wwPDB X-ray Structure Validation Summary Report (i)

Jun 13, 2024 – 10:45 AM EDT

PDB ID	:	4D1I
Title	:	The structure of the GH35 beta-galactosidase Bgl35A from Cellvibrio japonicus
Authors	:	Larsbrink, J.; Thompson, A.J.; Lundqvist, M.; Gardner, J.G.; Davies, G.J.; Brumer, H.
Deposited on	:	2014-05-02
Resolution	:	1.80 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:

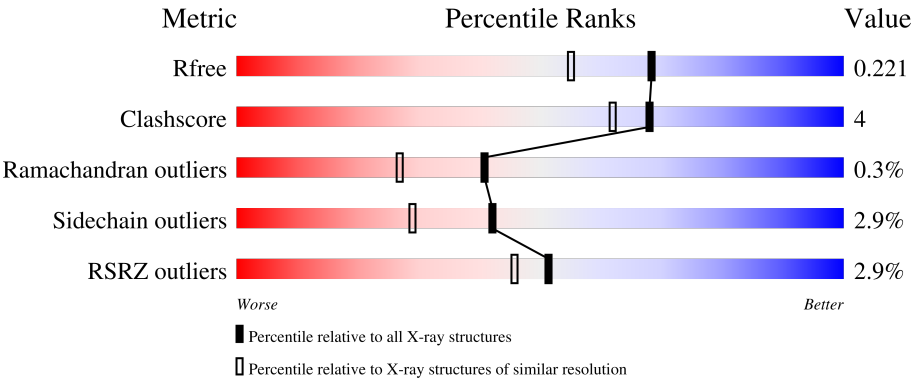
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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	540	4% 92% 7% •
1	B	540	4% 91% 7% •
1	C	540	3% 91% 8%
1	D	540	2% 92% 7% •
1	E	540	% 90% 7% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	540	2% 89% 9%
1	G	540	3% 90% 7%
1	H	540	4% 89% 8%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 37302 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-GALACTOSIDASE, PUTATIVE, BGL35A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	539	Total	C	N	O	S	0	2	0
			4196	2687	713	780	16			
1	B	539	Total	C	N	O	S	0	3	0
			4206	2695	714	781	16			
1	C	539	Total	C	N	O	S	0	6	0
			4216	2702	718	780	16			
1	D	539	Total	C	N	O	S	0	6	0
			4238	2714	719	789	16			
1	E	539	Total	C	N	O	S	0	3	0
			4201	2688	714	784	15			
1	F	539	Total	C	N	O	S	0	4	0
			4231	2706	717	793	15			
1	G	540	Total	C	N	O	S	0	4	0
			4233	2713	721	783	16			
1	H	539	Total	C	N	O	S	0	4	0
			4159	2658	702	783	16			

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

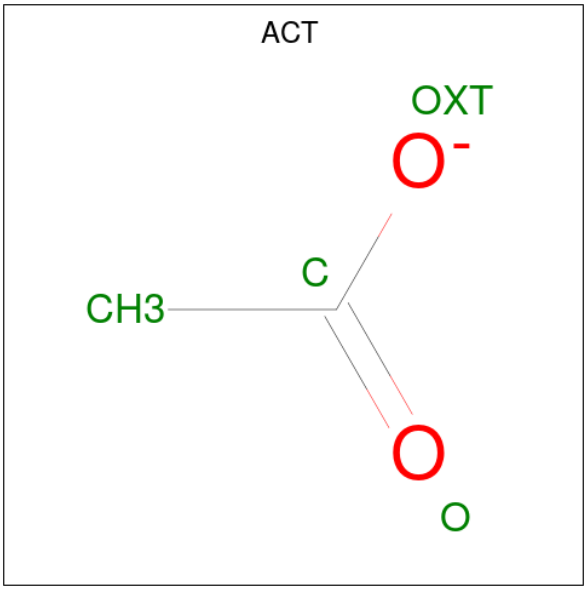
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Na	0	0
			1	1		
2	B	1	Total	Na	0	0
			1	1		
2	C	2	Total	Na	0	0
			2	2		
2	D	3	Total	Na	0	0
			3	3		
2	E	3	Total	Na	0	0
			3	3		
2	F	3	Total	Na	0	0
			3	3		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	G	3	Total Na 3 3	0	0
2	H	1	Total Na 1 1	0	0

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	D	1	Total C O 4 2 2	0	0
3	D	1	Total C O 8 4 4	0	1
3	H	1	Total C O 4 2 2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	373	Total O 375 375	0	2
4	B	445	Total O 445 445	0	0
4	C	419	Total O 420 420	0	1
4	D	542	Total O 543 543	0	1

Continued on next page...

Continued from previous page...

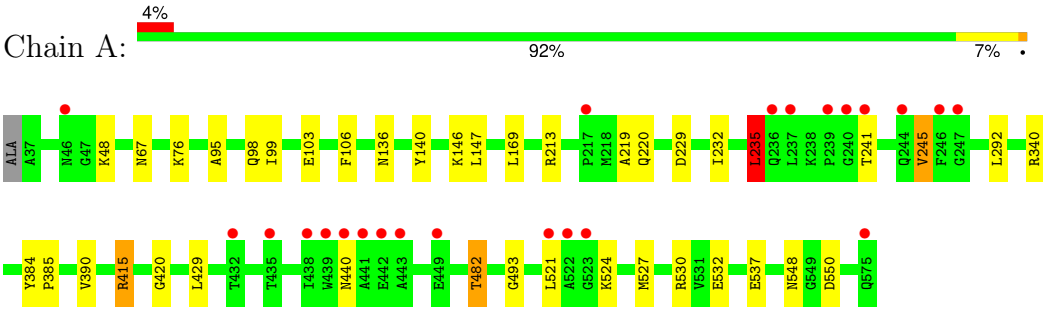
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	520	Total O 520 520	0	0
4	F	475	Total O 476 476	0	1
4	G	426	Total O 427 427	0	1
4	H	381	Total O 383 383	0	2

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.

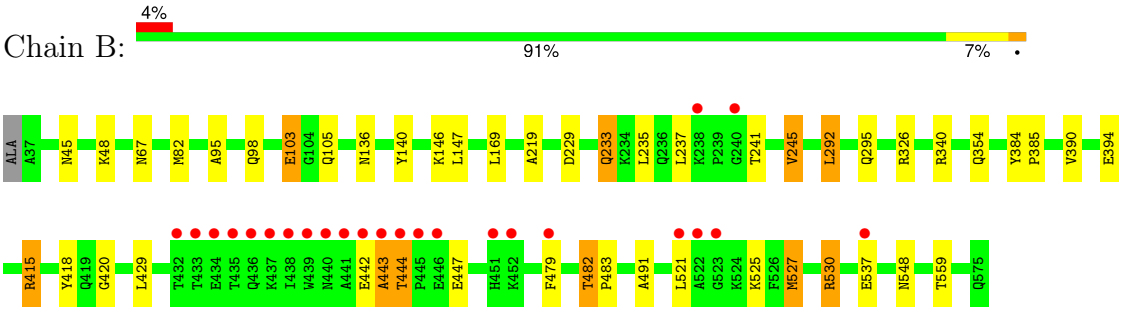
• Molecule 1: BETA-GALACTOSIDASE, PUTATIVE, BGL35A

Chain A:



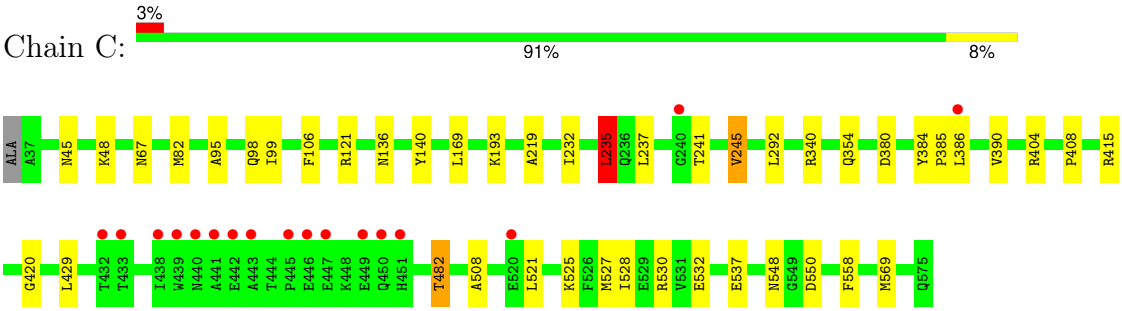
• Molecule 1: BETA-GALACTOSIDASE, PUTATIVE, BGL35A

Chain B:



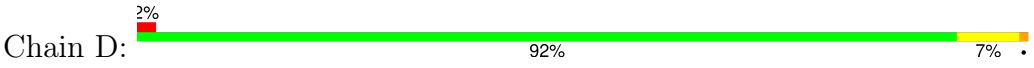
• Molecule 1: BETA-GALACTOSIDASE, PUTATIVE, BGL35A

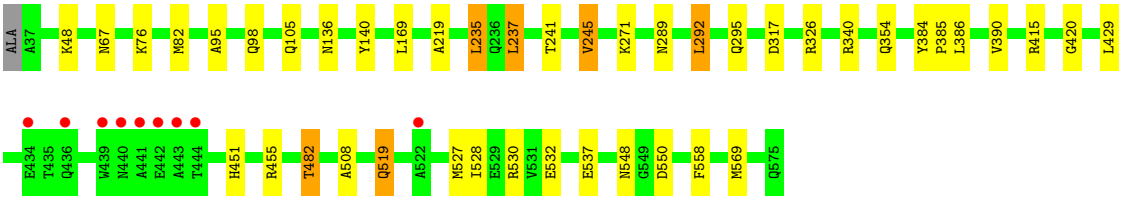
Chain C:



• Molecule 1: BETA-GALACTOSIDASE, PUTATIVE, BGL35A

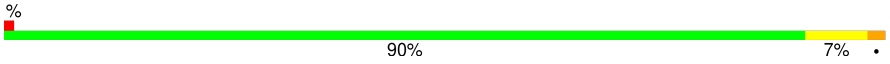
Chain D:

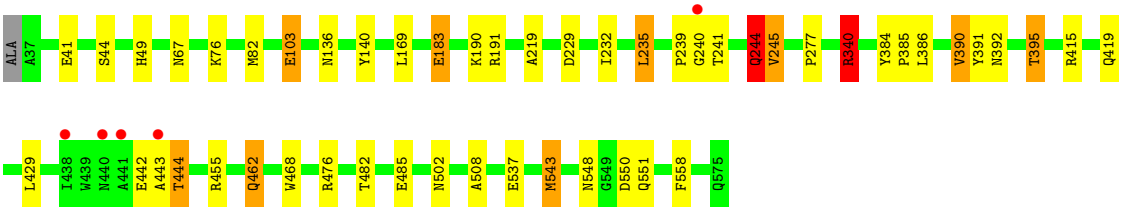




• Molecule 1: BETA-GALACTOSIDASE, PUTATIVE, BGL35A

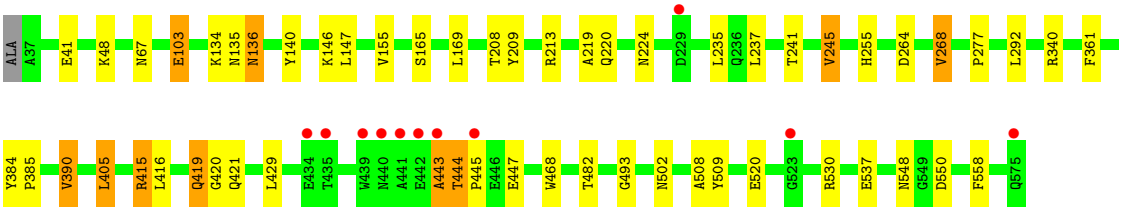
Chain E:





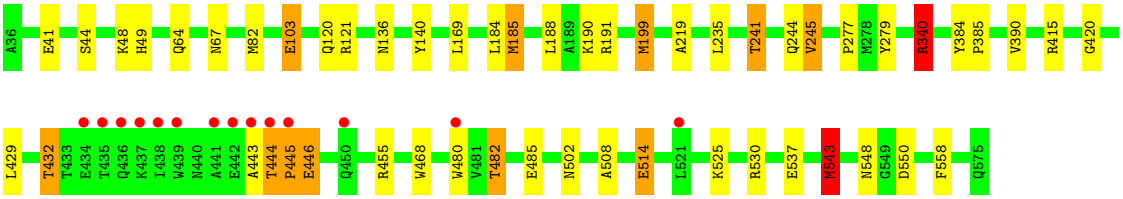
• Molecule 1: BETA-GALACTOSIDASE, PUTATIVE, BGL35A

Chain F:



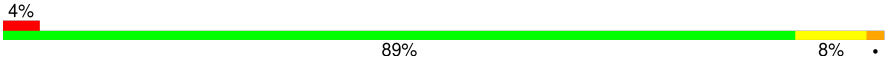
• Molecule 1: BETA-GALACTOSIDASE, PUTATIVE, BGL35A

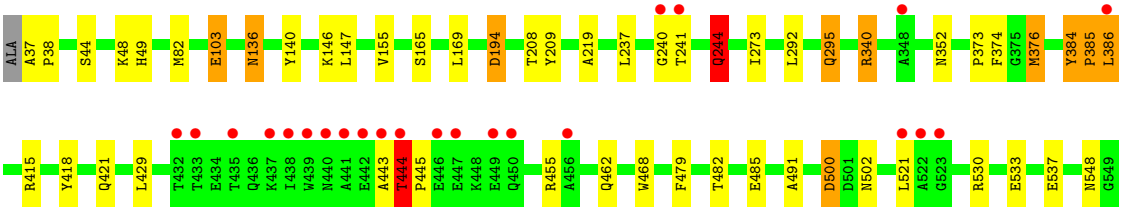
Chain G:



• Molecule 1: BETA-GALACTOSIDASE, PUTATIVE, BGL35A

Chain H:






D550
H563
Q575

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	98.91Å 115.78Å 116.04Å 90.21° 90.25° 90.38°	Depositor
Resolution (Å)	116.03 – 1.80 46.02 – 1.80	Depositor EDS
% Data completeness (in resolution range)	96.4 (116.03-1.80) 96.4 (46.02-1.80)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.68 (at 1.79Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.185 , 0.212 0.195 , 0.221	Depositor DCC
R_{free} test set	23101 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	23.8	Xtriage
Anisotropy	0.134	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 44.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.075 for h,l,-k 0.075 for h,-l,k 0.053 for h,-k,-l 0.036 for -h,k,-l 0.036 for -h,-k,l 0.037 for -h,l,k 0.038 for -h,-l,-k	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	37302	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.20% 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

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

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	0.55	0/4313	0.73	6/5882 (0.1%)
1	B	0.55	0/4326	0.73	10/5898 (0.2%)
1	C	0.55	0/4344	0.74	5/5922 (0.1%)
1	D	0.59	0/4364	0.73	3/5949 (0.1%)
1	E	0.63	0/4320	0.92	16/5894 (0.3%)
1	F	0.62	0/4351	0.76	8/5932 (0.1%)
1	G	0.60	0/4356	0.86	13/5936 (0.2%)
1	H	0.62	2/4280 (0.0%)	0.89	18/5846 (0.3%)
All	All	0.59	2/34654 (0.0%)	0.80	79/47259 (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	A	0	1
1	B	0	1
1	E	0	2
1	F	0	2
1	G	0	1
1	H	0	3
All	All	0	10

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	H	194[A]	ASP	CB-CG	8.35	1.69	1.51
1	H	194[B]	ASP	CB-CG	8.35	1.69	1.51

The worst 5 of 79 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	340	ARG	NE-CZ-NH1	19.27	129.94	120.30
1	H	340	ARG	NE-CZ-NH2	-18.22	111.19	120.30
1	E	340[A]	ARG	NE-CZ-NH1	17.74	129.17	120.30
1	E	340[B]	ARG	NE-CZ-NH1	17.74	129.17	120.30
1	G	340[A]	ARG	NE-CZ-NH1	16.64	128.62	120.30

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	103	GLU	Peptide
1	B	103	GLU	Peptide
1	E	103	GLU	Peptide
1	E	442	GLU	Peptide
1	F	103	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	4196	0	4010	33	0
1	B	4206	0	4029	36	0
1	C	4216	0	4041	31	0
1	D	4238	0	4071	37	0
1	E	4201	0	4019	29	0
1	F	4231	0	4044	38	0
1	G	4233	0	4084	41	0
1	H	4159	0	3913	47	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	2	0	0	0	0
2	D	3	0	0	0	0
2	E	3	0	0	0	0
2	F	3	0	0	0	0
2	G	3	0	0	0	0
2	H	1	0	0	0	0
3	D	12	0	9	0	0
3	H	4	0	3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	375	0	0	10	0
4	B	445	0	0	9	0
4	C	420	0	0	8	0
4	D	543	0	0	14	0
4	E	520	0	0	8	0
4	F	476	0	0	8	0
4	G	427	0	0	11	0
4	H	383	0	0	4	0
All	All	37302	0	32223	250	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 4.

The worst 5 of 250 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:A:146:LYS:HE3	1:C:527[A]:MET:HE3	1.35	1.07
1:B:527[A]:MET:HE3	1:H:146:LYS:HE3	1.39	1.05
1:A:527[A]:MET:HE3	1:B:146:LYS:HE3	1.42	0.99
1:E:468:TRP:HE1	1:E:502:ASN:HD22	1.13	0.94
1:F:468:TRP:HE1	1:F:502:ASN:HD22	1.15	0.93

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	539/540 (100%)	521 (97%)	18 (3%)	0	100 100
1	B	540/540 (100%)	521 (96%)	15 (3%)	4 (1%)	22 10
1	C	543/540 (101%)	526 (97%)	17 (3%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	543/540 (101%)	527 (97%)	16 (3%)	0	100	100
1	E	540/540 (100%)	519 (96%)	20 (4%)	1 (0%)	47	33
1	F	541/540 (100%)	521 (96%)	19 (4%)	1 (0%)	47	33
1	G	542/540 (100%)	523 (96%)	16 (3%)	3 (1%)	25	12
1	H	541/540 (100%)	519 (96%)	18 (3%)	4 (1%)	22	10
All	All	4329/4320 (100%)	4177 (96%)	139 (3%)	13 (0%)	41	27

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	444	THR
1	G	445	PRO
1	H	385	PRO
1	H	444	THR
1	H	445	PRO

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	426/452 (94%)	416 (98%)	10 (2%)	50	37
1	B	428/452 (95%)	416 (97%)	12 (3%)	43	30
1	C	428/452 (95%)	415 (97%)	13 (3%)	41	27
1	D	435/452 (96%)	422 (97%)	13 (3%)	41	27
1	E	428/452 (95%)	412 (96%)	16 (4%)	34	19
1	F	433/452 (96%)	421 (97%)	12 (3%)	43	30
1	G	433/452 (96%)	417 (96%)	16 (4%)	34	19
1	H	417/452 (92%)	403 (97%)	14 (3%)	37	22
All	All	3428/3616 (95%)	3322 (97%)	106 (3%)	42	25

5 of 106 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	419	GLN
1	F	419	GLN
1	H	340	ARG
1	E	462[A]	GLN
1	F	136	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 58 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	462	GLN
1	H	244	GLN
1	E	551	GLN
1	H	136	ASN
1	G	401	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 21 ligands modelled in this entry, 17 are monoatomic - leaving 4 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	ACT	D	604[A]	-	3,3,3	0.91	0	3,3,3	0.72	0
3	ACT	H	601	-	3,3,3	0.87	0	3,3,3	0.54	0
3	ACT	D	603	-	3,3,3	0.80	0	3,3,3	0.98	0
3	ACT	D	604[B]	-	3,3,3	0.87	0	3,3,3	0.50	0

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 (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	539/540 (99%)	-0.05	23 (4%)	35	29	21, 31, 60, 73	0
1	B	539/540 (99%)	-0.11	24 (4%)	33	27	21, 28, 57, 100	0
1	C	539/540 (99%)	-0.18	17 (3%)	47	41	19, 28, 60, 96	0
1	D	539/540 (99%)	-0.29	9 (1%)	70	66	16, 24, 48, 74	0
1	E	539/540 (99%)	-0.28	5 (0%)	84	82	17, 25, 46, 76	0
1	F	539/540 (99%)	-0.21	11 (2%)	65	61	17, 26, 52, 94	0
1	G	540/540 (100%)	-0.15	14 (2%)	56	51	18, 27, 54, 102	0
1	H	539/540 (99%)	-0.11	23 (4%)	35	29	21, 29, 58, 104	0
All	All	4313/4320 (99%)	-0.17	126 (2%)	51	46	16, 27, 55, 104	0

The worst 5 of 126 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	438	ILE	7.2
1	H	441	ALA	6.5
1	G	438	ILE	6.3
1	C	439	TRP	6.2
1	B	443	ALA	6.2

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 (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, 95^{th} 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(Å^2)	Q<0.9
3	ACT	D	604[A]	4/4	0.80	0.28	40,42,43,45	4
3	ACT	D	604[B]	4/4	0.80	0.28	46,47,48,49	4
2	NA	F	600	1/1	0.89	0.10	43,43,43,43	0
3	ACT	D	603	4/4	0.89	0.20	36,37,38,39	0
2	NA	C	600	1/1	0.90	0.16	46,46,46,46	0
2	NA	C	601	1/1	0.91	0.11	50,50,50,50	0
3	ACT	H	601	4/4	0.92	0.20	49,51,52,52	0
2	NA	G	601	1/1	0.93	0.14	34,34,34,34	0
2	NA	G	600	1/1	0.95	0.06	41,41,41,41	0
2	NA	B	600	1/1	0.95	0.12	35,35,35,35	0
2	NA	A	600	1/1	0.95	0.12	37,37,37,37	0
2	NA	F	601	1/1	0.96	0.12	40,40,40,40	0
2	NA	H	600	1/1	0.96	0.07	36,36,36,36	0
2	NA	D	601	1/1	0.97	0.16	36,36,36,36	0
2	NA	E	601	1/1	0.98	0.08	26,26,26,26	0
2	NA	G	602	1/1	0.99	0.10	23,23,23,23	0
2	NA	D	602	1/1	0.99	0.06	27,27,27,27	0
2	NA	E	600	1/1	0.99	0.07	26,26,26,26	0
2	NA	F	602	1/1	0.99	0.06	24,24,24,24	0
2	NA	D	600	1/1	0.99	0.09	25,25,25,25	0
2	NA	E	602	1/1	0.99	0.04	29,29,29,29	0

6.5 Other polymers (i)

There are no such residues in this entry.