wwPDB X-ray Structure Validation Summary Report (i)

Aug 20, 2023 – 10:07 AM EDT

PDB ID	:	2NOJ
Title	:	Crystal structure of Ehp / C3d complex
Authors	:	Hammel, M.; Geisbrecht, B.V.
Deposited on	:	2006-10-25
Resolution	:	2.70 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

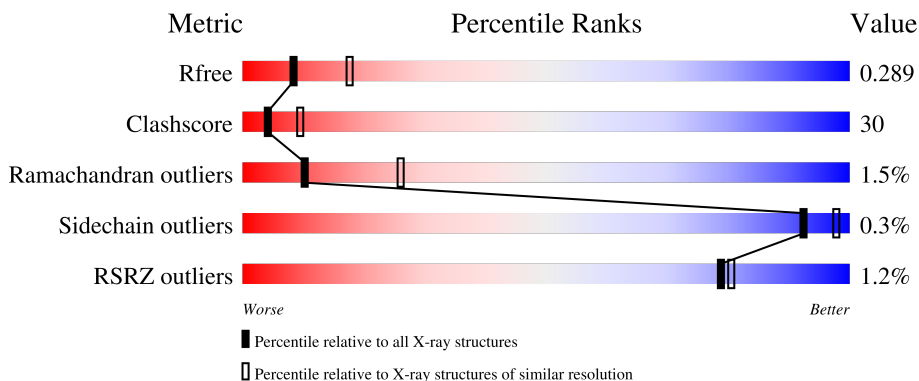|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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	Percentile Ranks	Value
Rfree		0.289
Clashscore		30
Ramachandran outliers		1.5%
Sidechain outliers		0.3%
RSRZ outliers		1.2%

Worse ← → *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	A	297	% 51% 42% • 7%
1	C	297	% 52% 39% • 7%
1	E	297	% 53% 38% • 7%
1	G	297	2% 56% 38% • •
2	B	80	40% 32% 28%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	80	56% 16% 28%
2	F	80	% 44% 29% 28%
2	H	80	50% 22% 28%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9768 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 Complement C3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	277	Total	C	N	O	S	0	0	0
			2007	1294	334	371	8			
1	C	275	Total	C	N	O	S	0	0	0
			1946	1251	322	366	7			
1	E	276	Total	C	N	O	S	0	0	0
			1943	1251	327	358	7			
1	G	284	Total	C	N	O	S	0	0	0
			2002	1284	334	375	9			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	991	GLY	-	expression tag	UNP P01024
A	992	SER	-	expression tag	UNP P01024
A	993	ARG	-	expression tag	UNP P01024
A	994	SER	-	expression tag	UNP P01024
A	995	THR	-	expression tag	UNP P01024
A	1010	ALA	CYS	engineered mutation	UNP P01024
C	991	GLY	-	expression tag	UNP P01024
C	992	SER	-	expression tag	UNP P01024
C	993	ARG	-	expression tag	UNP P01024
C	994	SER	-	expression tag	UNP P01024
C	995	THR	-	expression tag	UNP P01024
C	1010	ALA	CYS	engineered mutation	UNP P01024
E	991	GLY	-	expression tag	UNP P01024
E	992	SER	-	expression tag	UNP P01024
E	993	ARG	-	expression tag	UNP P01024
E	994	SER	-	expression tag	UNP P01024
E	995	THR	-	expression tag	UNP P01024
E	1010	ALA	CYS	engineered mutation	UNP P01024
G	991	GLY	-	expression tag	UNP P01024
G	992	SER	-	expression tag	UNP P01024
G	993	ARG	-	expression tag	UNP P01024

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	994	SER	-	expression tag	UNP P01024
G	995	THR	-	expression tag	UNP P01024
G	1010	ALA	CYS	engineered mutation	UNP P01024

- Molecule 2 is a protein called Efb homologous protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	58	Total 454	C 285	N 89	O 80	0	0	0
2	D	58	Total 449	C 282	N 87	O 80	0	0	0
2	F	58	Total 454	C 287	N 89	O 78	0	0	0
2	H	58	Total 442	C 278	N 86	O 78	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	63	GLU	ASN	engineered mutation	UNP Q99UV2
D	63	GLU	ASN	engineered mutation	UNP Q99UV2
F	63	GLU	ASN	engineered mutation	UNP Q99UV2
H	63	GLU	ASN	engineered mutation	UNP Q99UV2

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	12	Total 12	O 12	0	0
3	B	7	Total 7	O 7	0	0
3	C	11	Total 11	O 11	0	0
3	D	6	Total 6	O 6	0	0
3	E	12	Total 12	O 12	0	0
3	F	5	Total 5	O 5	0	0
3	G	14	Total 14	O 14	0	0

Continued on next page...

Continued from previous page...

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	H	4	Total 4	O 4	0	0

3 Residue-property plots ⓘ

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.

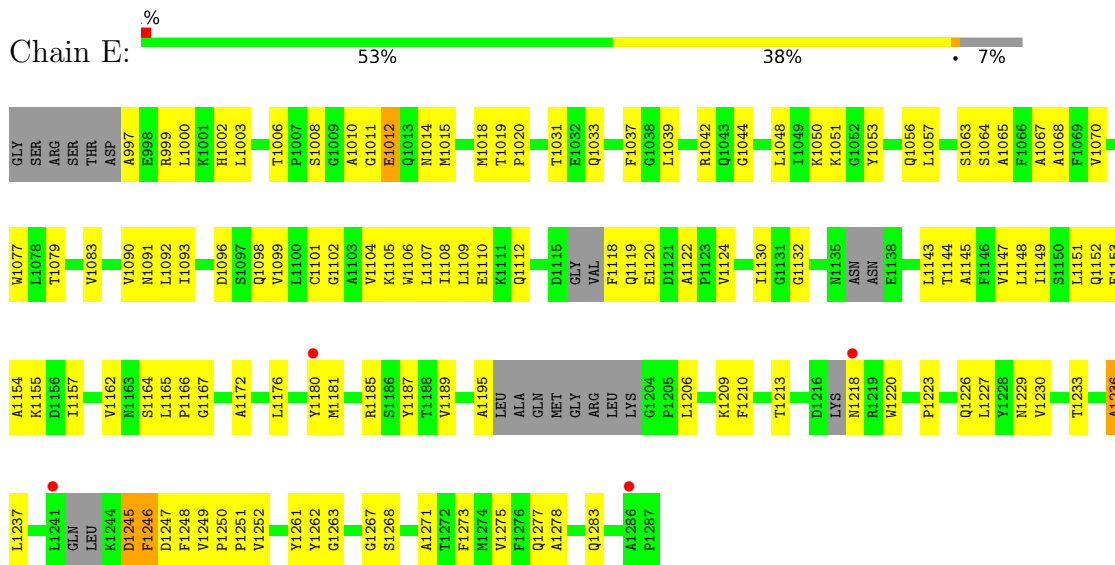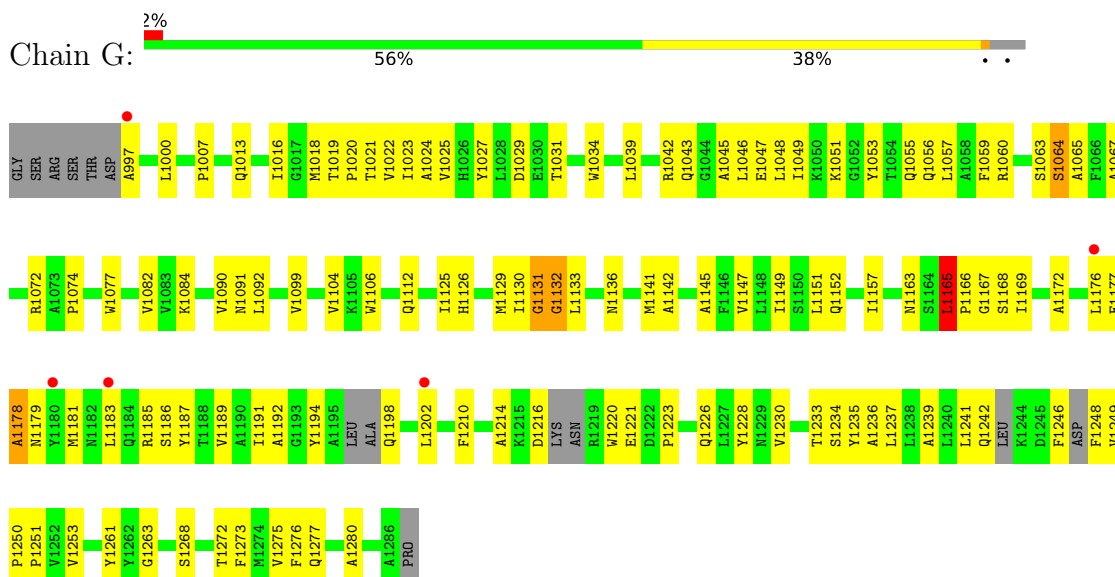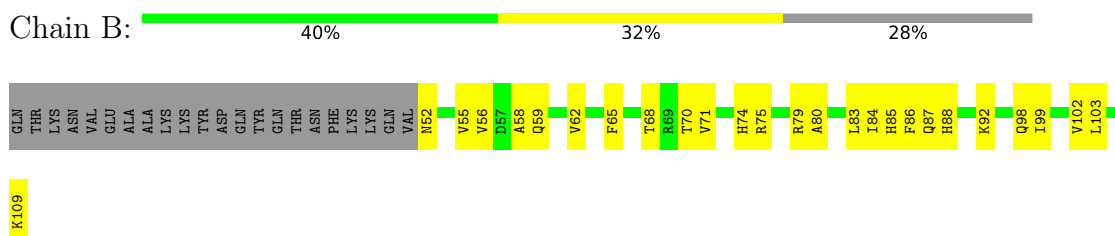- Molecule 1: Complement C3
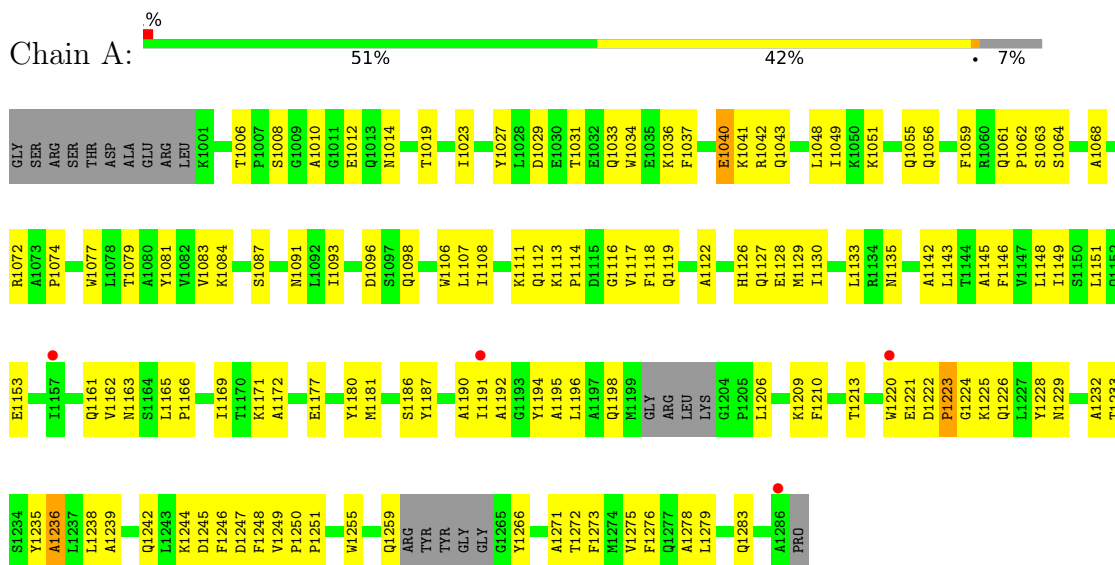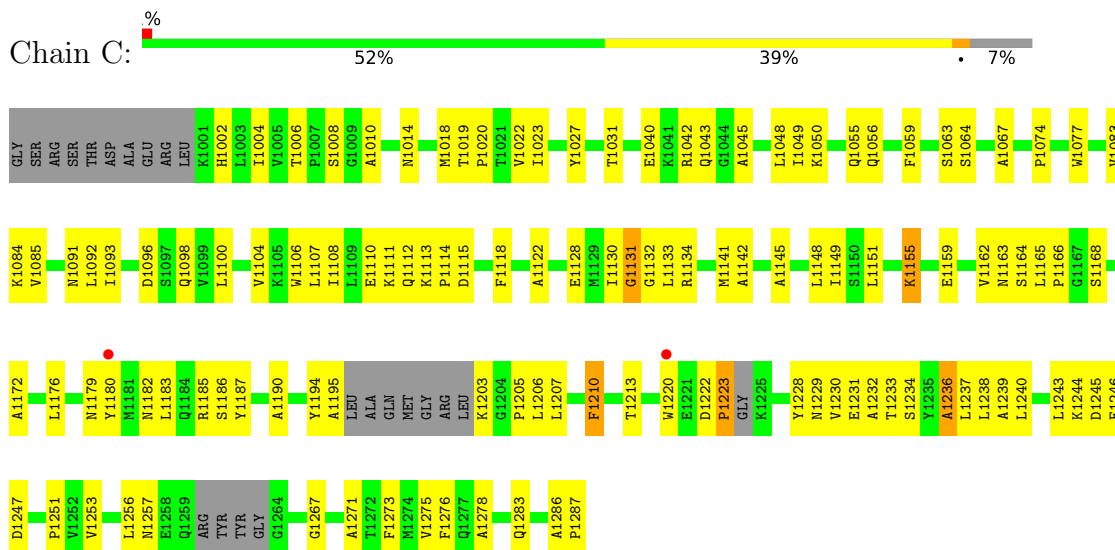
Chain A:

- Molecule 1: Complement C3

Chain C:

- Molecule 1: Complement C3

Chain E:

• Molecule 1: Complement C3

Chain G:

• Molecule 2: Efb homologous protein

Chain B:

• Molecule 2: Efb homologous protein

Chain D:

GLN THR LYS ASN VAL GLU ALA ALA LYS LYS TYR ASP TYR GLN THR ASN PHE LYS LYS GLN VAL **N52** V55 A58 T68 R69 T70 V71 A72 T73 H74 R79 A80 I84 L103 L108 K109

• Molecule 2: Efb homologous protein

%

Chain F:
44%   29%   28%

GLN THR LYS ASN VAL GLU ALA ALA LYS LYS TYR ASP TYR GLN THR ASN PHE LYS LYS GLN VAL **N52** K53 K54 V55 V56 D57 A58 Q59 V62 F65 K66 R67 T68 R69 T70 V71 A72 T73 H74 R75 A80 V81 N82 F86 Q87 H88 S89 Y90 L95 Q98 V102

T107 L108 K109

• Molecule 2: Efb homologous protein

Chain H:
50%   22%   28%

GLN THR LYS ASN VAL GLU ALA ALA LYS LYS TYR ASP TYR GLN THR ASN PHE LYS LYS GLN VAL **N52** K53 K54 Q59 F65 T68 R69 T70 V71 A72 T73 H74 Q78 F86 Q87 H88 L95 Q96 V102 M106 K109
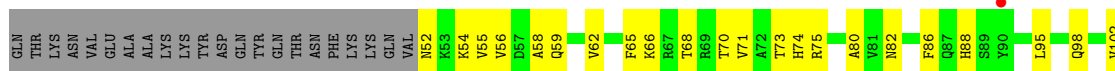
4   Data and refinement statistics   (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	67.89Å 91.03Å 122.59Å 90.00° 89.93° 90.00°	Depositor
Resolution (Å)	50.00 – 2.70 49.76 – 2.70	Depositor EDS
% Data completeness (in resolution range)	80.6 (50.00-2.70) 80.6 (49.76-2.70)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.05 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.291 , 0.284 0.291 , 0.289	Depositor DCC
R_{free} test set	1655 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	39.6	Xtriage
Anisotropy	0.502	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 73.4	EDS
L-test for twinning[2]	$< |L| > = 0.54$, $< L^2 > = 0.38$	Xtriage
Estimated twinning fraction	0.390 for h,-k,-l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	9768	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 47.24 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.0204e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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	A	0.51	1/2048 (0.0%)	0.63	0/2800
1	C	0.53	1/1985 (0.1%)	0.67	3/2718 (0.1%)
1	E	0.48	1/1979 (0.1%)	0.65	1/2703 (0.0%)
1	G	0.50	0/2040	0.67	1/2789 (0.0%)
2	B	0.50	0/460	0.66	0/620
2	D	0.44	0/454	0.63	0/612
2	F	0.46	0/460	0.61	0/620
2	H	0.44	0/447	0.63	0/604
All	All	0.50	3/9873 (0.0%)	0.65	5/13466 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1236	ALA	CA-CB	-7.10	1.37	1.52
1	C	1236	ALA	CA-CB	-6.03	1.39	1.52
1	E	1236	ALA	CA-CB	-5.00	1.42	1.52

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	1210	PHE	N-CA-C	-7.33	91.20	111.00
1	G	1223	PRO	N-CA-CB	6.26	110.81	103.30
1	E	1223	PRO	N-CA-CB	5.79	110.25	103.30
1	C	1223	PRO	N-CA-CB	5.44	109.83	103.30
1	C	1251	PRO	N-CA-CB	5.12	109.45	103.30

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	2007	0	1872	137	0
1	C	1946	0	1743	121	0
1	E	1943	0	1759	122	0
1	G	2002	0	1788	109	0
2	B	454	0	441	34	0
2	D	449	0	436	10	0
2	F	454	0	448	25	0
2	H	442	0	423	18	0
3	A	12	0	0	1	0
3	B	7	0	0	0	0
3	C	11	0	0	1	0
3	D	6	0	0	0	0
3	E	12	0	0	0	0
3	F	5	0	0	0	0
3	G	14	0	0	0	0
3	H	4	0	0	0	0
All	All	9768	0	8910	550	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 30.

The worst 5 of 550 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:1220:TRP:HD1	1:C:1233:THR:HA	1.11	1.15
1:A:1161:GLN:NE2	2:B:85:HIS:HB3	1.62	1.15
1:G:1043:GLN:O	1:G:1047:GLU:HG2	1.47	1.14
1:E:1148:LEU:HD23	1:E:1195:ALA:HB1	1.33	1.09
1:C:1141:MET:HG3	1:C:1183:LEU:HD21	1.38	1.05

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	A	271/297 (91%)	240 (89%)	28 (10%)	3 (1%)	14	34
1	C	267/297 (90%)	228 (85%)	35 (13%)	4 (2%)	10	26
1	E	264/297 (89%)	221 (84%)	40 (15%)	3 (1%)	14	34
1	G	274/297 (92%)	233 (85%)	31 (11%)	10 (4%)	3	7
2	B	56/80 (70%)	53 (95%)	3 (5%)	0	100	100
2	D	56/80 (70%)	53 (95%)	3 (5%)	0	100	100
2	F	56/80 (70%)	52 (93%)	4 (7%)	0	100	100
2	H	56/80 (70%)	53 (95%)	3 (5%)	0	100	100
All	All	1300/1508 (86%)	1133 (87%)	147 (11%)	20 (2%)	10	26

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	1155	LYS
1	A	1223	PRO
1	C	1131	GLY
1	G	1131	GLY
1	G	1202	LEU

5.3.2 Protein sidechains ⓘ

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	A	186/245 (76%)	185 (100%)	1 (0%)	88	96

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	171/245 (70%)	170 (99%)	1 (1%)	86	95
1	E	168/245 (69%)	168 (100%)	0	100	100
1	G	174/245 (71%)	173 (99%)	1 (1%)	86	95
2	B	44/73 (60%)	44 (100%)	0	100	100
2	D	43/73 (59%)	43 (100%)	0	100	100
2	F	44/73 (60%)	44 (100%)	0	100	100
2	H	41/73 (56%)	41 (100%)	0	100	100
All	All	871/1272 (68%)	868 (100%)	3 (0%)	92	98

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1040	GLU
1	C	1040	GLU
1	G	1165	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	1014	ASN
2	H	59	GLN
1	G	1056	GLN
1	G	1257	ASN
2	H	98	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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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.

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, 95th 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	A	277/297 (93%)	-0.05	4 (1%)	75	77	19, 49, 90, 109	0
1	C	275/297 (92%)	-0.08	2 (0%)	87	89	16, 46, 93, 114	0
1	E	276/297 (92%)	-0.09	4 (1%)	75	77	16, 48, 99, 121	0
1	G	284/297 (95%)	-0.12	5 (1%)	68	70	14, 48, 95, 118	0
2	B	58/80 (72%)	-0.40	0	100	100	16, 35, 60, 79	0
2	D	58/80 (72%)	-0.19	0	100	100	18, 34, 54, 80	0
2	F	58/80 (72%)	-0.30	1 (1%)	70	72	13, 36, 63, 82	0
2	H	58/80 (72%)	-0.27	0	100	100	15, 38, 67, 72	0
All	All	1344/1508 (89%)	-0.12	16 (1%)	79	80	13, 44, 93, 121	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1180	TYR	4.6
1	C	1220	TRP	4.3
1	A	1220	TRP	3.6
1	E	1218	ASN	3.5
1	G	1202	LEU	2.9

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.4   Ligands ⓘ

There are no ligands in this entry.

6.5   Other polymers ⓘ

There are no such residues in this entry.