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

Dec 12, 2018 – 02:17 PM EST

PDB ID	:	6GYP
EMDB ID:	:	EMD-0095
Title	:	Cryo-EM structure of the CBF3-core-Ndc10-DBD complex of the budding yeast kinetochore
Authors	:	Yan, K.; Zhang, Z.; Yang, J.; McLaughlin, S.H.; Barford, D.
Deposited on	:	2018-07-01
Resolution	:	3.60 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.

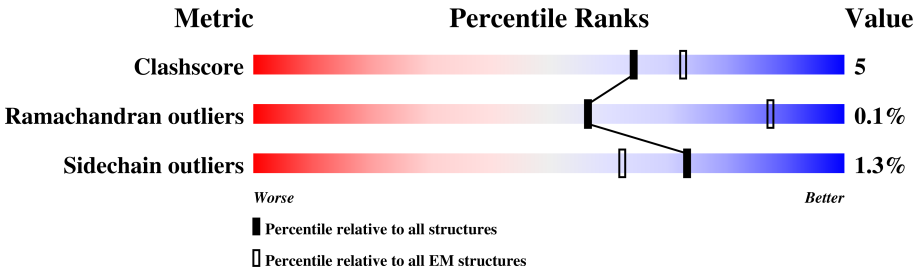
MolProbity	:	4.02b-467
Mogul	:	1.7.3 (157068), CSD as539be (2018)
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031633

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 3.60 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	B	608	76% 11% 13%
2	A	478	70% 18% • 12%
3	C	564	84% 9% 6%
4	D	194	64% 14% 22%
5	E	956	46% 8% 46%

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 17811 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 Centromere DNA-binding protein complex CBF3 subunit B.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	B	527	4374	2844	704	805	21	0	0

- Molecule 2 is a protein called Centromere DNA-binding protein complex CBF3 subunit C.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	A	423	3526	2299	594	621	12	0	0

- Molecule 3 is a protein called Centromere DNA-binding protein complex CBF3 subunit B.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	528	4362	2839	702	800	21	0	1

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MET	-	initiating methionine	UNP P40969
C	2	PHE	-	expression tag	UNP P40969
C	3	ASN	-	expression tag	UNP P40969
C	4	ARG	-	expression tag	UNP P40969

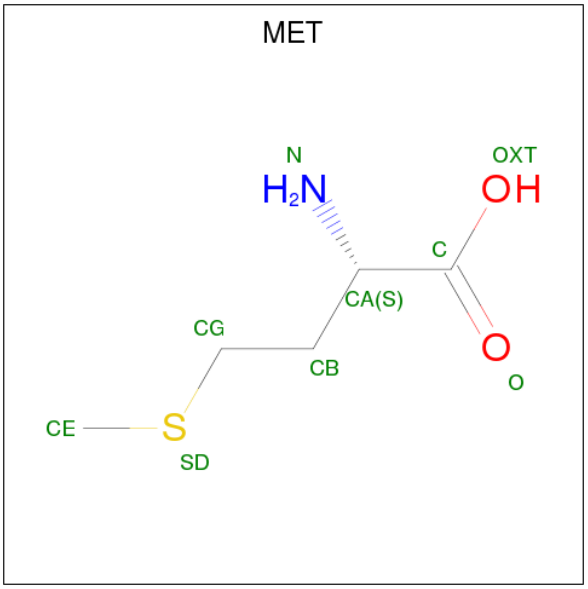
- Molecule 4 is a protein called Suppressor of kinetochore protein 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	151	1226	767	214	241	4	0	0

- Molecule 5 is a protein called Centromere DNA-binding protein complex CBF3 subunit A.

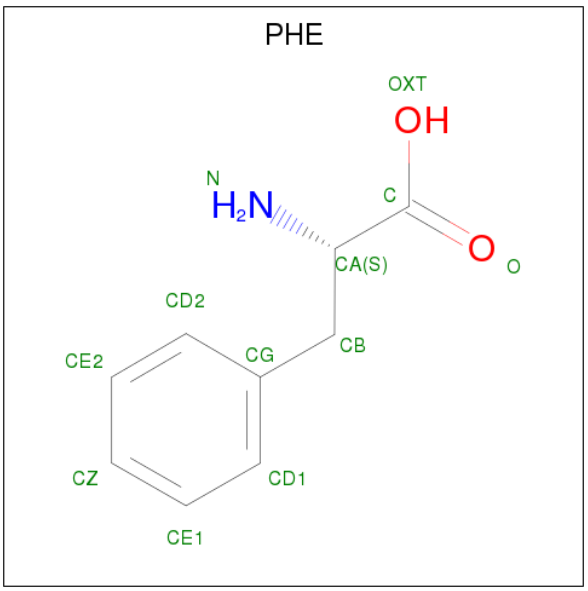
Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	512	Total	C	N	O	S	0	0
			4265	2755	706	794	10		

- Molecule 6 is METHIONINE (three-letter code: MET) (formula: $C_5H_{11}NO_2S$).



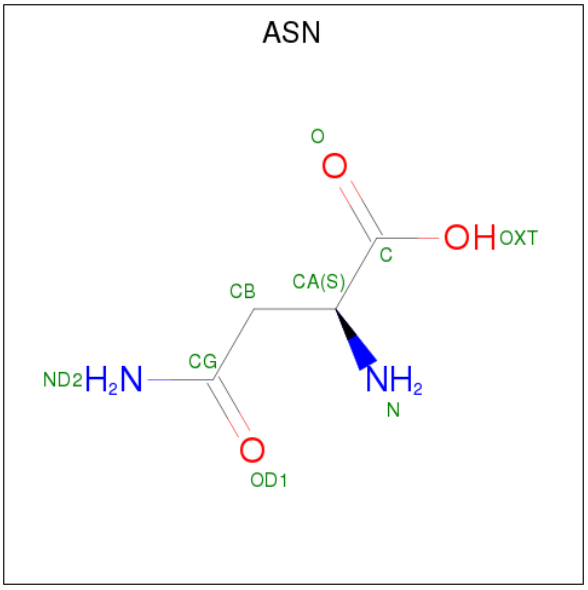
Mol	Chain	Residues	Atoms				AltConf
6	B	1	Total	C	N	O	0
			5	3	1	1	

- Molecule 7 is PHENYLALANINE (three-letter code: PHE) (formula: $C_9H_{11}NO_2$).



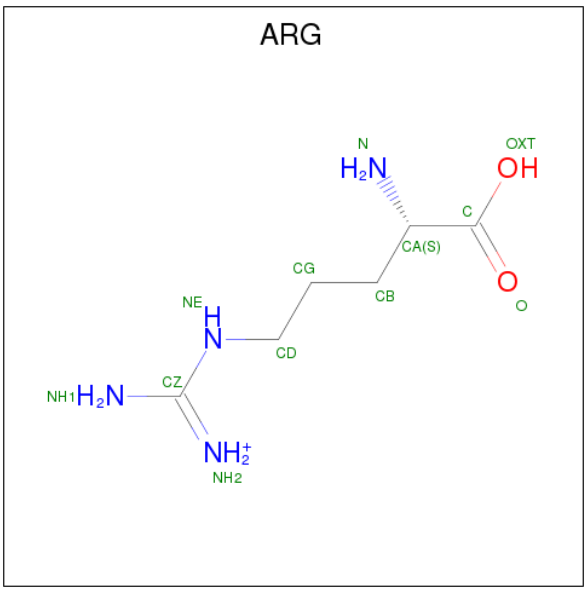
Mol	Chain	Residues	Atoms				AltConf
7	B	1	Total	C	N	O	0
			11	9	1	1	

- Molecule 8 is ASPARAGINE (three-letter code: ASN) (formula: $C_4H_8N_2O_3$).



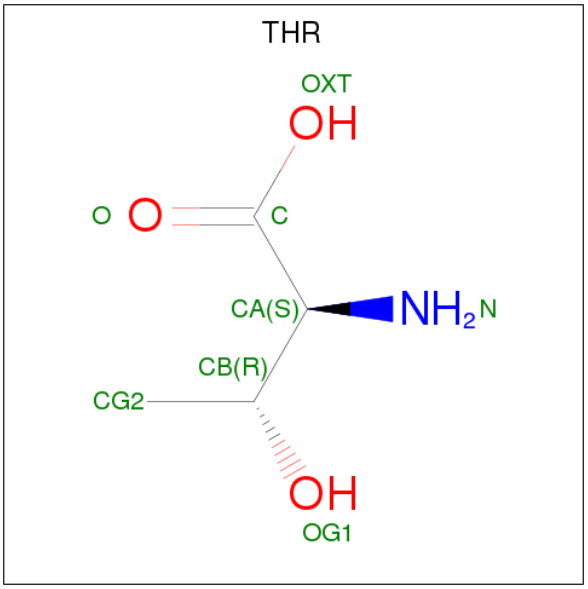
Mol	Chain	Residues	Atoms				AltConf
8	B	1	Total	C	N	O	0
			8	4	2	2	

- Molecule 9 is ARGININE (three-letter code: ARG) (formula: $C_6H_{15}N_4O_2$).



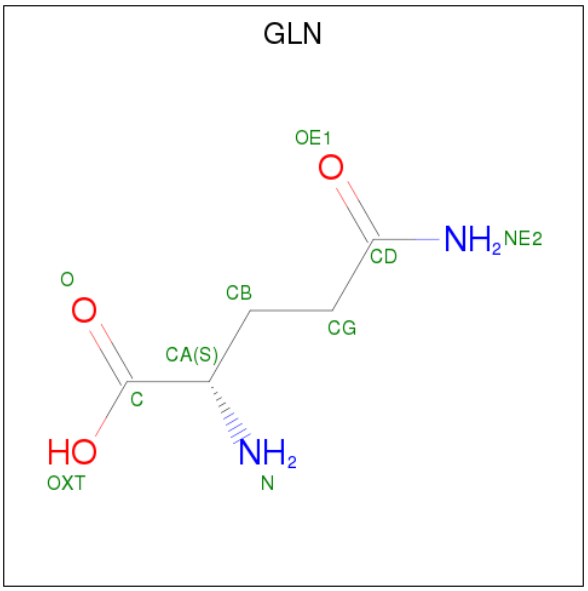
Mol	Chain	Residues	Atoms				AltConf
9	B	1	Total	C	N	O	0
			11	6	4	1	

- Molecule 10 is THREONINE (three-letter code: THR) (formula: $C_4H_9NO_3$).



Mol	Chain	Residues	Atoms				AltConf
10	B	1	Total	C	N	O	0
			14	8	2	4	
10	B	1	Total	C	N	O	0
			14	8	2	4	

- Molecule 11 is GLUTAMINE (three-letter code: GLN) (formula: $C_5H_{10}N_2O_3$).



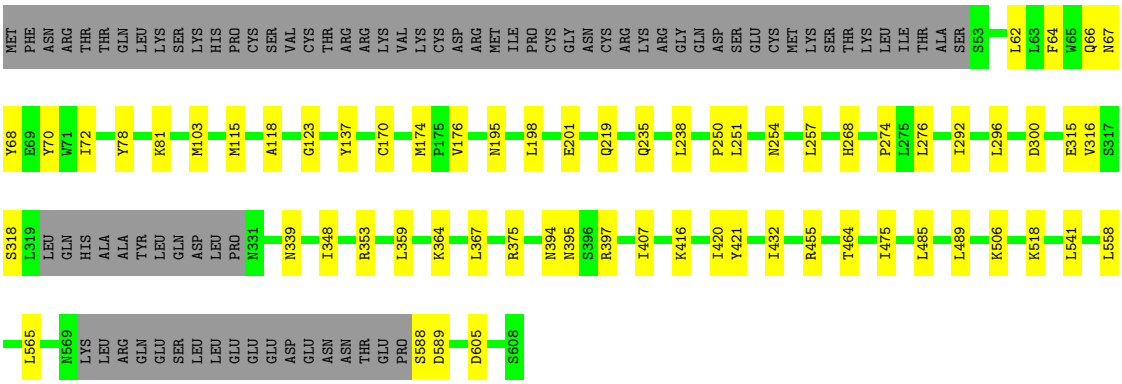
Mol	Chain	Residues	Atoms				AltConf
11	B	1	Total	C	N	O	0
			9	5	2	2	

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.

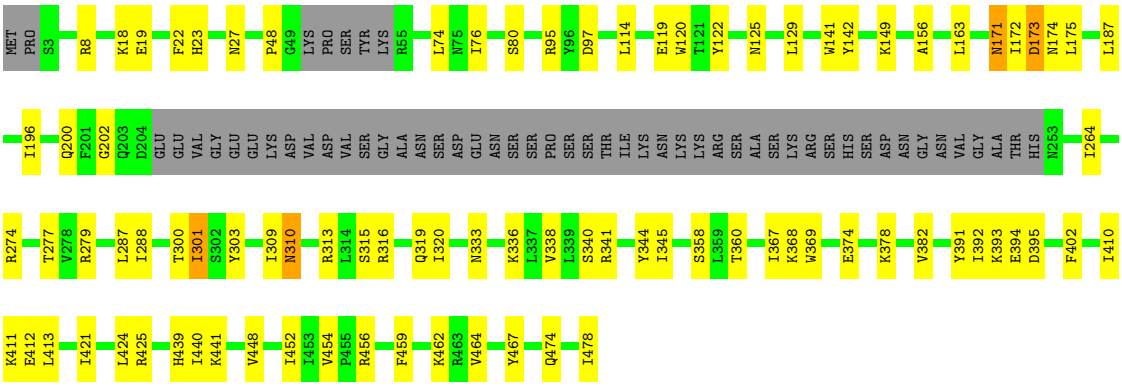
• Molecule 1: Centromere DNA-binding protein complex CBF3 subunit B

Chain B:



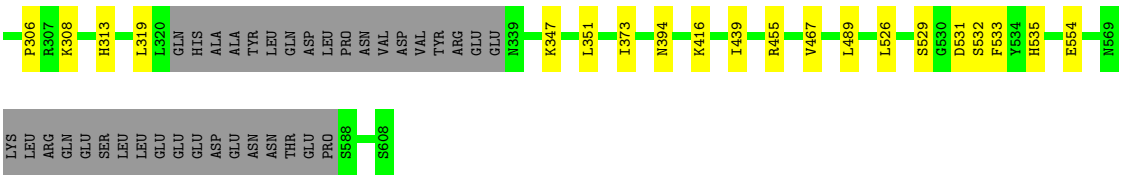
• Molecule 2: Centromere DNA-binding protein complex CBF3 subunit C

Chain A:



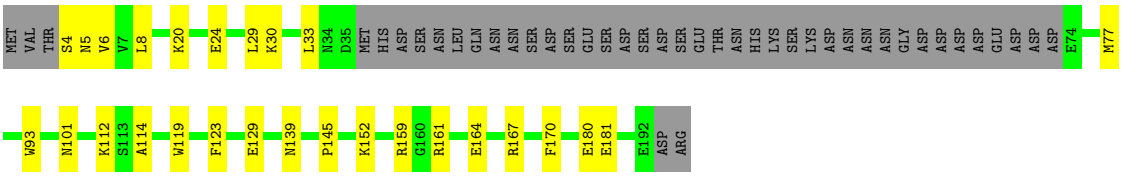
• Molecule 3: Centromere DNA-binding protein complex CBF3 subunit B

Chain C:



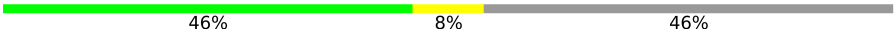
• Molecule 4: Suppressor of kinetochore protein 1

Chain D:



• Molecule 5: Centromere DNA-binding protein complex CBF3 subunit A

Chain E:



4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	73894	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	27	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	FEI FALCON III (4k x 4k)	Depositor

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	#$\|Z\|>2$	RMSZ	#$\|Z\|>2$
1	B	0.36	0/4480	0.54	0/6068
2	A	0.37	0/3604	0.67	3/4869 (0.1%)
3	C	0.31	0/4467	0.49	0/6050
4	D	0.33	0/1247	0.52	0/1687
5	E	0.27	0/4391	0.49	1/5960 (0.0%)
All	All	0.33	0/18189	0.55	4/24634 (0.0%)

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
2	A	0	4

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	173	ASP	CB-CG-OD1	8.46	125.91	118.30
2	A	74	LEU	CA-CB-CG	5.80	128.64	115.30
2	A	172	ILE	C-N-CA	5.56	135.59	121.70
5	E	90	LEU	CA-CB-CG	5.04	126.89	115.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	A	300	THR	Peptide
2	A	301	ILE	Peptide
2	A	368	LYS	Peptide
2	A	421	ILE	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	B	4374	0	4334	36	0
2	A	3526	0	3558	54	0
3	C	4362	0	4332	29	0
4	D	1226	0	1194	18	0
5	E	4265	0	4150	38	0
6	B	5	0	1	0	0
7	B	11	0	9	0	0
8	B	8	0	6	0	0
9	B	11	0	13	0	0
10	B	14	0	14	0	0
11	B	9	0	8	0	0
All	All	17811	0	17619	162	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 162 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)
3:C:57:TYR:O	3:C:61:LEU:HB2	1.92	0.70
2:A:287:LEU:HD12	2:A:301:ILE:HD12	1.75	0.69
2:A:18:LYS:HG2	2:A:114:LEU:HD11	1.80	0.64
5:E:297:ILE:HB	5:E:301:SER:H	1.61	0.63
1:B:315:GLU:O	1:B:353:ARG:NH2	2.31	0.63

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	B	521/608 (86%)	496 (95%)	25 (5%)	0	100 100
2	A	417/478 (87%)	358 (86%)	57 (14%)	2 (0%)	31 72
3	C	520/564 (92%)	503 (97%)	17 (3%)	0	100 100
4	D	147/194 (76%)	140 (95%)	7 (5%)	0	100 100
5	E	510/956 (53%)	467 (92%)	43 (8%)	0	100 100
All	All	2115/2800 (76%)	1964 (93%)	149 (7%)	2 (0%)	56 86

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	369	TRP
2	A	303	TYR

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	B	492/569 (86%)	488 (99%)	4 (1%)	83 93
2	A	388/449 (86%)	382 (98%)	6 (2%)	67 87
3	C	490/527 (93%)	483 (99%)	7 (1%)	69 88
4	D	136/179 (76%)	134 (98%)	2 (2%)	67 87
5	E	481/897 (54%)	475 (99%)	6 (1%)	74 89
All	All	1987/2621 (76%)	1962 (99%)	25 (1%)	73 89

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	146	ARG
3	C	199	ASN
5	E	429	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	C	158	TYR
3	C	232	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	82	GLN
3	C	394	ASN
5	E	267	HIS
3	C	67	ASN
5	E	379	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

7 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
6	MET	B	701	-	4,4,8	0.99	0	1,4,9	0.25	0
7	PHE	B	702	-	11,11,12	0.73	0	12,13,15	0.52	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
8	ASN	B	703	-	7,7,8	0.83	0	7,8,10	0.60	0
9	ARG	B	704	-	10,10,11	0.80	1 (10%)	7,11,13	0.69	0
10	THR	B	705	-	6,6,7	1.29	1 (16%)	6,7,9	0.80	0
10	THR	B	706	-	6,6,7	1.02	1 (16%)	6,7,9	0.70	0
11	GLN	B	707	-	8,8,9	0.81	1 (12%)	6,9,11	0.72	0

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
6	MET	B	701	-	-	0/0/2/8	0/0/0/0
7	PHE	B	702	-	-	0/4/6/8	0/1/1/1
8	ASN	B	703	-	-	0/4/6/8	0/0/0/0
9	ARG	B	704	-	-	0/7/9/11	0/0/0/0
10	THR	B	705	-	-	0/4/6/8	0/0/0/0
10	THR	B	706	-	-	0/4/6/8	0/0/0/0
11	GLN	B	707	-	-	0/5/7/9	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	B	707	GLN	CA-C	2.01	1.52	1.50
9	B	704	ARG	CA-C	2.22	1.53	1.50
10	B	706	THR	CA-C	2.23	1.53	1.50
10	B	705	THR	CA-C	2.96	1.54	1.50

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.