wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Sep 24, 2018 – 12:46 PM EDT

PDB ID	:	5A1A
EMDB ID:	:	EMD-2984
Title	:	2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor
Authors	:	Bartesaghi, A.; Merk, A.; Banerjee, S.; Matthies, D.; Wu, X.; Milne, J.; Subramaniam, S.
Deposited on	:	2015-04-29
Resolution	:	2.20 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

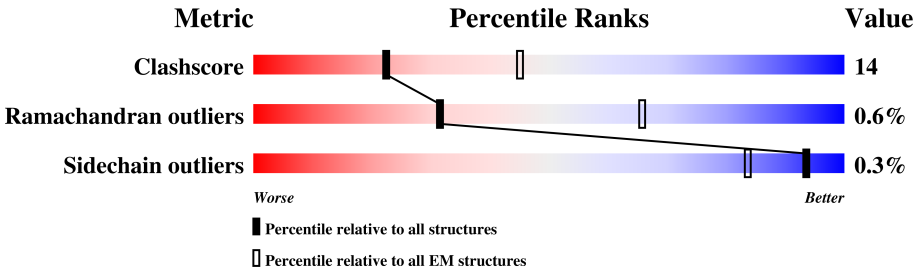
MolProbity	:	4.02b-467
Mogul	:	1.7.3 (157068), CSD as539be (2018)
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031172

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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%

Mol	Chain	Length	Quality of chain
1	A	1022	72% 27%
1	B	1022	71% 28%
1	C	1022	72% 27%
1	D	1022	71% 28%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	PTQ	A	2001	-	-	X	-
2	PTQ	B	2001	-	-	X	-
2	PTQ	C	2001	-	-	X	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	PTQ	D	2001	-	-	X	-

2 Entry composition (i)

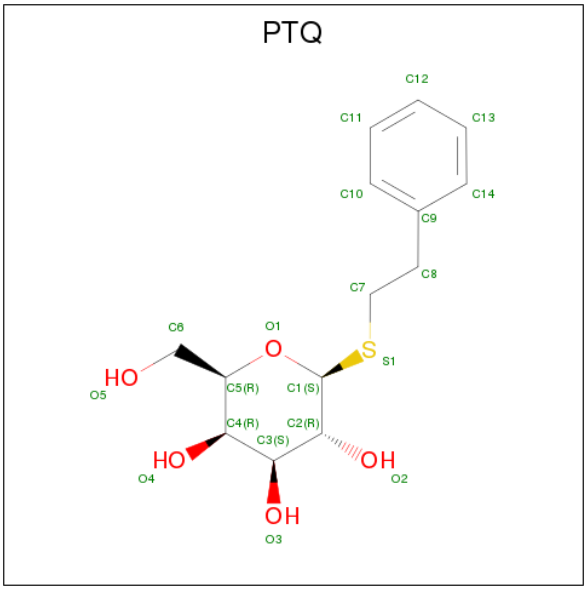
There are 5 unique types of molecules in this entry. The entry contains 33696 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1022	Total	C	N	O	S	0	0
			8206	5190	1452	1525	39		
1	B	1022	Total	C	N	O	S	0	0
			8206	5190	1452	1525	39		
1	C	1022	Total	C	N	O	S	0	0
			8206	5190	1452	1525	39		
1	D	1022	Total	C	N	O	S	0	0
			8206	5190	1452	1525	39		

- Molecule 2 is 2-phenylethyl 1-thio-beta-D-galactopyranoside (three-letter code: PTQ) (formula: $C_{14}H_{20}O_5S$).



Mol	Chain	Residues	Atoms				AltConf
2	A	1	Total	C	O	S	0
			20	14	5	1	
2	B	1	Total	C	O	S	0
			20	14	5	1	

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				AltConf
2	C	1	Total	C	O	S	0
			20	14	5	1	
2	D	1	Total	C	O	S	0
			20	14	5	1	

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
3	B	2	Total	Mg	0
			2	2	
3	A	2	Total	Mg	0
			2	2	
3	D	2	Total	Mg	0
			2	2	
3	C	2	Total	Mg	0
			2	2	

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		AltConf
4	B	2	Total	Na	0
			2	2	
4	A	2	Total	Na	0
			2	2	
4	D	2	Total	Na	0
			2	2	
4	C	2	Total	Na	0
			2	2	

- Molecule 5 is water.

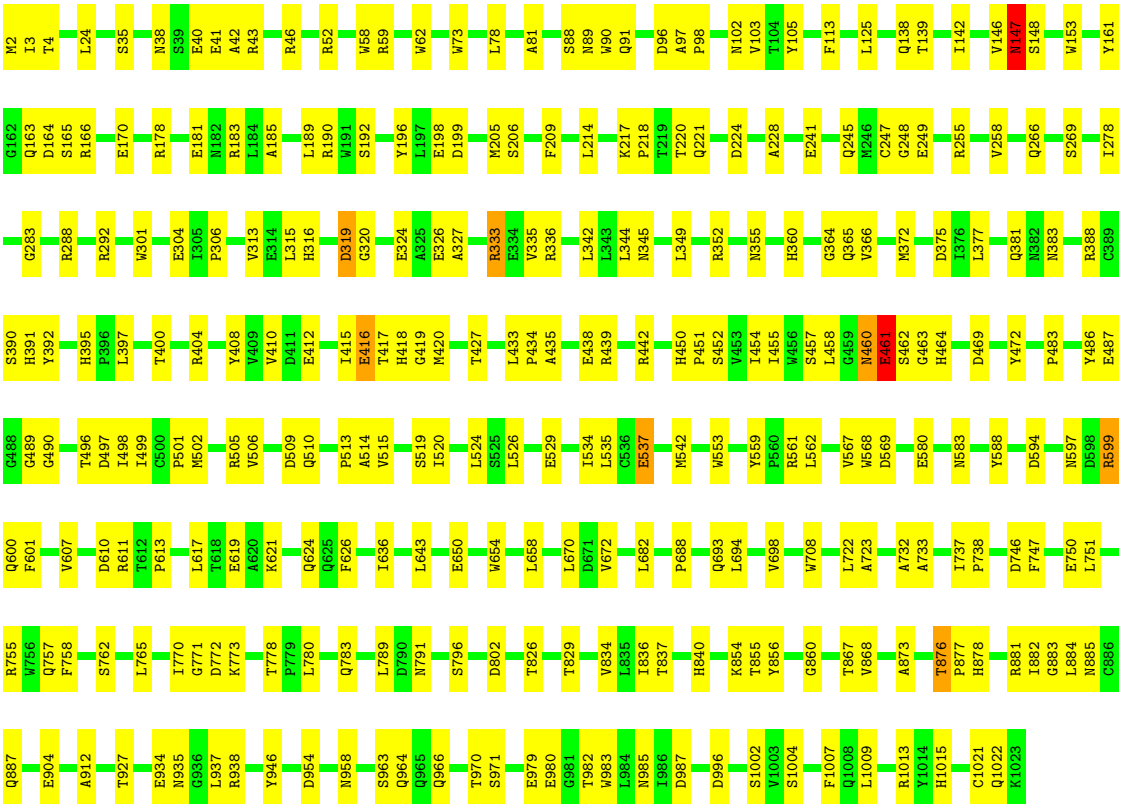
Mol	Chain	Residues	Atoms		AltConf
5	A	194	Total	O	0
			194	194	
5	B	194	Total	O	0
			194	194	
5	C	194	Total	O	0
			194	194	
5	D	194	Total	O	0
			194	194	

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 the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.

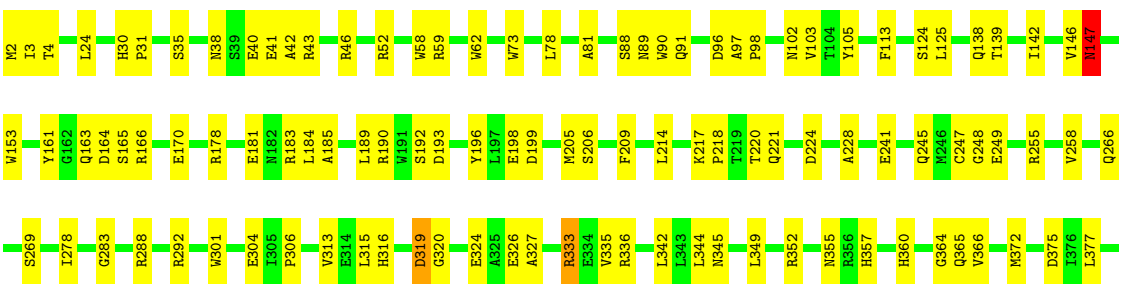
• Molecule 1: BETA-GALACTOSIDASE

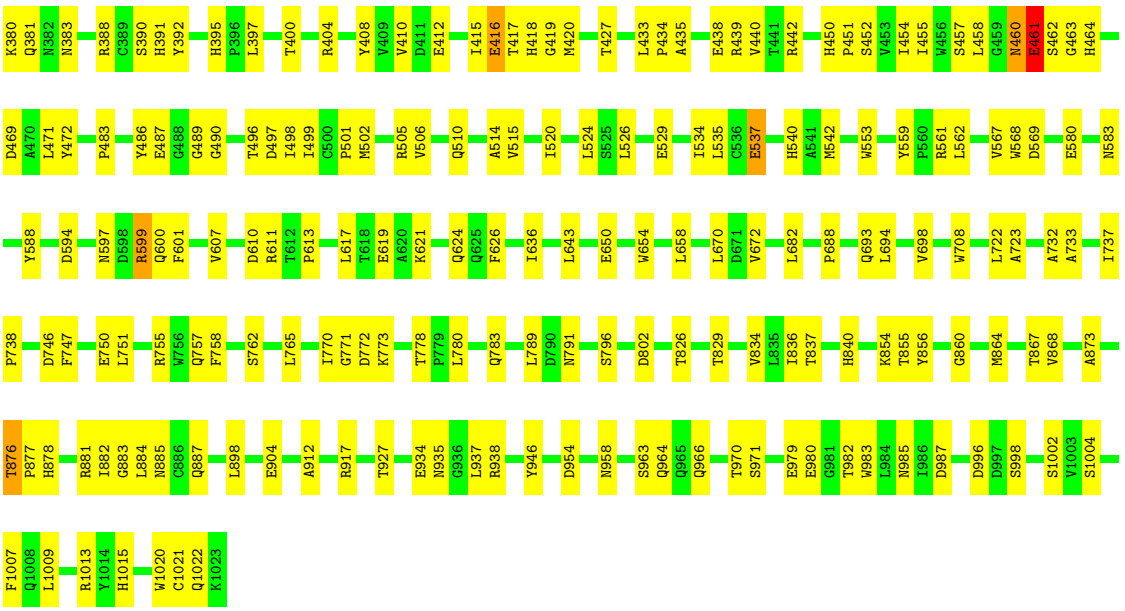
Chain A:



• Molecule 1: BETA-GALACTOSIDASE

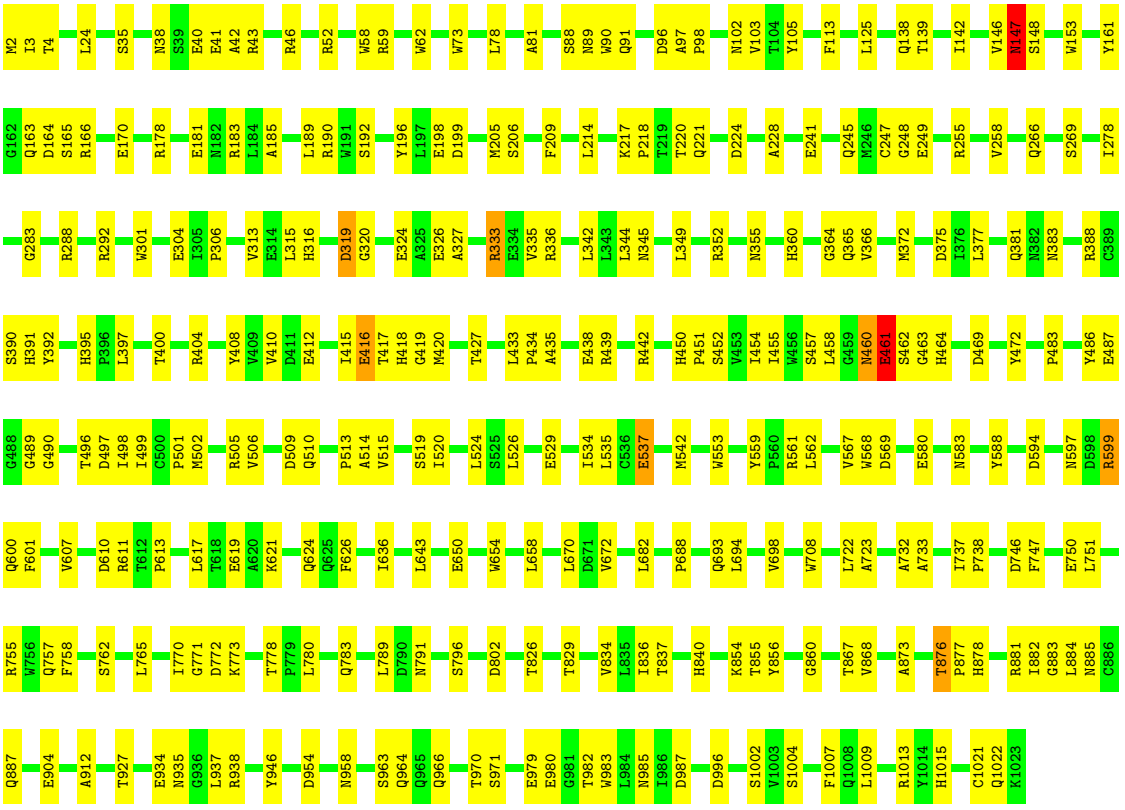
Chain B:





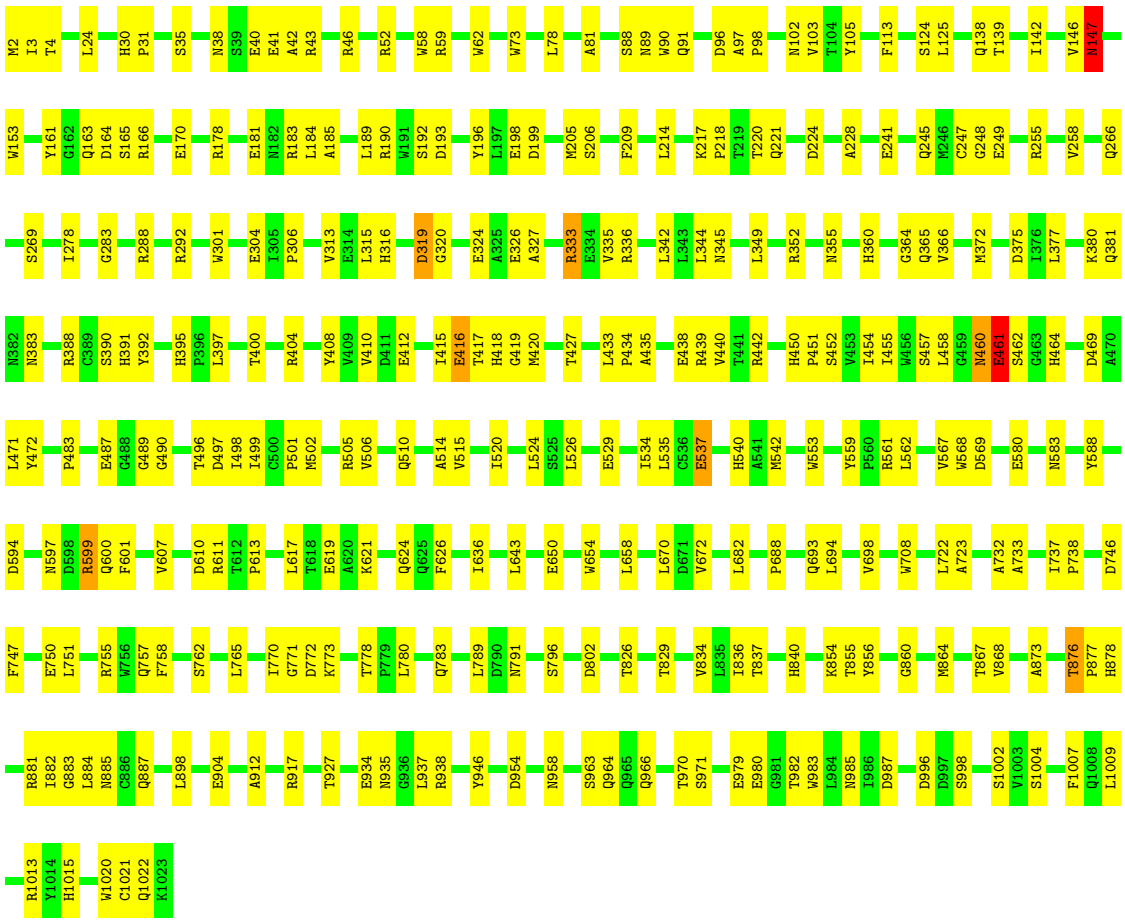
- Molecule 1: BETA-GALACTOSIDASE

Chain C:



- Molecule 1: BETA-GALACTOSIDASE

Chain D:


M2 I3 T4 L24 H30 P31 S35 N38 S39 E40 E41 A42 R43 R46 R52 W58 R59 W62 W73 L78 A81 S88 N89 W90 Q91 D96 A97 P98 N102 V103 T104 Y105 F113 S124 L125 Q138 T139 I142 V146 N147
W153 Y161 G162 Q163 D164 S165 R166 E170 R178 E181 N182 R183 L184 A185 L189 R190 W191 S192 D193 Y196 L197 E198 D199 M205 S206 F209 L214 K217 P218 T219 T220 Q221 D224 A228 E241 Q245 M246 C247 G248 E249 R255 V258 Q266
S269 I278 G283 R288 R292 W301 E304 I305 P306 V313 E314 L315 H316 D319 G320 E324 A325 E326 A327 R333 E334 V335 R336 L342 L343 L344 N345 L349 R352 N355 H360 G364 Q365 V366 M372 D375 I376 L377 K380 Q381
N382 N383 R388 C389 S390 H391 Y392 H395 P396 L397 T400 R404 Y408 V409 V410 D411 E412 I415 E416 T417 H418 G419 M420 T427 L433 P434 A435 E438 R439 V440 T441 R442 H450 P451 S452 V453 I454 I455 W456 S457 L458 G459 N460 E461 S462 G463 H464 D469 A470
L471 Y472 P483 E487 G488 G489 G490 T496 D497 I498 I499 C500 P501 M502 R505 V506 Q510 A514 V515 I520 L524 S525 L526 E529 I534 L535 C536 E537 H540 A541 M542 W553 Y559 P560 R561 L562 V567 W568 D569 E580 N583 Y588
D594 N597 D598 R599 Q600 F601 V607 D610 R611 T612 P613 L617 T618 E619 A620 K621 Q624 Q625 F626 I636 L643 E650 W654 L658 L670 D671 V672 L682 P688 Q693 L694 V698 W708 L722 A723 A732 A733 I737 P738 D746
F747 E750 L751 R755 W756 Q757 F758 S762 L765 I770 G771 D772 K773 T778 P779 L780 Q783 L789 D790 N791 S796 D802 T826 T829 V834 L835 I836 T837 H840 K854 T855 Y856 G860 M864 T867 V868 A873 T876 P877 H878
R881 I882 G883 L884 N885 C886 Q887 L898 E904 A912 R917 T927 E934 N935 G936 L937 R938 Y946 D954 N958 S963 Q964 Q965 Q966 T970 S971 E979 E980 G981 T982 W983 L984 N985 I986 D987 D996 D997 S998 S1002 V1003 S1004 F1007 Q1008 L1009
R1013 Y1014 H1015 W1020 C1021 Q1022 K1023

4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, D2	Depositor
Number of particles used	41123	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	EACH PARTICLE	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	45	Depositor
Minimum defocus (nm)	600	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	215000	Depositor
Image detector	GATAN K2 (4k x 4k)	Depositor

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: NA, MG, PTQ

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	#$\|Z\|$ >2	RMSZ	#$\|Z\|$ >2
1	A	0.40	6/8448 (0.1%)	0.52	14/11526 (0.1%)
1	B	0.40	6/8448 (0.1%)	0.52	14/11526 (0.1%)
1	C	0.40	6/8448 (0.1%)	0.52	14/11526 (0.1%)
1	D	0.40	6/8448 (0.1%)	0.52	14/11526 (0.1%)
All	All	0.40	24/33792 (0.1%)	0.52	56/46104 (0.1%)

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	3
1	B	0	3
1	C	0	3
1	D	0	3
All	All	0	12

The worst 5 of 24 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	146	VAL	C-N	-18.42	0.91	1.34
1	B	146	VAL	C-N	-18.42	0.91	1.34
1	C	146	VAL	C-N	-18.42	0.91	1.34
1	D	146	VAL	C-N	-18.42	0.91	1.34
1	A	461	GLU	C-N	-15.25	0.98	1.34

The worst 5 of 56 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	461	GLU	O-C-N	19.17	153.38	122.70
1	B	461	GLU	O-C-N	19.17	153.38	122.70
1	C	461	GLU	O-C-N	19.17	153.38	122.70
1	D	461	GLU	O-C-N	19.17	153.38	122.70
1	A	147	ASN	O-C-N	-14.44	99.59	122.70

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	147	ASN	Mainchain
1	A	460	ASN	Mainchain
1	A	882	ILE	Peptide
1	B	147	ASN	Mainchain
1	B	460	ASN	Mainchain

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	8206	0	7802	218	0
1	B	8206	0	7802	221	0
1	C	8206	0	7802	218	0
1	D	8206	0	7802	218	0
2	A	20	0	18	20	0
2	B	20	0	18	20	0
2	C	20	0	18	20	0
2	D	20	0	18	19	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
3	C	2	0	0	0	0
3	D	2	0	0	0	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
4	C	2	0	0	0	0
4	D	2	0	0	0	0
5	A	194	0	0	19	0
5	B	194	0	0	19	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	194	0	0	19	0
5	D	194	0	0	19	0
All	All	33696	0	31280	878	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 14.

The worst 5 of 878 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:601:PHE:CG	2:C:2001:PTQ:H11	1.82	1.15
1:A:601:PHE:CG	2:A:2001:PTQ:H11	1.82	1.15
1:D:601:PHE:CG	2:D:2001:PTQ:H11	1.82	1.14
1:B:601:PHE:CG	2:B:2001:PTQ:H11	1.82	1.14
1:B:601:PHE:CD1	2:B:2001:PTQ:H11	1.97	1.00

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 PDB entries followed by that with respect to all EM entries.

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	1020/1022 (100%)	974 (96%)	40 (4%)	6 (1%)	27	28
1	B	1020/1022 (100%)	974 (96%)	40 (4%)	6 (1%)	27	28
1	C	1020/1022 (100%)	974 (96%)	40 (4%)	6 (1%)	27	28
1	D	1020/1022 (100%)	974 (96%)	40 (4%)	6 (1%)	27	28
All	All	4080/4088 (100%)	3896 (96%)	160 (4%)	24 (1%)	31	28

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	3	ILE
1	A	461	GLU
1	B	3	ILE
1	B	461	GLU
1	C	3	ILE

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 PDB entries followed by that with respect to all EM entries.

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	874/874 (100%)	871 (100%)	3 (0%)	93 97
1	B	874/874 (100%)	871 (100%)	3 (0%)	93 97
1	C	874/874 (100%)	871 (100%)	3 (0%)	93 97
1	D	874/874 (100%)	871 (100%)	3 (0%)	93 97
All	All	3496/3496 (100%)	3484 (100%)	12 (0%)	93 97

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	537	GLU
1	C	333	ARG
1	D	333	ARG
1	B	416	GLU
1	C	537	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 76 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	775	GLN
1	C	163	GLN
1	D	653	HIS
1	B	935	ASN
1	C	25	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 20 ligands modelled in this entry, 16 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	$\#\|Z\| > 2$
2	PTQ	A	2001	4	21,21,21	2.55	4 (19%)	27,28,28	1.40	2 (7%)
2	PTQ	B	2001	4	21,21,21	2.55	4 (19%)	27,28,28	1.40	2 (7%)
2	PTQ	C	2001	4	21,21,21	2.55	4 (19%)	27,28,28	1.40	2 (7%)
2	PTQ	D	2001	4	21,21,21	2.55	4 (19%)	27,28,28	1.40	2 (7%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	PTQ	A	2001	4	-	0/8/28/28	0/2/2/2
2	PTQ	B	2001	4	-	0/8/28/28	0/2/2/2
2	PTQ	C	2001	4	-	0/8/28/28	0/2/2/2
2	PTQ	D	2001	4	-	0/8/28/28	0/2/2/2

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	2001	PTQ	C7-S1	-2.83	1.77	1.81
2	C	2001	PTQ	C7-S1	-2.83	1.77	1.81
2	B	2001	PTQ	C7-S1	-2.83	1.77	1.81
2	D	2001	PTQ	C7-S1	-2.83	1.77	1.81
2	A	2001	PTQ	C12-C11	5.68	1.51	1.38

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	2001	PTQ	C7-C8-C9	-3.27	108.41	114.05
2	C	2001	PTQ	C7-C8-C9	-3.27	108.41	114.05
2	B	2001	PTQ	C7-C8-C9	-3.27	108.41	114.05
2	D	2001	PTQ	C7-C8-C9	-3.27	108.41	114.05
2	A	2001	PTQ	C7-S1-C1	5.64	109.54	100.13

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 79 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	2001	PTQ	20	0
2	B	2001	PTQ	20	0
2	C	2001	PTQ	20	0
2	D	2001	PTQ	19	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	3
1	A	3
1	D	3

Continued on next page...

Continued from previous page...

Mol	Chain	Number of breaks
1	C	3

The worst 5 of 12 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	139:THR	C	140:ARG	N	1.04
1	B	139:THR	C	140:ARG	N	1.04
1	C	139:THR	C	140:ARG	N	1.04
1	D	139:THR	C	140:ARG	N	1.04
1	A	461:GLU	C	462:SER	N	0.99