wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 01:51 PM BST

PDB ID	:	3J7H
EMDB ID:	:	EMD-5995
Title	:	Structure of beta-galactosidase at 3.2-A resolution obtained by cryo-electron microscopy
Authors	:	Bartesaghi, A.; Matthies, D.; Banerjee, S.; Merk, A.; Subramaniam, S.
Deposited on	:	2014-06-30
Resolution	:	3.20 Å(reported)

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

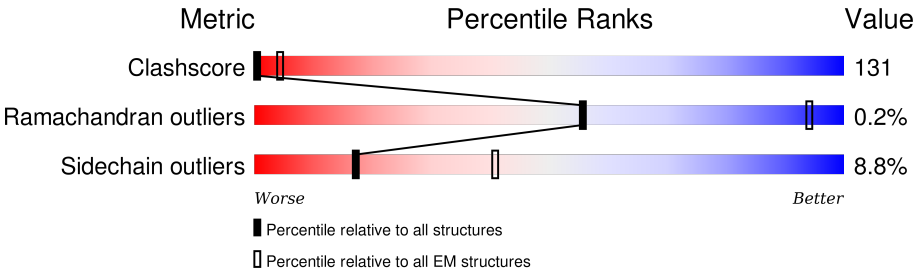
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

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 3.20 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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	1024	22% 71% 7%
1	B	1024	22% 71% 7%
1	C	1024	22% 71% 7%
1	D	1024	22% 71% 7%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 32828 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 MAGNESIUM ION (three-letter code: MG) (formula: Mg).

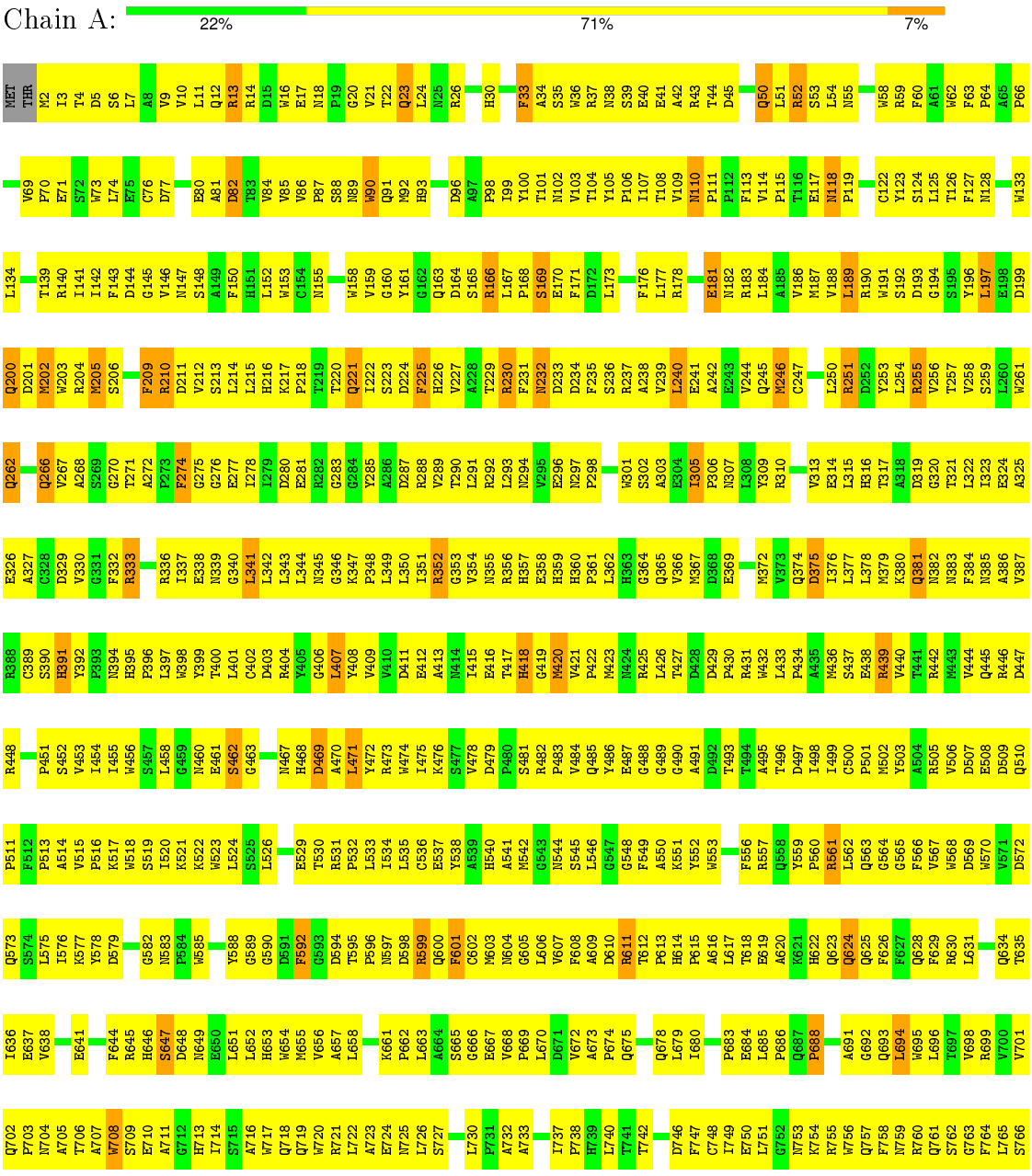
Mol	Chain	Residues	Atoms		AltConf
2	B	1	Total	Mg	0
			1	1	
2	A	1	Total	Mg	0
			1	1	
2	D	1	Total	Mg	0
			1	1	
2	C	1	Total	Mg	0
			1	1	

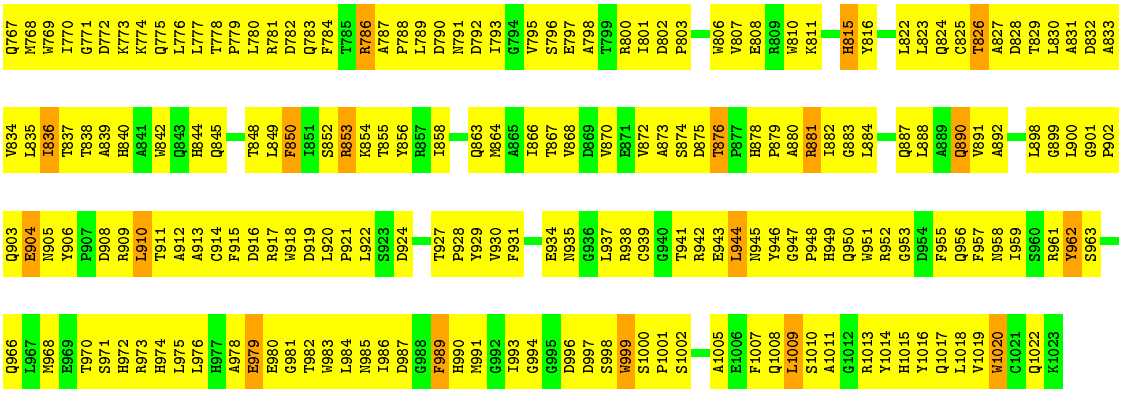
3 Residue-property plots (i)

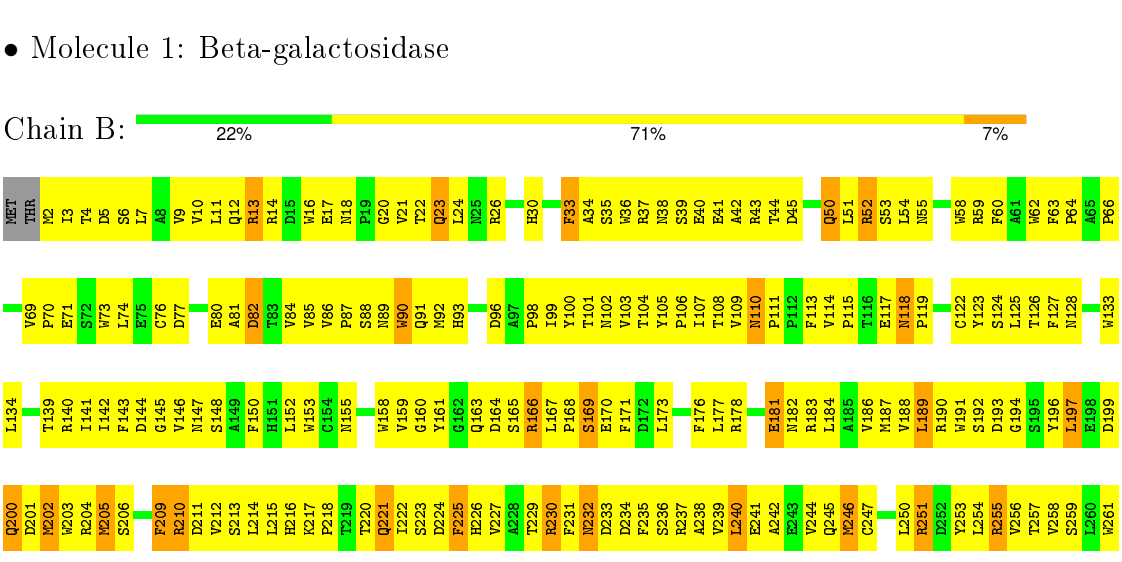
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric 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:

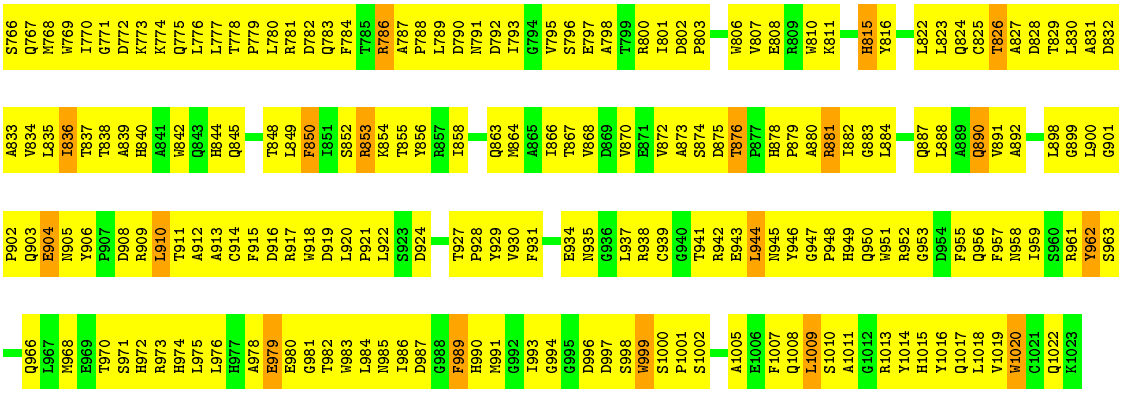


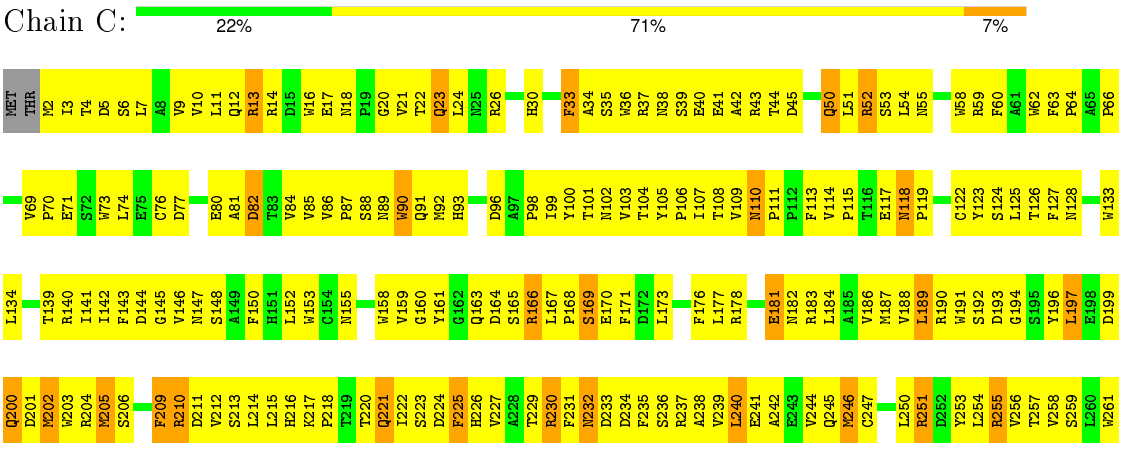




• Molecule 1: Beta-galactosidase

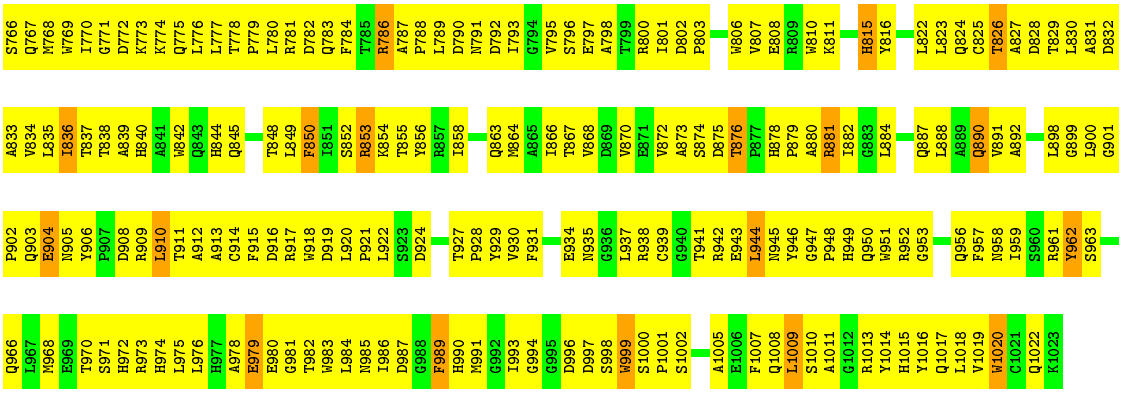
Chain B:





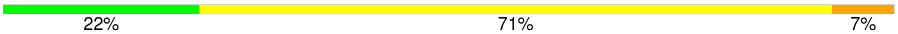
- Molecule 1: Beta-galactosidase

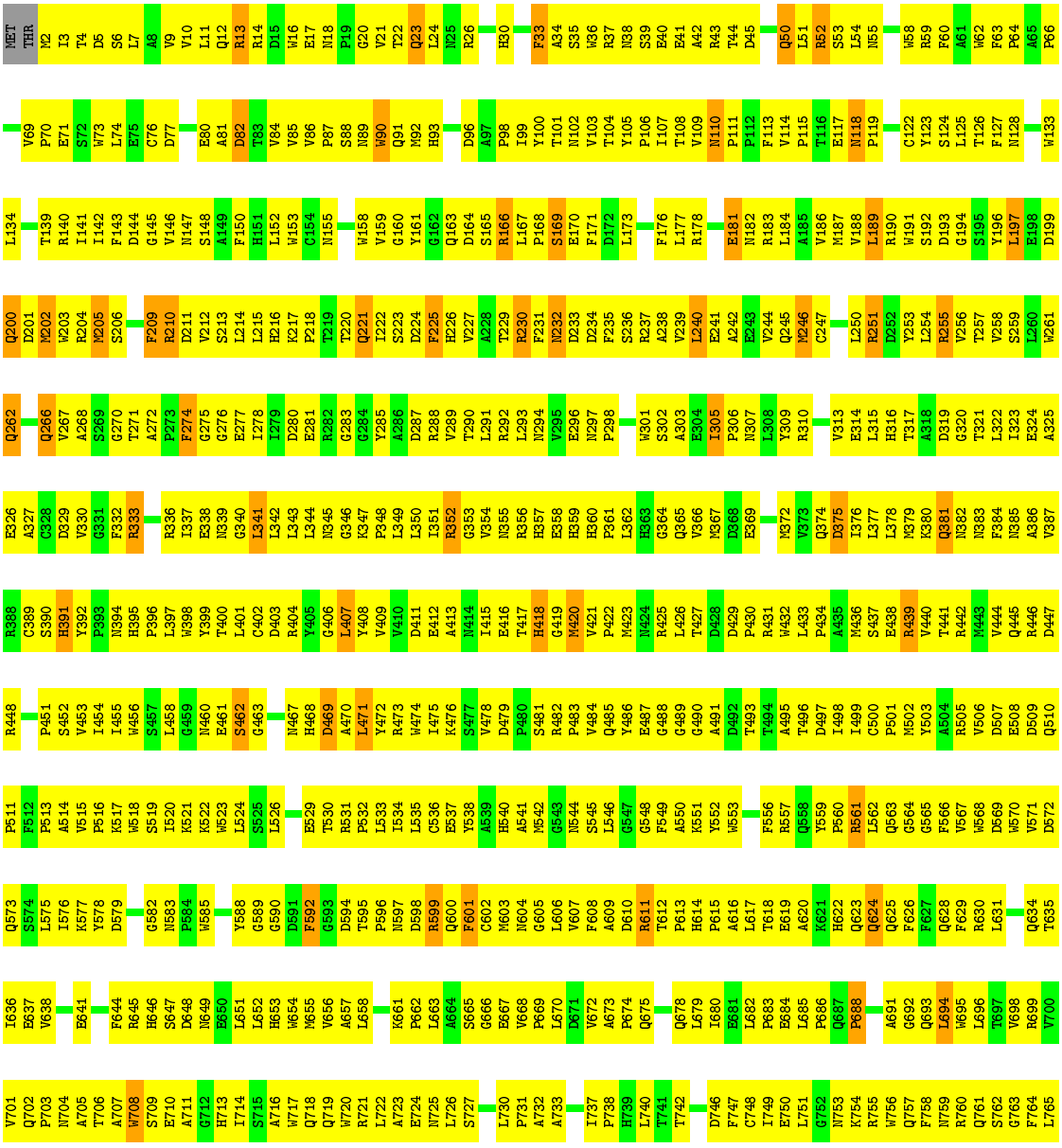
Chain C:

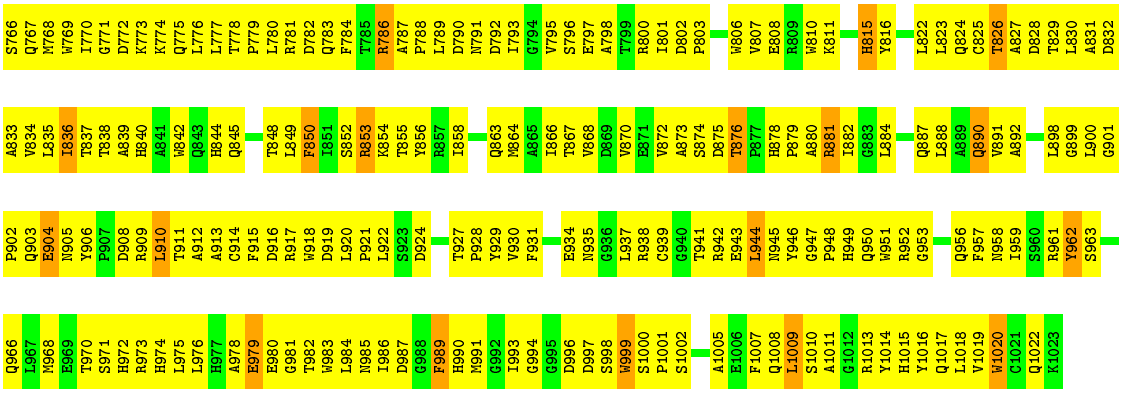


- Molecule 1: Beta-galactosidase

Chain D:






S766 Q767 M768 W769 I770 G771 D772 K773 K774 Q775 L776 L777 T778 P779 L780 R781 D782 Q783 F784 T785 R786 A787 P788 L789 D790 N791 D792 I793 G794 V795 S796 E797 A798 T799 R800 I801 D802 P803 W806 V807 E808 R809 W810 K811 H815 Y816 L822 L823 Q824 C825 T826 A827 D828 T829 L830 A831 D832
A833 V834 L835 I836 T837 T838 A839 H840 A841 W842 Q843 H844 Q845 T848 L849 F850 I851 S852 R853 K854 T855 Y856 R857 I858 Q863 M864 A865 I866 T867 V868 D869 V870 E871 V872 A873 S874 D875 T876 P877 H878 P879 A880 R881 I882 G883 L884 Q887 L888 A889 Q890 V891 A892 L898 G899 L900 G901
P902 Q903 E904 N905 Y906 P907 D908 R909 L910 T911 A912 A913 C914 F915 D916 R917 W918 D919 L920 P921 L922 S923 D924 T927 P928 Y929 V930 F931 E934 N935 G936 L937 R938 C939 G940 T941 R942 E943 L944 N945 Y946 G947 P948 H949 Q950 W951 R952 G953 Q956 F957 N958 I959 S960 R961 Y962 S963
Q966 L967 M968 E969 T970 S971 H972 R973 H974 L975 L976 H977 A978 E979 E980 G981 T982 W983 L984 N985 I986 D987 G988 F989 H990 M991 G992 I993 G994 G995 D996 D997 S998 W999 S1000 P1001 S1002 A1005 E1006 F1007 Q1008 L1009 S1010 A1011 G1012 R1013 Y1014 H1015 Y1016 Q1017 L1018 V1019 W1020 C1021 Q1022 K1023

4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	11726	Depositor
Resolution determination method	FSC 0.143	Depositor
CTF correction method	Individual frames of each movie were aligned by cross-correlation using the cumulative average of previously aligned frames as a reference to align the remaining frames. Parameters of the contrast transfer function for each micrograph were estimated from power spectra obtained using periodogram averaging with tiles of size 512x512 pixels extracted from all frames of each movie. These power spectra were then radially averaged and used to estimate the defocus for each image using frequencies in the 15.0-3.0 Angstrom range. CTF correction was done for each particle as implemented in FREALIGN's reconstruction protocol.	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	45	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2500	Depositor
Magnification	105000	Depositor
Image detector	GATAN K2 (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

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

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	0.26	0/8448	0.47	2/11526 (0.0%)
1	B	0.26	1/8448 (0.0%)	0.47	2/11526 (0.0%)
1	C	0.26	0/8448	0.47	2/11526 (0.0%)
1	D	0.26	0/8448	0.47	2/11526 (0.0%)
All	All	0.26	1/33792 (0.0%)	0.47	8/46104 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	306	PRO	N-CD	5.03	1.54	1.47

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	305	ILE	C-N-CD	5.72	140.41	128.40
1	C	305	ILE	C-N-CD	5.72	140.41	128.40
1	D	305	ILE	C-N-CD	5.71	140.39	128.40
1	B	305	ILE	C-N-CD	5.71	140.39	128.40
1	B	110	ASN	C-N-CD	5.11	139.14	128.40

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	8206	0	7791	2164	0
1	B	8206	0	7791	2165	0
1	C	8206	0	7791	2156	0
1	D	8206	0	7791	2161	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
All	All	32828	0	31164	8379	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 131.

The worst 5 of 8379 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:427:THR:HA	1:C:436:MET:CE	1.21	1.67
1:D:427:THR:HA	1:D:436:MET:CE	1.21	1.67
1:C:159:VAL:HG22	1:C:176:PHE:CE1	1.25	1.64
1:D:159:VAL:HG22	1:D:176:PHE:CE1	1.25	1.64
1:A:159:VAL:HG22	1:A:176:PHE:CE1	1.25	1.63

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	1020/1024 (100%)	993 (97%)	25 (2%)	2 (0%)	52 88
1	B	1020/1024 (100%)	993 (97%)	25 (2%)	2 (0%)	52 88
1	C	1020/1024 (100%)	993 (97%)	25 (2%)	2 (0%)	52 88
1	D	1020/1024 (100%)	993 (97%)	25 (2%)	2 (0%)	52 88

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	4080/4096 (100%)	3972 (97%)	100 (2%)	8 (0%)	56 88

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	511	PRO
1	B	511	PRO
1	C	511	PRO
1	D	511	PRO
1	A	688	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 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	874/876 (100%)	797 (91%)	77 (9%)	12 45
1	B	874/876 (100%)	797 (91%)	77 (9%)	12 45
1	C	874/876 (100%)	797 (91%)	77 (9%)	12 45
1	D	874/876 (100%)	797 (91%)	77 (9%)	12 45
All	All	3496/3504 (100%)	3188 (91%)	308 (9%)	17 45

5 of 308 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	836	ILE
1	C	221	GLN
1	D	624	GLN
1	B	881	ARG
1	C	33	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 160 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	843	GLN
1	C	266	GLN
1	D	713	HIS
1	B	890	GLN
1	C	23	GLN

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 4 ligands modelled in this entry, 4 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

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.