wwPDB X-ray Structure Validation Summary Report (i)

Jan 16, 2024 – 01:51 am GMT

PDB ID	:	6TU5
Title	:	Influenza A/H7N9 polymerase core (apo)
Authors	:	Cusack, S.; Pflug, A.
Deposited on	:	2020-01-02
Resolution	:	3.33 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:

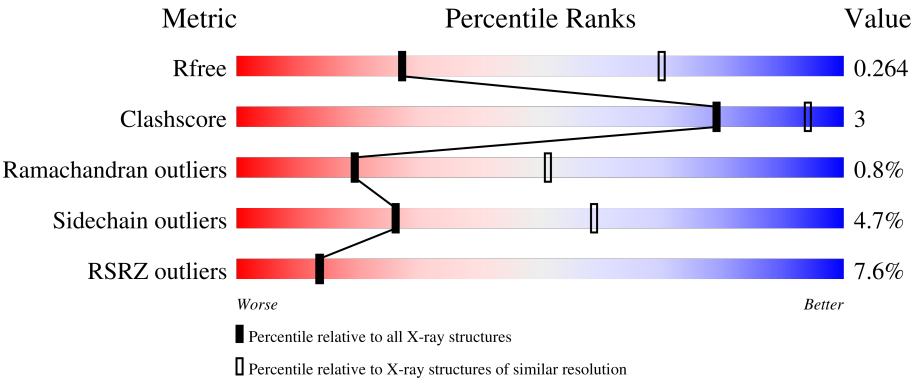
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.33 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.



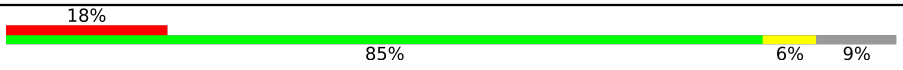
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1089 (3.36-3.28)
Clashscore	141614	1137 (3.36-3.28)
Ramachandran outliers	138981	1115 (3.36-3.28)
Sidechain outliers	138945	1114 (3.36-3.28)
RSRZ outliers	127900	1059 (3.36-3.28)

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	AAA	517	87% 11%
1	DDD	517	3% 91% 8%
2	BBB	757	8% 79% 11% 10%
2	EEE	757	9% 79% 10% 11%
3	CCC	137	18% 74% 12% 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	FFF	137	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 21144 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	515	Total	C	N	O	S	0	0	0
			4142	2622	703	787	30			
1	DDD	514	Total	C	N	O	S	0	0	0
			4131	2616	699	786	30			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	200	MET	-	initiating methionine	UNP M9TI86
DDD	200	MET	-	initiating methionine	UNP M9TI86

- Molecule 2 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	BBB	685	Total	C	N	O	S	0	0	0
			5472	3455	942	1031	44			
2	EEE	675	Total	C	N	O	S	0	0	0
			5385	3401	924	1017	43			

- Molecule 3 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	CCC	118	Total	C	N	O	S	0	0	0
			977	616	175	176	10			
3	FFF	125	Total	C	N	O	S	0	0	0
			1035	651	187	187	10			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
CCC	129	SER	-	linker	UNP X5F427
CCC	130	GLY	-	linker	UNP X5F427

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
CCC	131	SER	-	linker	UNP X5F427
CCC	132	GLU	-	expression tag	UNP X5F427
CCC	133	ASN	-	expression tag	UNP X5F427
CCC	134	LEU	-	expression tag	UNP X5F427
CCC	135	TYR	-	expression tag	UNP X5F427
CCC	136	PHE	-	expression tag	UNP X5F427
CCC	137	GLN	-	expression tag	UNP X5F427
FFF	129	SER	-	linker	UNP X5F427
FFF	130	GLY	-	linker	UNP X5F427
FFF	131	SER	-	linker	UNP X5F427
FFF	132	GLU	-	expression tag	UNP X5F427
FFF	133	ASN	-	expression tag	UNP X5F427
FFF	134	LEU	-	expression tag	UNP X5F427
FFF	135	TYR	-	expression tag	UNP X5F427
FFF	136	PHE	-	expression tag	UNP X5F427
FFF	137	GLN	-	expression tag	UNP X5F427

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

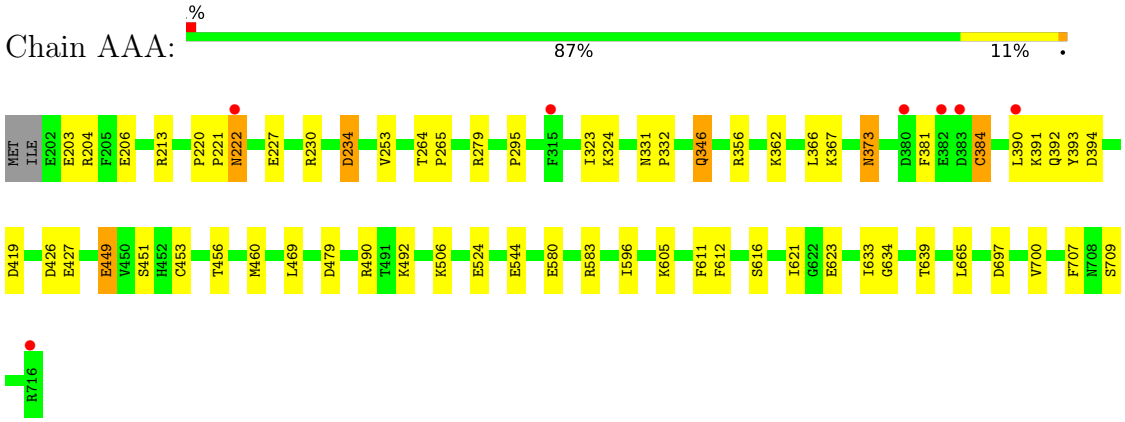
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	BBB	1	Total Mg 1 1	0	0
4	EEE	1	Total Mg 1 1	0	0

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.

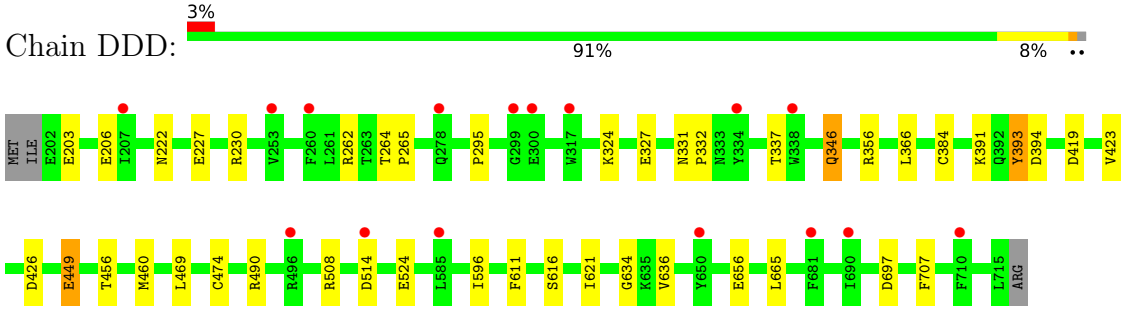
• Molecule 1: Polymerase acidic protein

Chain AAA:



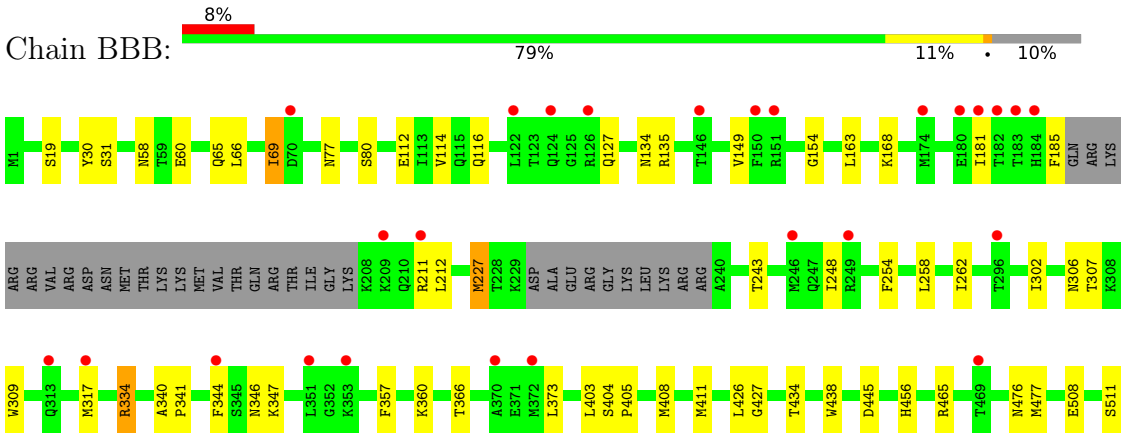
• Molecule 1: Polymerase acidic protein

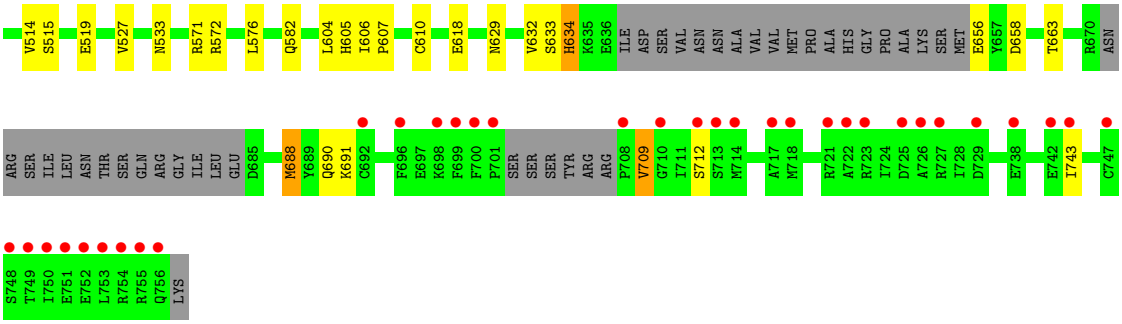
Chain DDD:

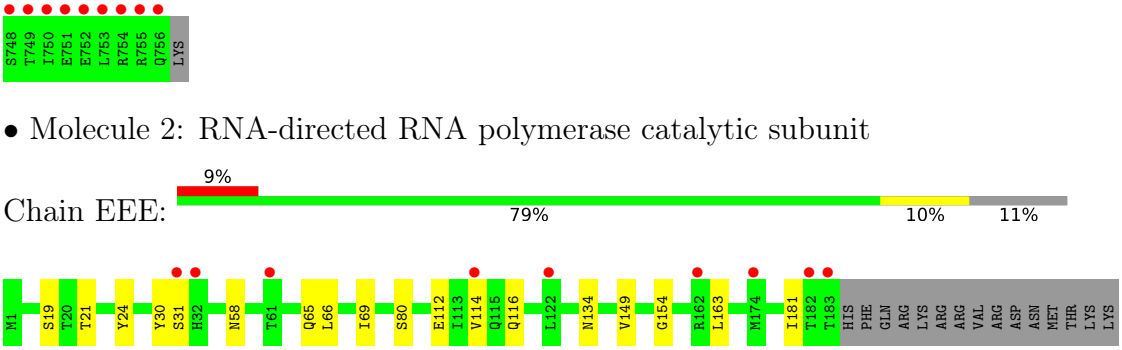


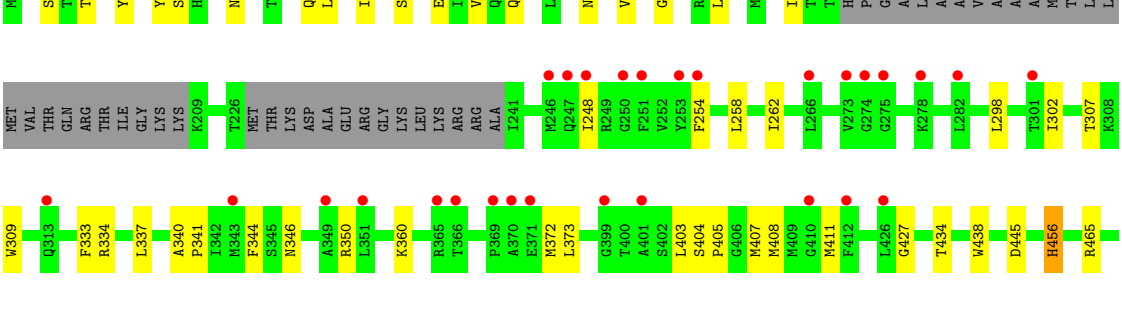
• Molecule 2: RNA-directed RNA polymerase catalytic subunit

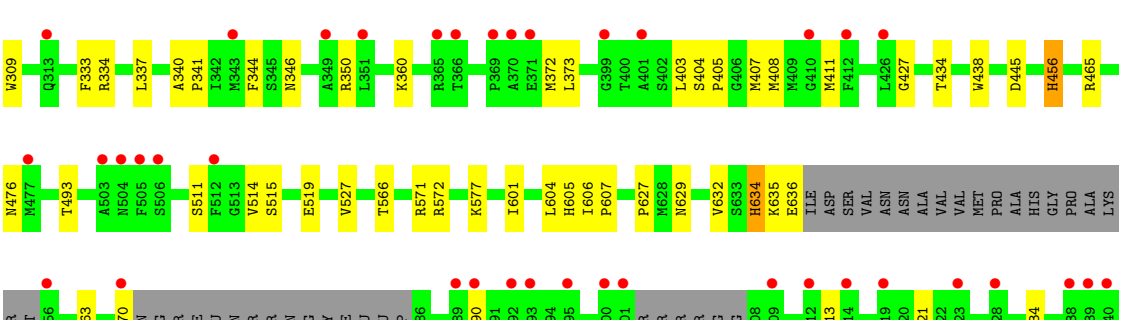
Chain BBB:









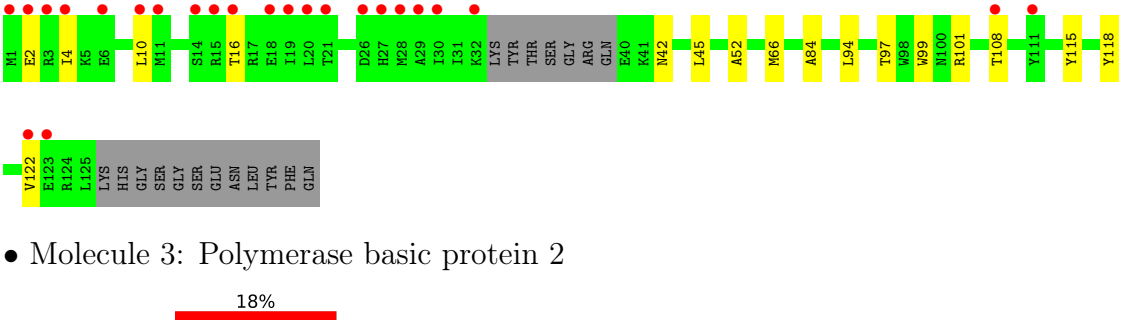


• Molecule 2: RNA-directed RNA polymerase catalytic subunit

Chain EEE:

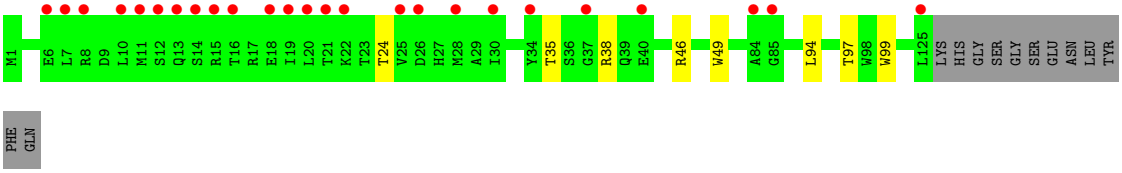
• Molecule 3: Polymerase basic protein 2

Chain CCC:



• Molecule 3: Polymerase basic protein 2

Chain FFF:


M1
E6
L7
R8
D9
L10
M11
S12
Q13
S14
R15
T16
R17
E18
I19
L20
T21
K22
T23
T24
V25
D26
H27
M28
A29
I30
Y34
T35
S36
G37
R38
Q39
E40
R46
W49
A84
G85
L94
T97
W98
W99
L125
LYS
HIS
GLY
SER
GLY
SER
GLU
ASN
LEU
TYR
PHE
GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	75.51Å 143.17Å 335.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	132.02 – 3.33 131.68 – 3.32	Depositor EDS
% Data completeness (in resolution range)	92.4 (132.02-3.33) 92.4 (131.68-3.32)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.78 (at 3.33Å)	Xtriage
Refinement program	REFMAC 5.8.0257	Depositor
R, R_{free}	0.204 , 0.268 0.205 , 0.264	Depositor DCC
R_{free} test set	2496 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	108.4	Xtriage
Anisotropy	0.024	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 78.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	21144	wwPDB-VP
Average B, all atoms (Å^2)	129.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.07% 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: 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	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	AAA	0.63	0/4231	0.73	0/5712
1	DDD	0.64	0/4220	0.72	0/5698
2	BBB	0.66	0/5580	0.72	0/7524
2	EEE	0.66	0/5491	0.72	0/7407
3	CCC	0.66	0/996	0.73	0/1340
3	FFF	0.66	0/1056	0.72	0/1421
All	All	0.65	0/21574	0.72	0/29102

There are no bond length outliers.

There are no bond angle outliers.

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	AAA	4142	0	4109	29	0
1	DDD	4131	0	4096	15	0
2	BBB	5472	0	5436	45	0
2	EEE	5385	0	5348	38	0
3	CCC	977	0	1003	12	0
3	FFF	1035	0	1062	8	0
4	BBB	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	EEE	1	0	0	0	0
All	All	21144	0	21054	120	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 3.

The worst 5 of 120 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:AAA:612:PHE:HA	1:AAA:633:ILE:HG21	1.64	0.78
2:BBB:572:ARG:HB3	2:BBB:576:LEU:HD13	1.74	0.69
2:EEE:519:GLU:OE2	2:EEE:572:ARG:NH1	2.27	0.68
2:BBB:519:GLU:OE2	2:BBB:572:ARG:NH1	2.26	0.67
2:EEE:629:ASN:HD21	2:EEE:632:VAL:HG23	1.59	0.67

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	AAA	513/517 (99%)	478 (93%)	30 (6%)	5 (1%)	15	47
1	DDD	512/517 (99%)	474 (93%)	34 (7%)	4 (1%)	19	51
2	BBB	673/757 (89%)	621 (92%)	47 (7%)	5 (1%)	22	55
2	EEE	663/757 (88%)	608 (92%)	49 (7%)	6 (1%)	17	49
3	CCC	114/137 (83%)	105 (92%)	8 (7%)	1 (1%)	17	49
3	FFF	123/137 (90%)	114 (93%)	9 (7%)	0	100	100
All	All	2598/2822 (92%)	2400 (92%)	177 (7%)	21 (1%)	19	51

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	EEE	634	HIS
1	AAA	222	ASN
1	AAA	394	ASP
3	CCC	84	ALA
1	DDD	222	ASN

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	AAA	462/464 (100%)	438 (95%)	24 (5%)	23	55
1	DDD	461/464 (99%)	440 (95%)	21 (5%)	27	60
2	BBB	604/668 (90%)	570 (94%)	34 (6%)	21	53
2	EEE	595/668 (89%)	567 (95%)	28 (5%)	26	59
3	CCC	109/125 (87%)	106 (97%)	3 (3%)	43	71
3	FFF	115/125 (92%)	114 (99%)	1 (1%)	78	88
All	All	2346/2514 (93%)	2235 (95%)	111 (5%)	26	59

5 of 111 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	BBB	709	VAL
3	FFF	35	THR
1	DDD	393	TYR
2	EEE	734	ARG
2	EEE	434	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 2 ligands modelled in this entry, 2 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 (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	AAA	515/517 (99%)	0.40	7 (1%) 75 75	59, 88, 175, 266	0
1	DDD	514/517 (99%)	0.34	16 (3%) 49 48	85, 127, 183, 221	0
2	BBB	685/757 (90%)	0.62	59 (8%) 10 10	60, 118, 207, 239	0
2	EEE	675/757 (89%)	0.65	70 (10%) 6 6	80, 139, 205, 241	0
3	CCC	118/137 (86%)	0.94	24 (20%) 1 0	80, 151, 216, 241	0
3	FFF	125/137 (91%)	1.04	25 (20%) 1 0	86, 148, 217, 233	0
All	All	2632/2822 (93%)	0.56	201 (7%) 13 14	59, 123, 204, 266	0

The worst 5 of 201 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	BBB	750	ILE	7.6
2	BBB	755	ARG	7.4
2	BBB	751	GLU	6.5
1	AAA	383	ASP	5.9
2	BBB	756	GLN	5.9

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

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
4	MG	EEE	801	1/1	0.94	0.08	103,103,103,103	0
4	MG	BBB	801	1/1	0.96	0.15	69,69,69,69	0

6.5 Other polymers

There are no such residues in this entry.