wwPDB X-ray Structure Validation Summary Report (i)

Aug 26, 2023 – 08:00 PM EDT

PDB ID	:	3FOA
Title	:	Crystal structure of the bacteriophage T4 tail sheath protein, deletion mutant gp18M
Authors	:	Aksyuk, A.A.; Leiman, P.G.; Kurochkina, L.P.; Shneider, M.M.; Kostyuchenko, V.A.; Mesyanzhinov, V.V.; Rossmann, M.G.
Deposited on	:	2008-12-29
Resolution	:	3.50 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:

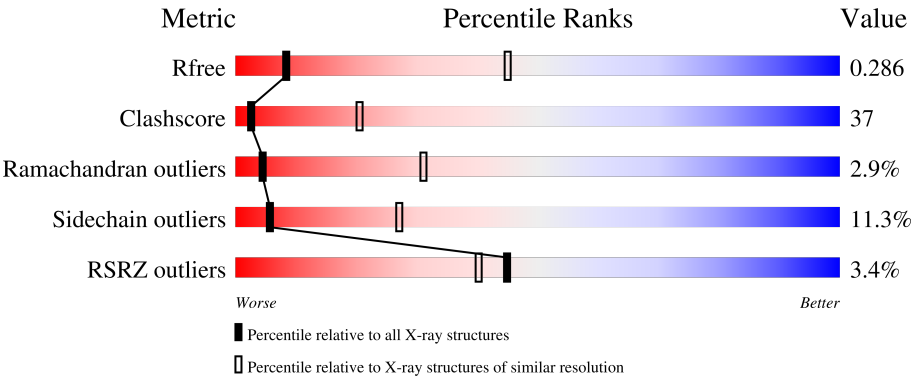
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.50 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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	510	39% 44% 10% • 6%
1	B	510	38% 43% 11% • 6%
1	C	510	42% 41% 10% • 6%
1	D	510	8% 42% 41% 9% • 7%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 14406 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 Tail sheath protein Gp18.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	479	Total	C	N	O	S	0	0	0
			3613	2279	600	726	8			
1	B	478	Total	C	N	O	S	0	0	0
			3605	2275	598	724	8			
1	C	477	Total	C	N	O	S	0	0	0
			3597	2271	596	722	8			
1	D	476	Total	C	N	O	S	0	0	0
			3591	2268	595	720	8			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	100	GLU	ASP	variant	UNP P13332
A	148	ALA	GLY	variant	UNP P13332
A	150	ILE	ASN	variant	UNP P13332
A	151	ILE	TYR	variant	UNP P13332
A	301	GLY	GLU	variant	UNP P13332
A	399	VAL	ALA	variant	UNP P13332
A	454	TYR	HIS	variant	UNP P13332
A	510	PRO	ARG	engineered mutation	UNP P13332
B	100	GLU	ASP	variant	UNP P13332
B	148	ALA	GLY	variant	UNP P13332
B	150	ILE	ASN	variant	UNP P13332
B	151	ILE	TYR	variant	UNP P13332
B	301	GLY	GLU	variant	UNP P13332
B	399	VAL	ALA	variant	UNP P13332
B	454	TYR	HIS	variant	UNP P13332
B	510	PRO	ARG	engineered mutation	UNP P13332
C	100	GLU	ASP	variant	UNP P13332
C	148	ALA	GLY	variant	UNP P13332
C	150	ILE	ASN	variant	UNP P13332
C	151	ILE	TYR	variant	UNP P13332
C	301	GLY	GLU	variant	UNP P13332

Continued on next page...

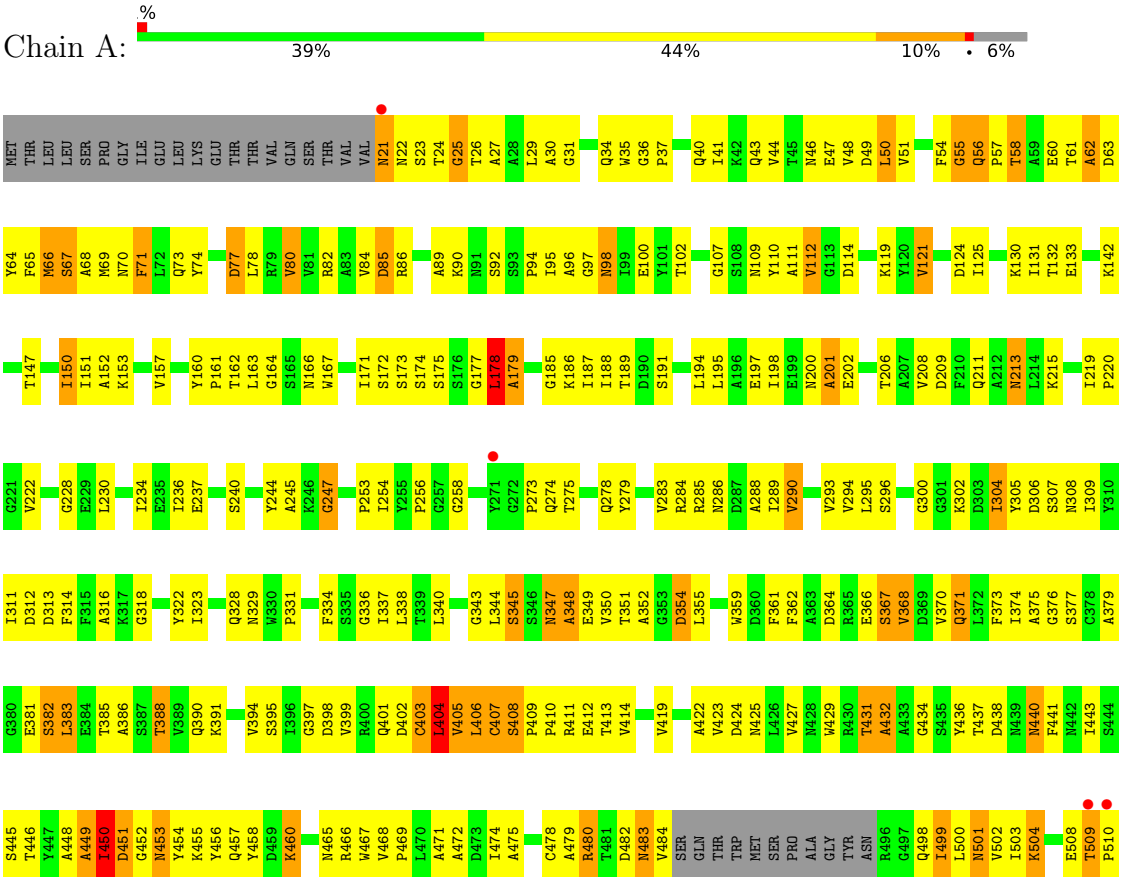
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	399	VAL	ALA	variant	UNP P13332
C	454	TYR	HIS	variant	UNP P13332
C	510	PRO	ARG	engineered mutation	UNP P13332
D	100	GLU	ASP	variant	UNP P13332
D	148	ALA	GLY	variant	UNP P13332
D	150	ILE	ASN	variant	UNP P13332
D	151	ILE	TYR	variant	UNP P13332
D	301	GLY	GLU	variant	UNP P13332
D	399	VAL	ALA	variant	UNP P13332
D	454	TYR	HIS	variant	UNP P13332
D	510	PRO	ARG	engineered mutation	UNP P13332

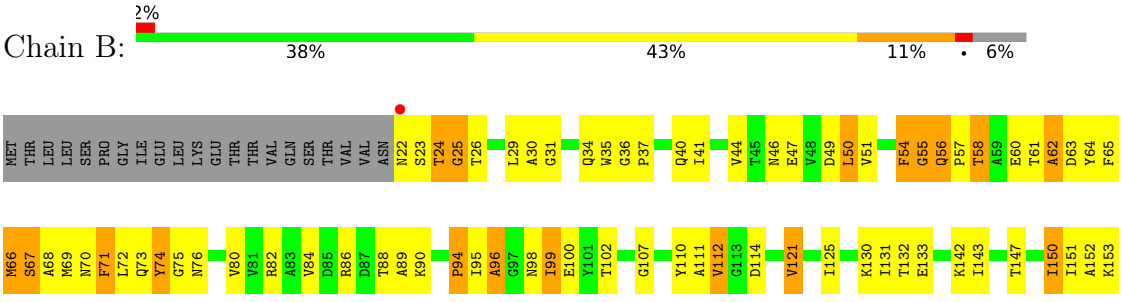
3 Residue-property plots (i)

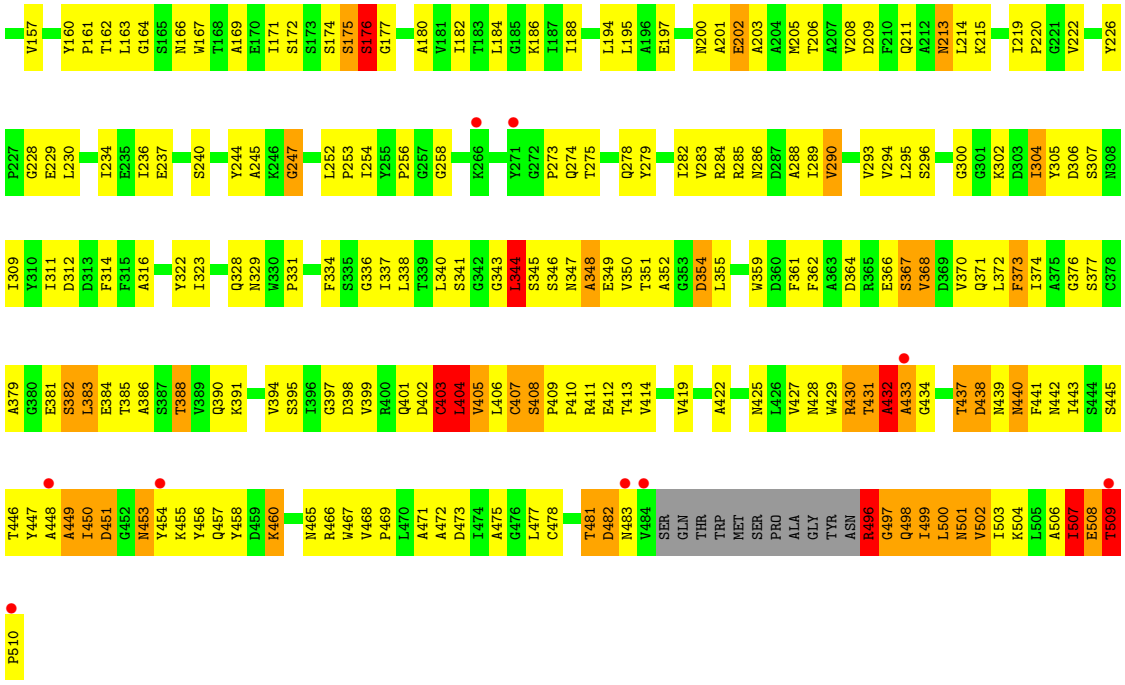
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Tail sheath protein Gp18



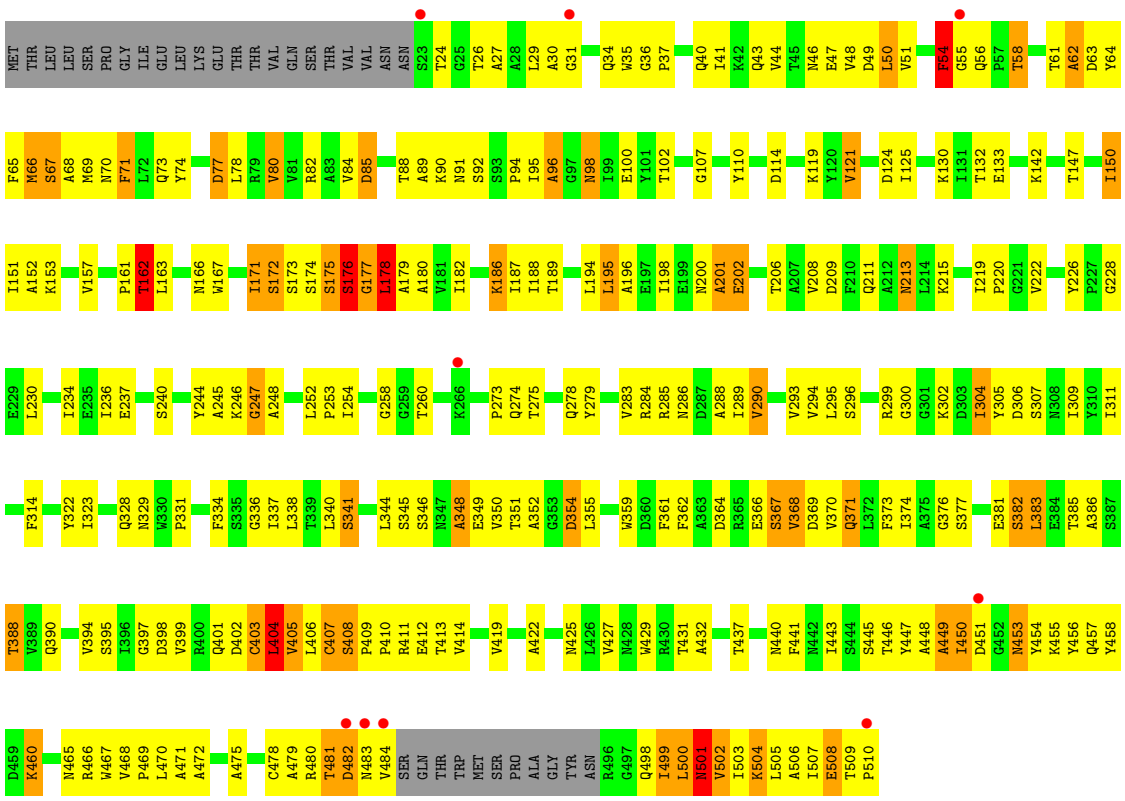
- Molecule 1: Tail sheath protein Gp18





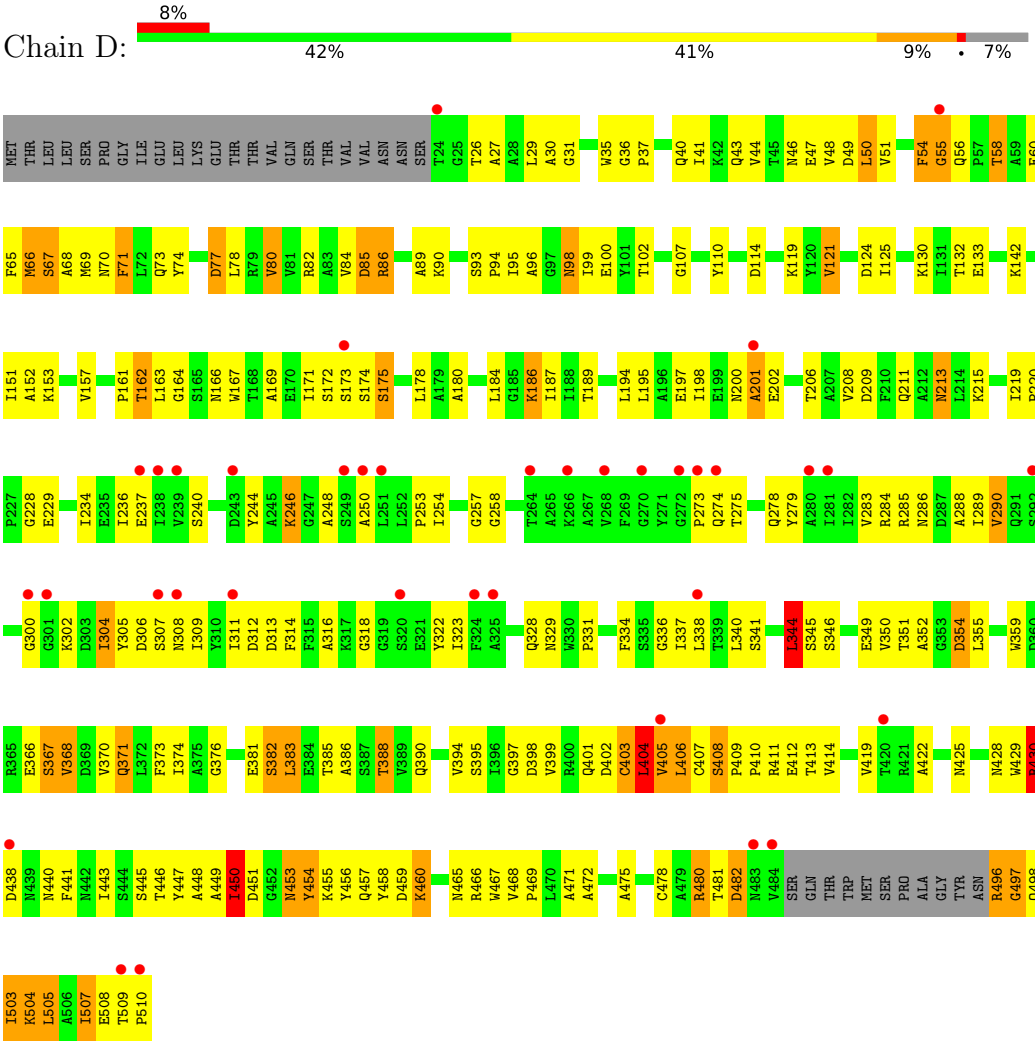
• Molecule 1: Tail sheath protein Gp18

Chain C:



• Molecule 1: Tail sheath protein Gp18

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	99.59Å 116.29Å 433.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.80 – 3.50 49.80 – 3.50	Depositor EDS
% Data completeness (in resolution range)	87.8 (49.80-3.50) 87.8 (49.80-3.50)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.68 (at 3.48Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.267 , 0.299 0.280 , 0.286	Depositor DCC
R_{free} test set	5405 reflections (16.72%)	wwPDB-VP
Wilson B-factor (Å^2)	68.5	Xtriage
Anisotropy	0.673	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 67.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.84	EDS
Total number of atoms	14406	wwPDB-VP
Average B, all atoms (Å^2)	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.77% 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	A	0.35	0/3677	0.78	21/5001 (0.4%)
1	B	0.29	0/3669	0.91	31/4990 (0.6%)
1	C	0.30	0/3661	0.70	19/4979 (0.4%)
1	D	0.28	0/3655	0.88	27/4971 (0.5%)
All	All	0.30	0/14662	0.82	98/19941 (0.5%)

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
1	A	0	4
1	B	0	13
1	C	0	5
1	D	0	2
All	All	0	24

There are no bond length outliers.

The worst 5 of 98 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	96	ALA	N-CA-CB	-22.35	78.81	110.10
1	D	500	LEU	CB-CA-C	-18.07	75.86	110.20
1	A	55	GLY	N-CA-C	18.00	158.11	113.10
1	D	55	GLY	N-CA-C	16.20	153.61	113.10
1	B	55	GLY	N-CA-C	13.58	147.05	113.10

There are no chirality outliers.

5 of 24 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	21	ASN	Peptide
1	A	450	ILE	Peptide
1	A	451	ASP	Peptide
1	A	452	GLY	Peptide
1	B	22	ASN	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	A	3613	0	3554	298	2
1	B	3605	0	3548	280	1
1	C	3597	0	3541	253	1
1	D	3591	0	3536	232	0
All	All	14406	0	14179	1050	2

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 37.

The worst 5 of 1050 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:A:409:PRO:O	1:A:454:TYR:CE1	1.65	1.48
1:A:409:PRO:CD	1:A:451:ASP:O	1.75	1.33
1:B:496:ARG:HG3	1:B:497:GLY:O	1.16	1.29
1:A:379:ALA:CB	1:A:454:TYR:OH	1.84	1.25
1:B:496:ARG:CG	1:B:497:GLY:O	1.89	1.21

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:509:THR:O	1:B:419:VAL:N[4_655]	1.92	0.28
1:A:22:ASN:ND2	1:C:260:THR:OG1[4_555]	2.00	0.20

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	A	475/510 (93%)	371 (78%)	87 (18%)	17 (4%)	3 26
1	B	474/510 (93%)	365 (77%)	94 (20%)	15 (3%)	4 29
1	C	473/510 (93%)	374 (79%)	84 (18%)	15 (3%)	4 29
1	D	472/510 (92%)	377 (80%)	87 (18%)	8 (2%)	9 42
All	All	1894/2040 (93%)	1487 (78%)	352 (19%)	55 (3%)	4 31

5 of 55 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	98	ASN
1	A	201	ALA
1	A	432	ALA
1	B	403	CYS
1	B	502	VAL

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	A	383/411 (93%)	340 (89%)	43 (11%)	6 27
1	B	382/411 (93%)	341 (89%)	41 (11%)	6 30
1	C	381/411 (93%)	337 (88%)	44 (12%)	5 26
1	D	380/411 (92%)	335 (88%)	45 (12%)	5 25
All	All	1526/1644 (93%)	1353 (89%)	173 (11%)	6 27

5 of 173 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	395	SER
1	D	178	LEU
1	C	446	THR
1	D	66	MET
1	D	346	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	43	GLN
1	D	453	ASN
1	D	70	ASN
1	D	329	ASN
1	B	213	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 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	A	479/510 (93%)	-0.06	4 (0%)	86	81	24, 65, 121, 227	0
1	B	478/510 (93%)	0.24	10 (2%)	63	58	20, 82, 162, 247	0
1	C	477/510 (93%)	0.12	9 (1%)	66	61	25, 80, 148, 263	0
1	D	476/510 (93%)	0.51	41 (8%)	10	11	26, 116, 254, 407	0
All	All	1910/2040 (93%)	0.20	64 (3%)	45	40	20, 84, 186, 407	0

The worst 5 of 64 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	292	SER	9.6
1	D	325	ALA	6.6
1	B	271	TYR	5.9
1	B	484	VAL	5.2
1	D	484	VAL	5.1

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 (i)

There are no such residues in this entry.