wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2023 – 11:31 PM EDT

PDB ID	:	7N9X
Title	:	CA-targeting nanobody is a tool for studying HIV-1 capsid lattice interactions
Authors	:	Gerber, E.E.; Digianantonio, K.M.; Tripler, T.N.; Smaga, S.S.; Summers, B.J.; Xiong, Y.
Deposited on	:	2021-06-18
Resolution	:	3.51 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 (1)) were used in the production of this report:

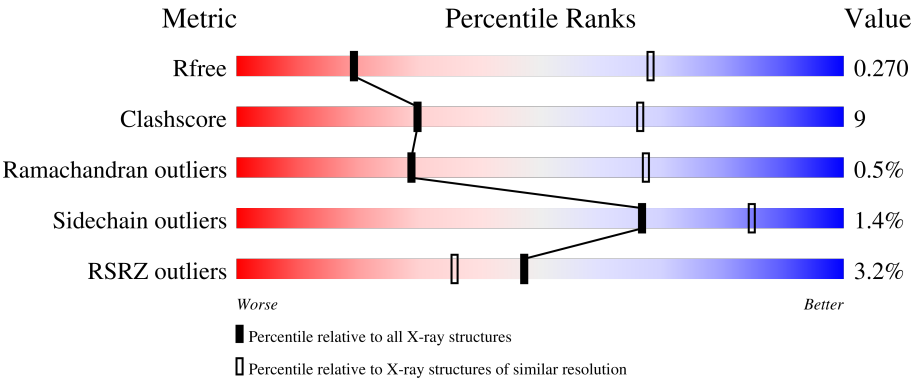
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.51 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.



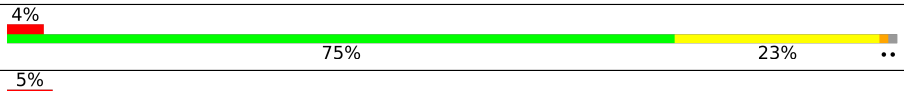
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1161 (3.60-3.44)
Clashscore	141614	1244 (3.60-3.44)
Ramachandran outliers	138981	1206 (3.60-3.44)
Sidechain outliers	138945	1207 (3.60-3.44)
RSRZ outliers	127900	1080 (3.60-3.44)

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	222	81% 17%
1	BBB	222	82% 16%
1	CCC	222	82% 18%
2	FFF	115	80% 17%
3	EEE	112	6% 70% 29%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	DDD	114	
5	GGG	165	
5	HHH	165	
5	III	165	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 11458 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 Capsid protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	BBB	219	Total	C	N	O	S	0	9	0
			1700	1069	298	319	14			
1	CCC	222	Total	C	N	O	S	0	9	0
			1715	1078	301	322	14			
1	AAA	219	Total	C	N	O	S	0	8	0
			1700	1069	298	319	14			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
BBB	14	CYS	ALA	conflict	UNP B6DRA0
BBB	45	CYS	GLU	conflict	UNP B6DRA0
BBB	184	ALA	TRP	conflict	UNP B6DRA0
BBB	185	ALA	MET	conflict	UNP B6DRA0
CCC	14	CYS	ALA	conflict	UNP B6DRA0
CCC	45	CYS	GLU	conflict	UNP B6DRA0
CCC	184	ALA	TRP	conflict	UNP B6DRA0
CCC	185	ALA	MET	conflict	UNP B6DRA0
AAA	14	CYS	ALA	conflict	UNP B6DRA0
AAA	45	CYS	GLU	conflict	UNP B6DRA0
AAA	184	ALA	TRP	conflict	UNP B6DRA0
AAA	185	ALA	MET	conflict	UNP B6DRA0

- Molecule 2 is a protein called Nanobody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	FFF	113	Total	C	N	O	S	0	0	0
			857	538	151	164	4			

- Molecule 3 is a protein called Nanobody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	EEE	112	Total	C	N	O	S	0	0	0
			851	535	150	162	4			

- Molecule 4 is a protein called Nanobody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	DDD	112	Total	C	N	O	S	0	0	0
			852	535	153	161	3			

- Molecule 5 is a protein called Peptidyl-prolyl cis-trans isomerase A.

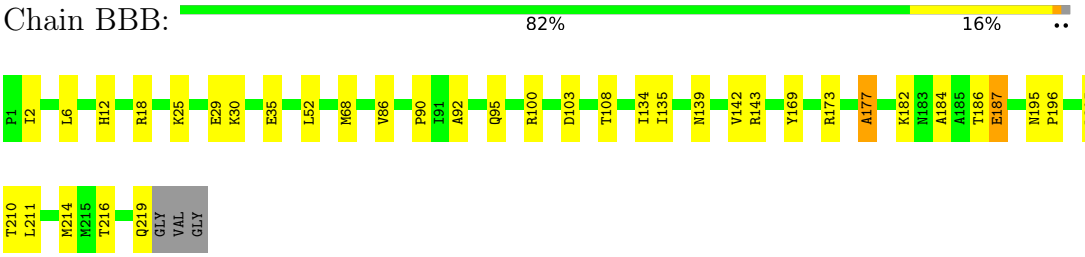
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	GGG	165	Total	C	N	O	S	0	0	0
			1266	802	218	237	9			
5	HHH	163	Total	C	N	O	S	12	0	0
			1251	792	216	235	8			
5	III	165	Total	C	N	O	S	0	0	0
			1266	802	218	237	9			

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.

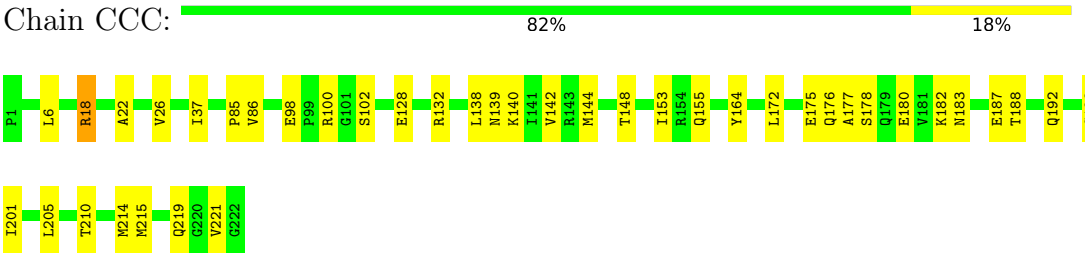
• Molecule 1: Capsid protein

Chain BBB:



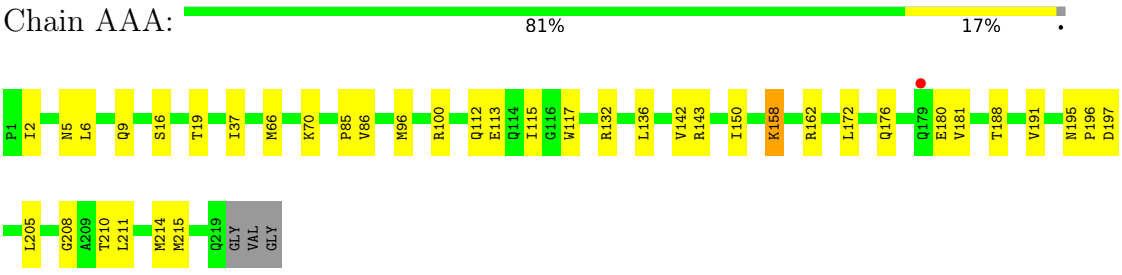
• Molecule 1: Capsid protein

Chain CCC:



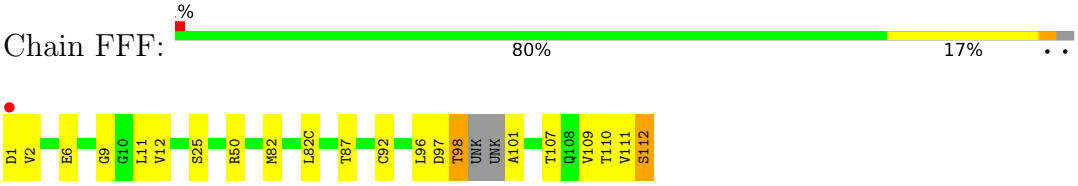
• Molecule 1: Capsid protein

Chain AAA:



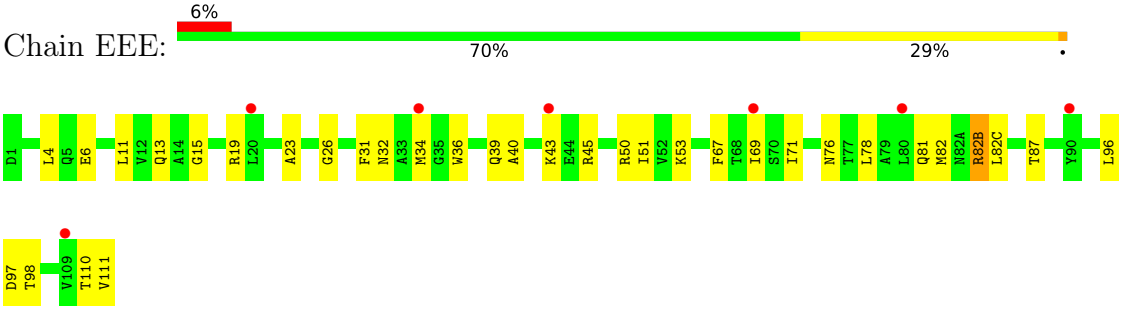
• Molecule 2: Nanobody

Chain FFF:



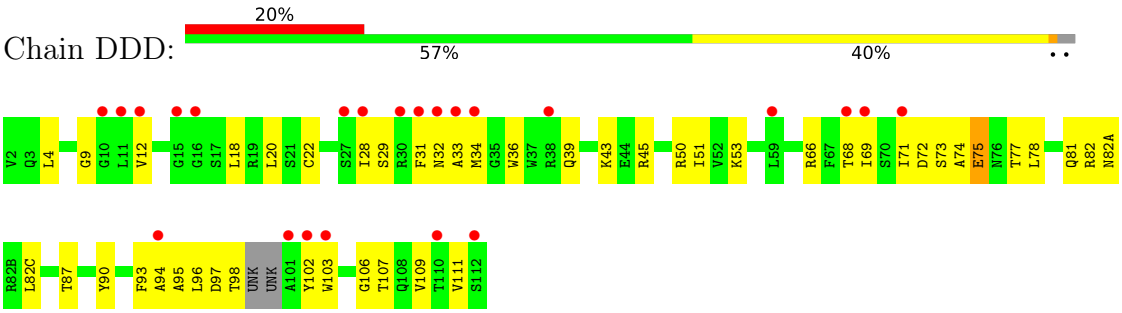
• Molecule 3: Nanobody

Chain EEE:



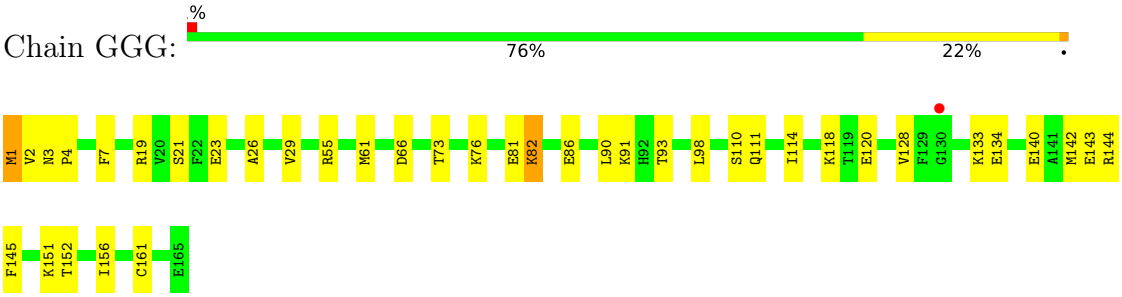
• Molecule 4: Nanobody

Chain DDD:



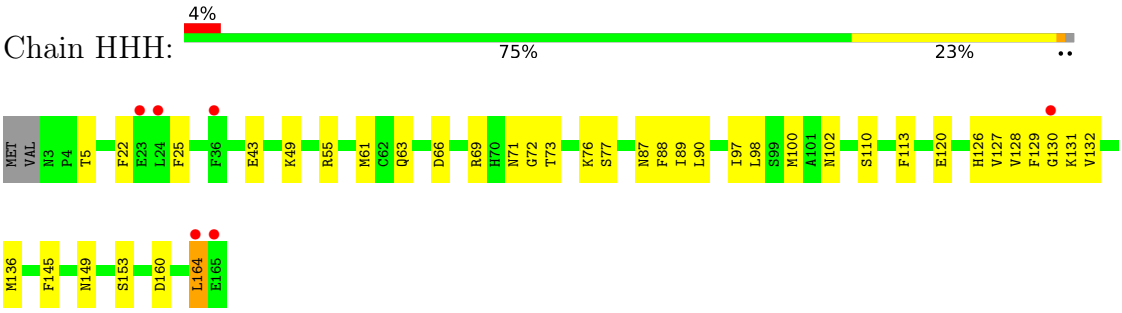
• Molecule 5: Peptidyl-prolyl cis-trans isomerase A

Chain GGG:



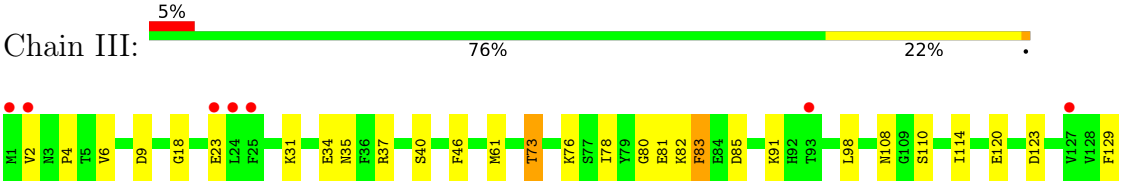
• Molecule 5: Peptidyl-prolyl cis-trans isomerase A

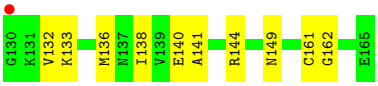
Chain HHH:



• Molecule 5: Peptidyl-prolyl cis-trans isomerase A

Chain III:




G130
K131
V132
K133
M136
N137
I138
V139
E140
A141
R144
N149
C161
G162
E165

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	127.95Å 137.85Å 313.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.58 – 3.51 49.58 – 3.51	Depositor EDS
% Data completeness (in resolution range)	94.4 (49.58-3.51) 94.4 (49.58-3.51)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.55 (at 3.48Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.245 , 0.270 0.245 , 0.270	Depositor DCC
R_{free} test set	1668 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	94.4	Xtriage
Anisotropy	0.296	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 106.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	11458	wwPDB-VP
Average B, all atoms (Å^2)	180.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% 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

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	0.30	0/1737	0.62	0/2361
1	BBB	0.31	0/1737	0.60	0/2361
1	CCC	0.30	0/1752	0.58	0/2381
2	FFF	0.30	0/872	0.75	2/1178 (0.2%)
3	EEE	0.32	0/867	0.69	0/1173
4	DDD	0.42	1/867 (0.1%)	0.87	1/1171 (0.1%)
5	GGG	0.37	1/1294 (0.1%)	0.74	2/1733 (0.1%)
5	HHH	0.33	0/1279	0.65	0/1713
5	III	0.35	0/1294	0.74	3/1733 (0.2%)
All	All	0.33	2/11699 (0.0%)	0.68	8/15804 (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
5	III	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	GGG	144	ARG	CG-CD	-5.76	1.37	1.51
4	DDD	22	CYS	CB-SG	-5.33	1.73	1.81

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	GGG	144	ARG	CB-CG-CD	-9.65	86.52	111.60
5	III	82	LYS	C-N-CA	-6.96	104.30	121.70
4	DDD	75	GLU	CA-CB-CG	6.28	127.21	113.40
2	FFF	11	LEU	CA-CB-CG	5.56	128.10	115.30
5	III	82	LYS	CA-CB-CG	5.51	125.53	113.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	III	46	PHE	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	AAA	1700	0	1697	25	0
1	BBB	1700	0	1691	26	0
1	CCC	1715	0	1707	24	0
2	FFF	857	0	838	14	0
3	EEE	851	0	834	24	0
4	DDD	852	0	835	41	0
5	GGG	1266	0	1237	26	0
5	HHH	1251	0	1216	26	0
5	III	1266	0	1237	22	0
All	All	11458	0	11292	214	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 9.

The worst 5 of 214 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)
4:DDD:9:GLY:HA3	4:DDD:107:THR:HG21	1.40	1.03
4:DDD:4:LEU:HD11	4:DDD:102:TYR:HB3	1.44	0.98
5:GGG:118:LYS:HE2	5:GGG:120:GLU:HG2	1.59	0.83
1:AAA:85:PRO:HB2	5:GGG:73:THR:HG21	1.61	0.81
1:AAA:86:VAL:HG21	1:AAA:100:ARG:HG2	1.64	0.78

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	217/222 (98%)	208 (96%)	9 (4%)	0	100	100
1	BBB	217/222 (98%)	211 (97%)	4 (2%)	2 (1%)	17	57
1	CCC	220/222 (99%)	215 (98%)	3 (1%)	2 (1%)	17	57
2	FFF	109/115 (95%)	107 (98%)	2 (2%)	0	100	100
3	EEE	110/112 (98%)	108 (98%)	1 (1%)	1 (1%)	17	57
4	DDD	108/114 (95%)	106 (98%)	1 (1%)	1 (1%)	17	57
5	GGG	163/165 (99%)	156 (96%)	7 (4%)	0	100	100
5	HHH	161/165 (98%)	153 (95%)	7 (4%)	1 (1%)	25	64
5	III	163/165 (99%)	156 (96%)	6 (4%)	1 (1%)	25	64
All	All	1468/1502 (98%)	1420 (97%)	40 (3%)	8 (0%)	29	67

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	EEE	97	ASP
4	DDD	97	ASP
1	BBB	177	ALA
1	CCC	176	GLN
5	III	73	THR

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	186/187 (100%)	184 (99%)	2 (1%)	73 87
1	BBB	186/187 (100%)	186 (100%)	0	100 100
1	CCC	187/187 (100%)	185 (99%)	2 (1%)	73 87
2	FFF	88/88 (100%)	86 (98%)	2 (2%)	50 77
3	EEE	87/87 (100%)	86 (99%)	1 (1%)	73 87
4	DDD	87/87 (100%)	87 (100%)	0	100 100
5	GGG	133/133 (100%)	129 (97%)	4 (3%)	41 71
5	HHH	131/133 (98%)	129 (98%)	2 (2%)	65 84
5	III	133/133 (100%)	129 (97%)	4 (3%)	41 71
All	All	1218/1222 (100%)	1201 (99%)	17 (1%)	67 85

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	III	61	MET
5	III	149	ASN
5	GGG	1	MET
5	GGG	55	ARG
5	GGG	61	MET

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

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	219/222 (98%)	-0.22	1 (0%) 91 84	83, 127, 235, 275	0
1	BBB	219/222 (98%)	-0.22	0 100 100	83, 119, 191, 304	0
1	CCC	222/222 (100%)	-0.28	0 100 100	81, 121, 177, 253	0
2	FFF	113/115 (98%)	0.23	1 (0%) 84 73	116, 183, 257, 280	0
3	EEE	112/112 (100%)	0.44	7 (6%) 20 15	134, 224, 292, 311	0
4	DDD	112/114 (98%)	1.09	23 (20%) 1 0	207, 299, 361, 423	0
5	GGG	165/165 (100%)	0.01	1 (0%) 89 81	122, 200, 289, 327	0
5	HHH	162/165 (98%)	0.10	6 (3%) 41 30	121, 207, 289, 389	2 (1%)
5	III	165/165 (100%)	0.15	8 (4%) 30 23	121, 204, 295, 391	0
All	All	1489/1502 (99%)	0.05	47 (3%) 47 36	81, 169, 299, 423	2 (0%)

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	HHH	165	GLU	8.2
4	DDD	27	SER	7.0
4	DDD	10	GLY	6.5
4	DDD	69	ILE	5.2
5	III	130	GLY	4.9

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.