S	Se	55	16	0
			3317	2137	552	619	2	7			
1	H	403	Total	C	N	O	S	Se	115	30	0
			3274	2111	542	612	2	7			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	20	Total	O	0	0
			20	20		
2	B	16	Total	O	0	0
			16	16		
2	C	17	Total	O	0	0
			17	17		
2	D	14	Total	O	0	0
			14	14		
2	E	13	Total	O	0	0
			13	13		
2	F	13	Total	O	0	0
			13	13		

Continued on next page...

Continued from previous page...

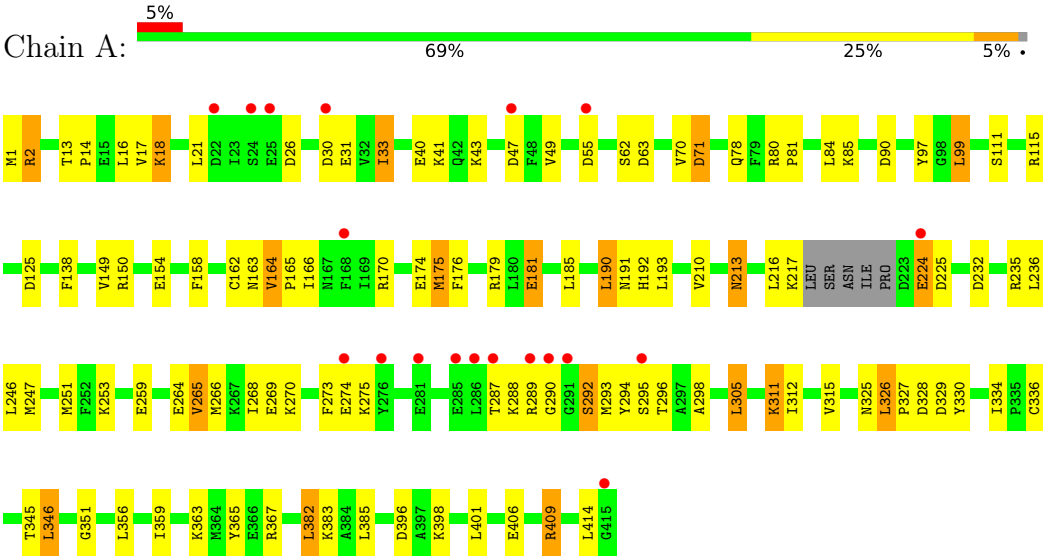
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	G	14	Total O 14 14	0	0
2	H	18	Total O 18 18	0	0

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.

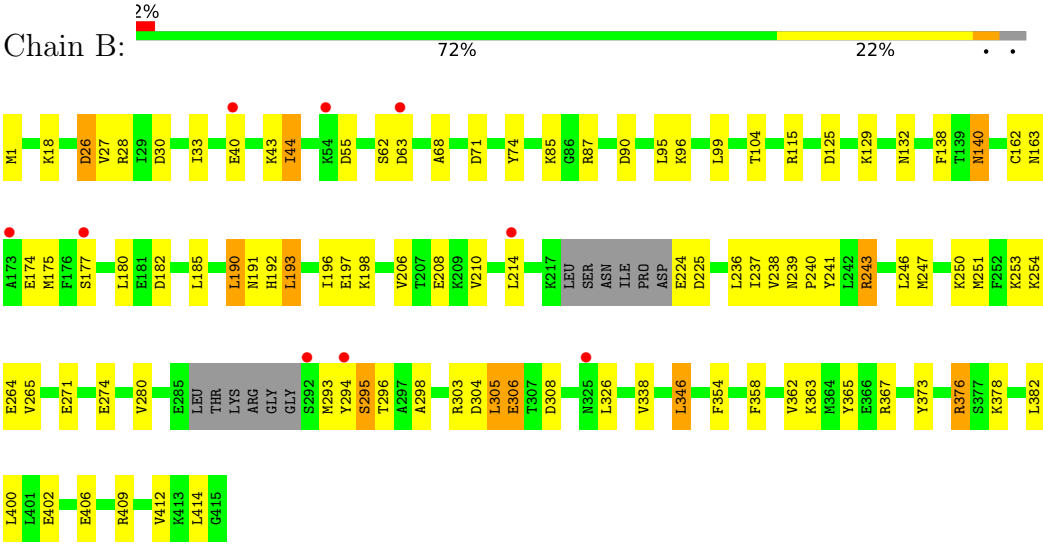
• Molecule 1: 6-PHOSPHO-BETA-GLUCOSIDASE

Chain A:



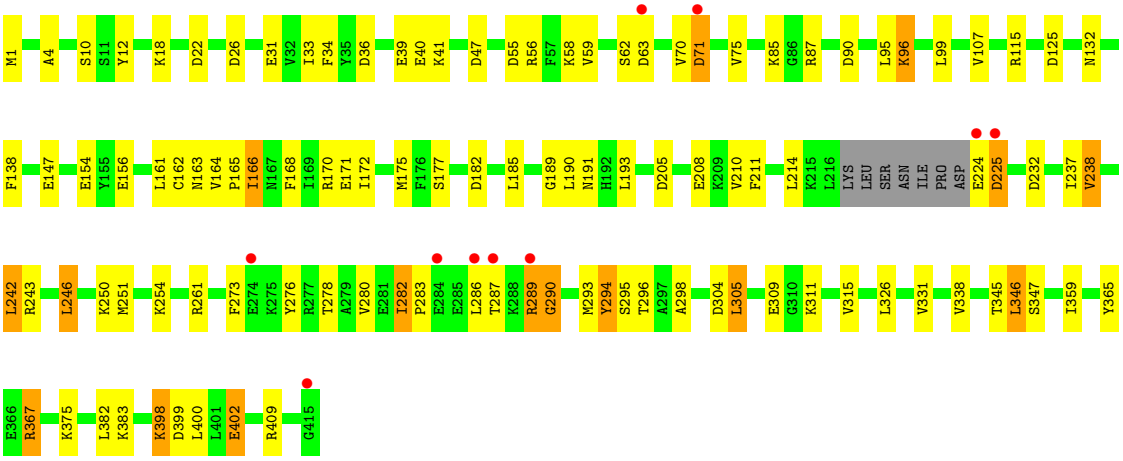
• Molecule 1: 6-PHOSPHO-BETA-GLUCOSIDASE

Chain B:



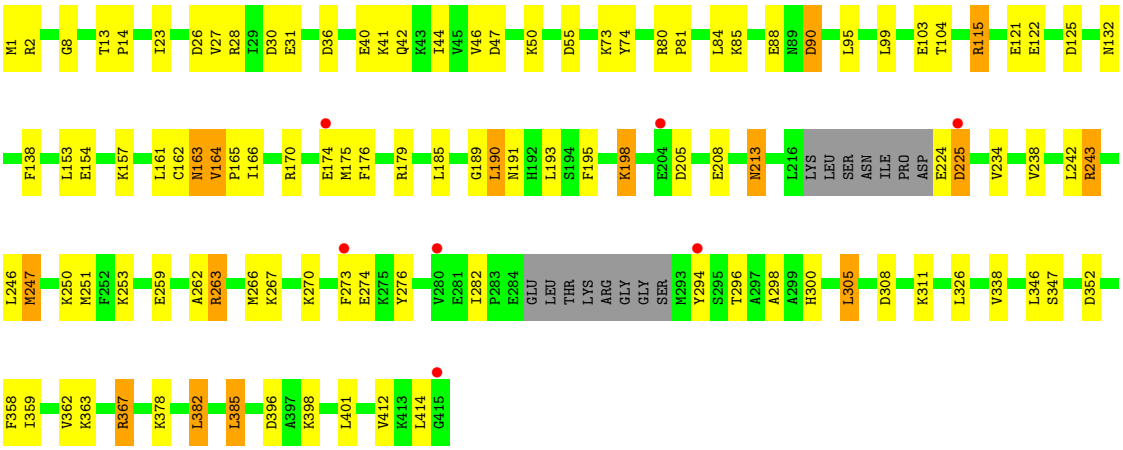
• Molecule 1: 6-PHOSPHO-BETA-GLUCOSIDASE

Chain C:



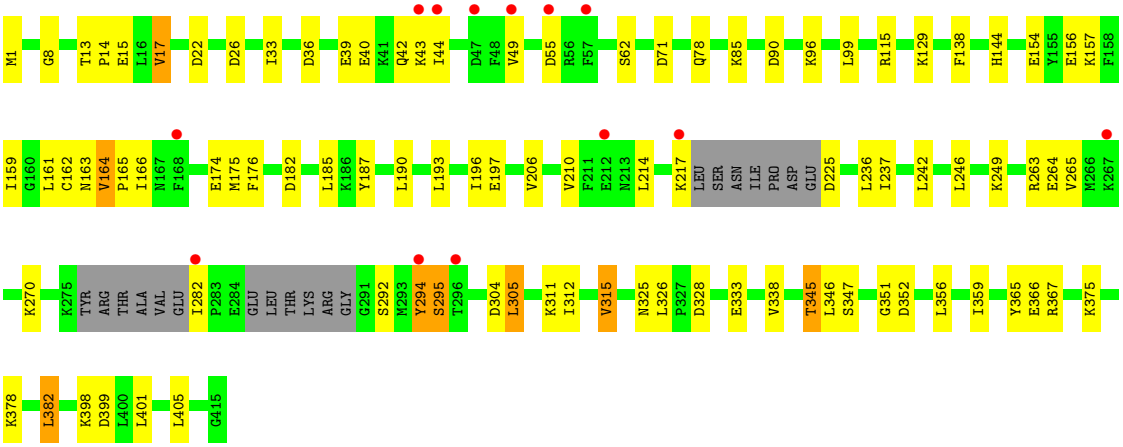
• Molecule 1: 6-PHOSPHO-BETA-GLUCOSIDASE

Chain D:



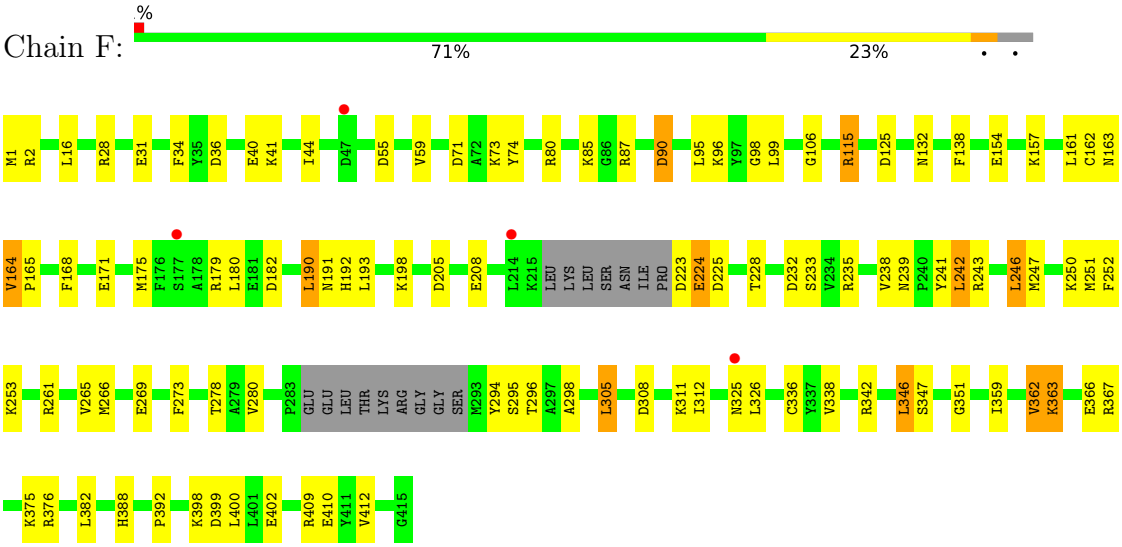
• Molecule 1: 6-PHOSPHO-BETA-GLUCOSIDASE

Chain E:



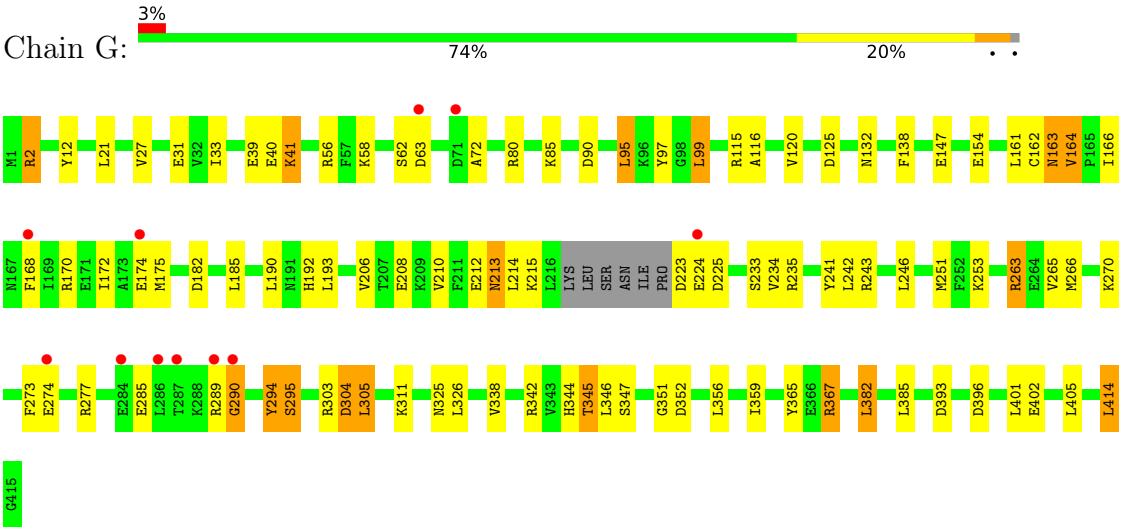
• Molecule 1: 6-PHOSPHO-BETA-GLUCOSIDASE

Chain F:



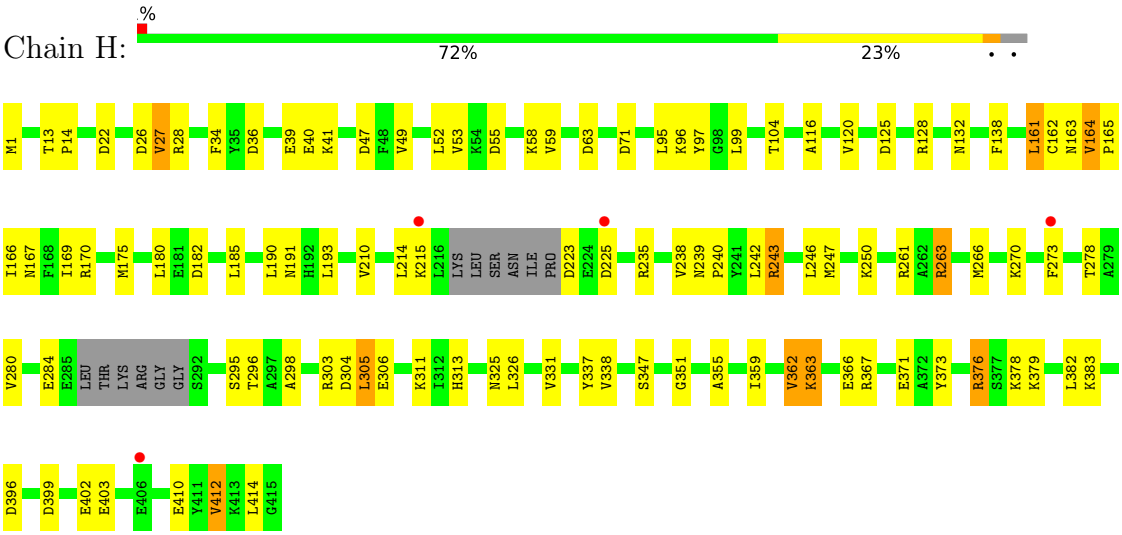
• Molecule 1: 6-PHOSPHO-BETA-GLUCOSIDASE

Chain G:



• Molecule 1: 6-PHOSPHO-BETA-GLUCOSIDASE

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	99.05Å 188.31Å 125.63Å 90.00° 104.52° 90.00°	Depositor
Resolution (Å)	40.00 – 2.85 40.00 – 2.85	Depositor EDS
% Data completeness (in resolution range)	99.6 (40.00-2.85) 99.5 (40.00-2.85)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.17 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.196 , 0.235 0.237 , 0.259	Depositor DCC
R_{free} test set	5154 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	37.2	Xtriage
Anisotropy	0.185	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 24.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	26328	wwPDB-VP
Average B, all atoms (Å^2)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.95% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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	0.63	0/3385	0.83	9/4550 (0.2%)
1	B	0.67	0/3337	0.83	8/4487 (0.2%)
1	C	0.58	0/3368	0.80	13/4528 (0.3%)
1	D	0.58	0/3310	0.83	12/4452 (0.3%)
1	E	0.54	0/3266	0.78	8/4387 (0.2%)
1	F	0.61	0/3301	0.82	9/4440 (0.2%)
1	G	0.66	0/3376	0.85	8/4539 (0.2%)
1	H	0.60	0/3332	0.82	11/4480 (0.2%)
All	All	0.61	0/26675	0.82	78/35863 (0.2%)

There are no bond length outliers.

The worst 5 of 78 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	63	ASP	CB-CG-OD2	8.05	125.55	118.30
1	H	125	ASP	CB-CG-OD2	7.50	125.05	118.30
1	E	182	ASP	CB-CG-OD2	6.74	124.36	118.30
1	C	232	ASP	CB-CG-OD2	6.55	124.20	118.30
1	H	399	ASP	CB-CG-OD2	6.54	124.19	118.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	3326	0	3359	58	0
1	B	3274	0	3302	39	0
1	C	3309	0	3342	40	0
1	D	3251	0	3280	46	0
1	E	3210	0	3245	34	0
1	F	3242	0	3267	42	0
1	G	3317	0	3299	50	0
1	H	3274	0	3208	36	0
2	A	20	0	0	0	0
2	B	16	0	0	0	0
2	C	17	0	0	0	0
2	D	14	0	0	1	0
2	E	13	0	0	0	0
2	F	13	0	0	1	0
2	G	14	0	0	0	0
2	H	18	0	0	0	0
All	All	26328	0	26302	331	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 7.

The worst 5 of 331 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:F:224:GLU:HG3	1:F:251:MSE:HE1	1.40	1.01
1:G:289:ARG:HG3	1:G:290:GLY:H	1.25	1.01
1:G:263:ARG:HH11	1:G:263:ARG:HG3	1.34	0.92
1:A:224:GLU:HG2	1:A:251:MSE:HE1	1.61	0.82
1:F:224:GLU:CG	1:F:251:MSE:HE1	2.11	0.80

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	406/415 (98%)	390 (96%)	14 (3%)	2 (0%)	25 43
1	B	398/415 (96%)	390 (98%)	8 (2%)	0	100 100
1	C	404/415 (97%)	392 (97%)	11 (3%)	1 (0%)	44 63
1	D	394/415 (95%)	386 (98%)	8 (2%)	0	100 100
1	E	388/415 (94%)	380 (98%)	8 (2%)	0	100 100
1	F	393/415 (95%)	386 (98%)	7 (2%)	0	100 100
1	G	405/415 (98%)	392 (97%)	12 (3%)	1 (0%)	44 63
1	H	397/415 (96%)	389 (98%)	8 (2%)	0	100 100
All	All	3185/3320 (96%)	3105 (98%)	76 (2%)	4 (0%)	48 69

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	290	GLY
1	G	290	GLY
1	A	292	SER
1	C	290	GLY

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	362/360 (101%)	305 (84%)	57 (16%)	2 3
1	B	357/360 (99%)	304 (85%)	53 (15%)	2 4
1	C	360/360 (100%)	306 (85%)	54 (15%)	2 4
1	D	354/360 (98%)	300 (85%)	54 (15%)	2 3
1	E	350/360 (97%)	303 (87%)	47 (13%)	3 5
1	F	353/360 (98%)	299 (85%)	54 (15%)	2 3
1	G	361/360 (100%)	316 (88%)	45 (12%)	3 7
1	H	357/360 (99%)	302 (85%)	55 (15%)	2 3
All	All	2854/2880 (99%)	2435 (85%)	419 (15%)	2 4

5 of 419 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	154	GLU
1	F	198	LYS
1	H	273	PHE
1	E	185	LEU
1	E	365	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	353	HIS
1	E	191	ASN
1	H	320	ASN
1	E	167	ASN
1	E	320	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 oligosaccharides 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

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	403/415 (97%)	0.19	19 (4%)	37	33	9, 19, 34, 43	22 (5%)
1	B	396/415 (95%)	0.12	9 (2%)	61	58	11, 19, 32, 43	20 (5%)
1	C	401/415 (96%)	0.05	10 (2%)	58	55	8, 19, 34, 43	24 (5%)
1	D	393/415 (94%)	0.06	7 (1%)	67	64	11, 19, 31, 43	25 (6%)
1	E	389/415 (93%)	0.21	13 (3%)	49	44	9, 19, 33, 43	35 (8%)
1	F	392/415 (94%)	-0.02	4 (1%)	79	78	10, 19, 31, 44	23 (5%)
1	G	402/415 (96%)	0.12	11 (2%)	56	53	11, 19, 34, 46	16 (3%)
1	H	396/415 (95%)	0.06	4 (1%)	79	78	9, 19, 32, 43	29 (7%)
All	All	3172/3320 (95%)	0.10	77 (2%)	59	57	8, 19, 33, 46	194 (6%)

The worst 5 of 77 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	287	THR	4.7
1	E	55	ASP	4.3
1	E	47	ASP	4.2
1	G	71[G]	ASP	3.5
1	E	282	ILE	3.4

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.