wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2016 – 09:24 PM GMT

PDB ID	:	4ZWJ
Title	:	Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser
Authors	:	Kang, Y.; Zhou, X.E.; Gao, X.; He, Y.; Liu, W.; Ishchenko, A.; Barty, A.; White, T.A.; Yefanov, O.; Han, G.W.; Xu, Q.; de Waal, P.W.; Ke, J.; Tan, M.H.E.; Zhang, C.; Moeller, A.; West, G.M.; Pascal, B.; Eps, N.V.; Caro, L.N.; Vishnivetskiy, S.A.; Lee, R.J.; Suino-Powell, K.M.; Gu, X.; Pal, K.; Ma, J.; Zhi, X.; Boutet, S.; Williams, G.J.; Messerschmidt, M.; Gati, C.; Zatsepin, N.A.; Wang, D.; James, D.; Basu, S.; Roy-Chowdhury, S.; Conrad, C.; Coe, J.; Liu, H.; Lisova, S.; Kupitz, C.; Grotjohann, I.; Fromme, R.; Jiang, Y.; Tan, M.; Yang, H.; Li, J.; Wang, M.; Zheng, Z.; Li, D.; Howe, N.; Zhao, Y.; Standfuss, J.; Diederichs, K.; Dong, Y.; Potter, C.S.; Carragher, B.; Caffrey, M.; Jiang, H.; Chapman, H.N.; Spence, J.C.H.; Fromme, P.; Weierstall, U.; Ernst, O.P.; Katritch, V.; Gurevich, V.V.; Griffin, P.R.; Hubbell, W.L.; Stevens, R.C.; Cherezov, V.; Melcher, K.; Xu, H.E.; GPCR Network (GPCR)
Deposited on	:	2015-05-19
Resolution	:	3.30 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

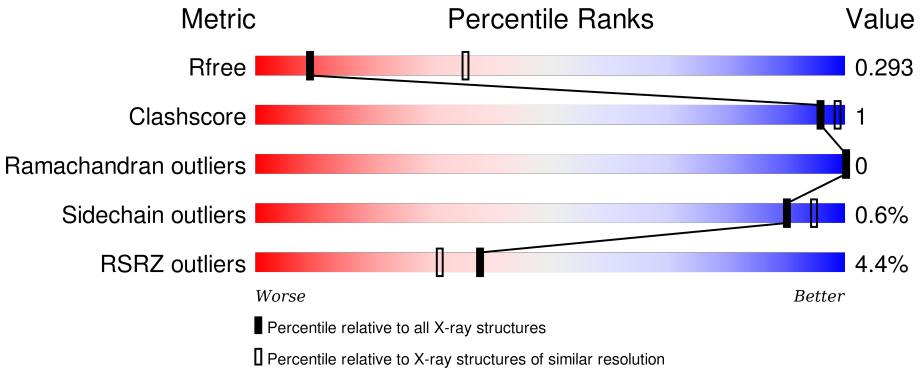
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026982
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135

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 3.30 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}	91344	2060 (3.40-3.20)
Clashscore	102246	1058 (3.38-3.22)
Ramachandran outliers	100387	1038 (3.38-3.22)
Sidechain outliers	100360	1037 (3.38-3.22)
RSRZ outliers	91569	2070 (3.40-3.20)

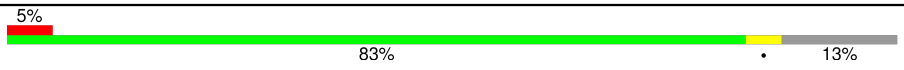
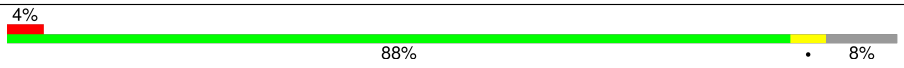
The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey 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% 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	906	4% 86% 5% 8%
1	B	906	2% 71% 26%

Continued on next page...

CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20026982

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	C	906	
1	D	906	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 24665 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 Chimera protein of human Rhodopsin, mouse S-arrestin, and T4 Endolysin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	833	Total	C	N	O	S	0	0	0
			6565	4264	1076	1187	38			
1	B	673	Total	C	N	O	S	0	0	0
			5296	3463	847	951	35			
1	C	789	Total	C	N	O	S	0	0	0
			6231	4052	1019	1121	39			
1	D	833	Total	C	N	O	S	0	0	0
			6573	4269	1079	1186	39			

There are 104 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1012	GLY	ARG	engineered mutation	UNP P00720
A	1054	THR	CYS	engineered mutation	UNP P00720
A	1097	ALA	CYS	engineered mutation	UNP P00720
A	1137	ARG	ILE	engineered mutation	UNP P00720
A	2	CYS	ASN	engineered mutation	UNP P08100
A	113	GLN	GLU	engineered mutation	UNP P08100
A	257	TYR	MET	engineered mutation	UNP P08100
A	282	CYS	ASN	engineered mutation	UNP P08100
A	1995	ALA	-	linker	UNP P08100
A	1996	ALA	-	linker	UNP P08100
A	1997	ALA	-	linker	UNP P08100
A	1998	GLY	-	linker	UNP P08100
A	1999	SER	-	linker	UNP P08100
A	2000	ALA	-	linker	UNP P08100
A	2001	GLY	-	linker	UNP P08100
A	2002	SER	-	linker	UNP P08100
A	2003	ALA	-	linker	UNP P08100
A	2004	GLY	-	linker	UNP P08100
A	2005	SER	-	linker	UNP P08100
A	2006	ALA	-	linker	UNP P08100

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	2007	GLY	-	linker	UNP P08100
A	2008	SER	-	linker	UNP P08100
A	2009	ALA	-	linker	UNP P08100
A	2374	ALA	LEU	engineered mutation	UNP P20443
A	2375	ALA	VAL	engineered mutation	UNP P20443
A	2376	ALA	PHE	engineered mutation	UNP P20443
B	1012	GLY	ARG	engineered mutation	UNP P00720
B	1054	THR	CYS	engineered mutation	UNP P00720
B	1097	ALA	CYS	engineered mutation	UNP P00720
B	1137	ARG	ILE	engineered mutation	UNP P00720
B	2	CYS	ASN	engineered mutation	UNP P08100
B	113	GLN	GLU	engineered mutation	UNP P08100
B	257	TYR	MET	engineered mutation	UNP P08100
B	282	CYS	ASN	engineered mutation	UNP P08100
B	1995	ALA	-	linker	UNP P08100
B	1996	ALA	-	linker	UNP P08100
B	1997	ALA	-	linker	UNP P08100
B	1998	GLY	-	linker	UNP P08100
B	1999	SER	-	linker	UNP P08100
B	2000	ALA	-	linker	UNP P08100
B	2001	GLY	-	linker	UNP P08100
B	2002	SER	-	linker	UNP P08100
B	2003	ALA	-	linker	UNP P08100
B	2004	GLY	-	linker	UNP P08100
B	2005	SER	-	linker	UNP P08100
B	2006	ALA	-	linker	UNP P08100
B	2007	GLY	-	linker	UNP P08100
B	2008	SER	-	linker	UNP P08100
B	2009	ALA	-	linker	UNP P08100
B	2374	ALA	LEU	engineered mutation	UNP P20443
B	2375	ALA	VAL	engineered mutation	UNP P20443
B	2376	ALA	PHE	engineered mutation	UNP P20443
C	1012	GLY	ARG	engineered mutation	UNP P00720
C	1054	THR	CYS	engineered mutation	UNP P00720
C	1097	ALA	CYS	engineered mutation	UNP P00720
C	1137	ARG	ILE	engineered mutation	UNP P00720
C	2	CYS	ASN	engineered mutation	UNP P08100
C	113	GLN	GLU	engineered mutation	UNP P08100
C	257	TYR	MET	engineered mutation	UNP P08100
C	282	CYS	ASN	engineered mutation	UNP P08100
C	1995	ALA	-	linker	UNP P08100
C	1996	ALA	-	linker	UNP P08100

Continued on next page...

Continued from previous page...

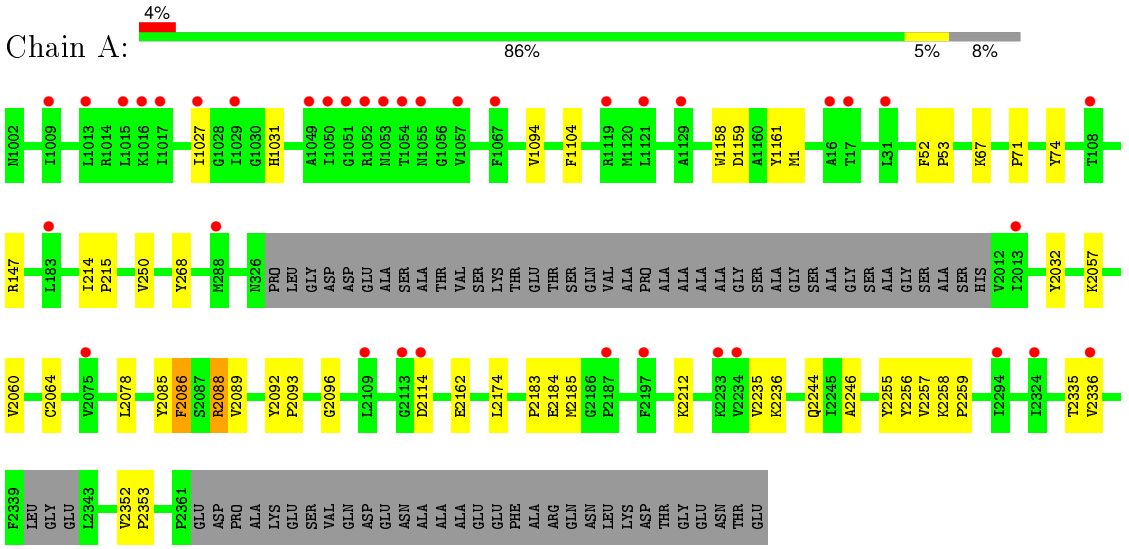
Chain	Residue	Modelled	Actual	Comment	Reference
C	1997	ALA	-	linker	UNP P08100
C	1998	GLY	-	linker	UNP P08100
C	1999	SER	-	linker	UNP P08100
C	2000	ALA	-	linker	UNP P08100
C	2001	GLY	-	linker	UNP P08100
C	2002	SER	-	linker	UNP P08100
C	2003	ALA	-	linker	UNP P08100
C	2004	GLY	-	linker	UNP P08100
C	2005	SER	-	linker	UNP P08100
C	2006	ALA	-	linker	UNP P08100
C	2007	GLY	-	linker	UNP P08100
C	2008	SER	-	linker	UNP P08100
C	2009	ALA	-	linker	UNP P08100
C	2374	ALA	LEU	engineered mutation	UNP P20443
C	2375	ALA	VAL	engineered mutation	UNP P20443
C	2376	ALA	PHE	engineered mutation	UNP P20443
D	1012	GLY	ARG	engineered mutation	UNP P00720
D	1054	THR	CYS	engineered mutation	UNP P00720
D	1097	ALA	CYS	engineered mutation	UNP P00720
D	1137	ARG	ILE	engineered mutation	UNP P00720
D	2	CYS	ASN	engineered mutation	UNP P08100
D	113	GLN	GLU	engineered mutation	UNP P08100
D	257	TYR	MET	engineered mutation	UNP P08100
D	282	CYS	ASN	engineered mutation	UNP P08100
D	1995	ALA	-	linker	UNP P08100
D	1996	ALA	-	linker	UNP P08100
D	1997	ALA	-	linker	UNP P08100
D	1998	GLY	-	linker	UNP P08100
D	1999	SER	-	linker	UNP P08100
D	2000	ALA	-	linker	UNP P08100
D	2001	GLY	-	linker	UNP P08100
D	2002	SER	-	linker	UNP P08100
D	2003	ALA	-	linker	UNP P08100
D	2004	GLY	-	linker	UNP P08100
D	2005	SER	-	linker	UNP P08100
D	2006	ALA	-	linker	UNP P08100
D	2007	GLY	-	linker	UNP P08100
D	2008	SER	-	linker	UNP P08100
D	2009	ALA	-	linker	UNP P08100
D	2374	ALA	LEU	engineered mutation	UNP P20443
D	2375	ALA	VAL	engineered mutation	UNP P20443
D	2376	ALA	PHE	engineered mutation	UNP P20443

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 errors 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.

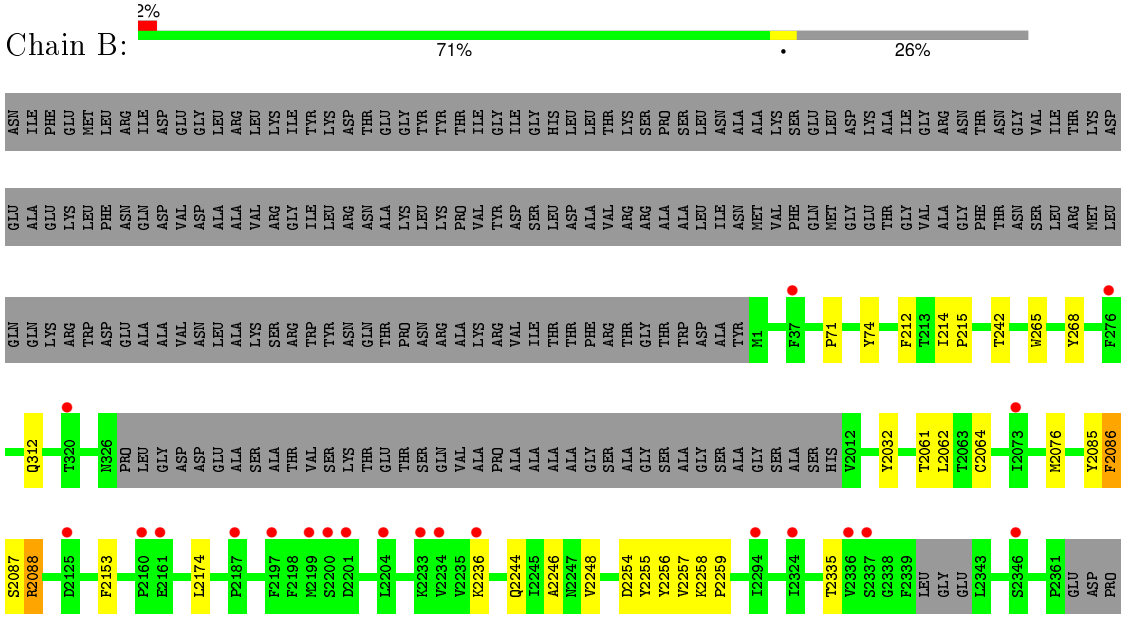
• Molecule 1: Chimera protein of human Rhodopsin, mouse S-arrestin, and T4 Endolysin

Chain A:



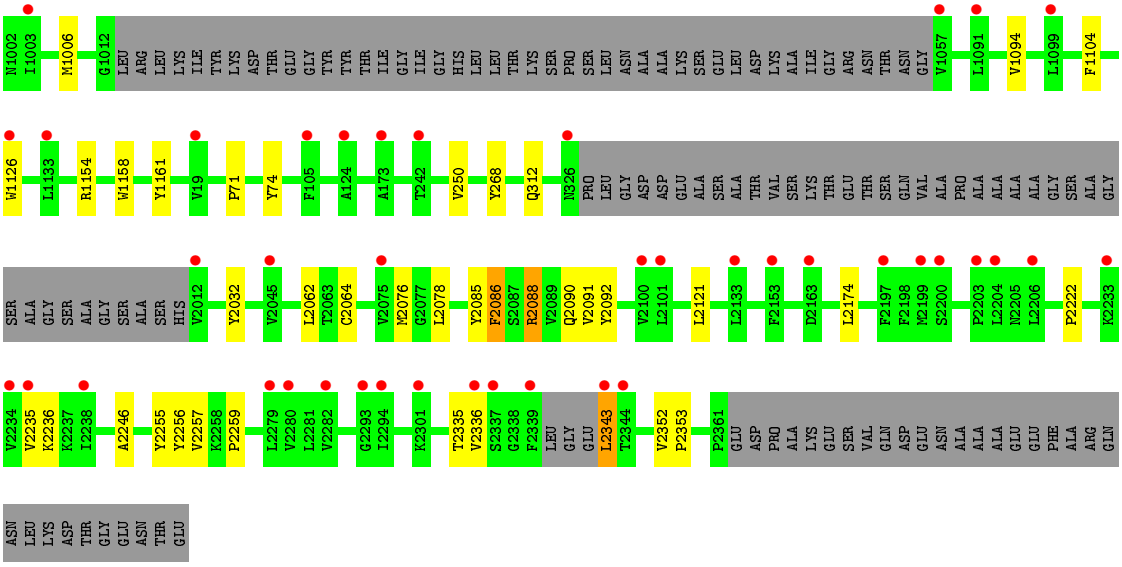
• Molecule 1: Chimera protein of human Rhodopsin, mouse S-arrestin, and T4 Endolysin

Chain B:



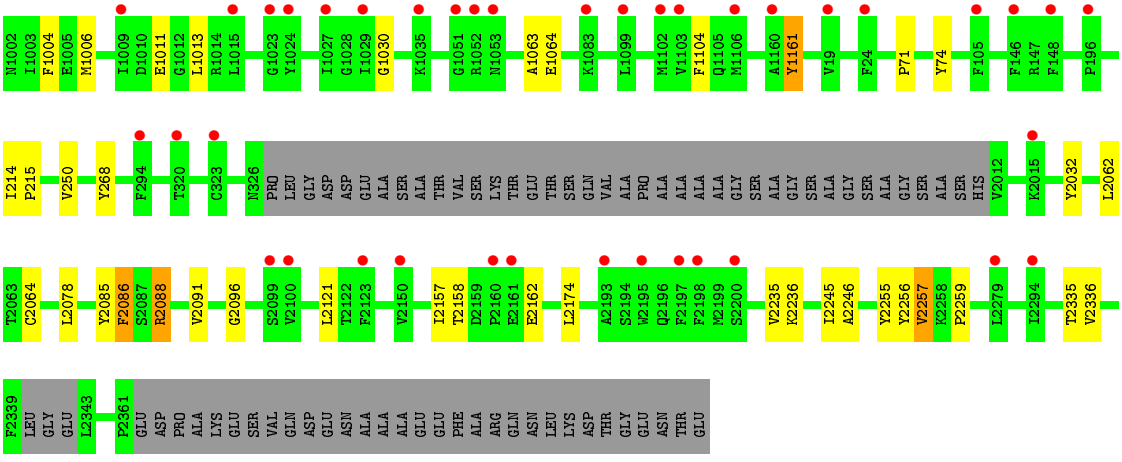
• Molecule 1: Chimera protein of human Rhodopsin, mouse S-arrestin, and T4 Endolysin

Chain C:



• Molecule 1: Chimera protein of human Rhodopsin, mouse S-arrestin, and T4 Endolysin

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	109.24Å 109.24Å 452.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	31.04 – 3.30 31.04 – 3.30	Depositor EDS
% Data completeness (in resolution range)	76.1 (31.04-3.30) 76.1 (31.04-3.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.19	Depositor
$< I/\sigma(I) >$ [1]	2.05 (at 3.31Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.253 , 0.293 0.253 , 0.293	Depositor DCC
R_{free} test set	3099 reflections (4.95%)	DCC
Wilson B-factor (Å^2)	96.5	Xtriage
Anisotropy	0.663	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 81.1	EDS
Estimated twinning fraction	0.500 for k,h,-l 0.499 for k,h,-l	Xtriage
Reported twinning fraction	0.500 for k,h,-l	Depositor
L-test for twinning[2]	$< \|L\| > = 0.37$, $< L^2 > = 0.20$	Xtriage
Outliers	0 of 62613 reflections	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	24665	wwPDB-VP
Average B, all atoms (Å^2)	159.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.74% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.42	1/6723 (0.0%)	0.52	1/9138 (0.0%)
1	B	0.44	2/5433 (0.0%)	0.54	1/7395 (0.0%)
1	C	0.42	1/6383 (0.0%)	0.54	2/8676 (0.0%)
1	D	0.42	3/6731 (0.0%)	0.53	1/9146 (0.0%)
All	All	0.42	7/25270 (0.0%)	0.53	5/34355 (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
1	A	0	6
1	B	0	3
1	C	0	6
1	D	0	4
All	All	0	19

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	2086	PHE	C-O	-7.97	1.08	1.23
1	B	2086	PHE	C-O	-7.42	1.09	1.23
1	A	2086	PHE	C-O	-6.96	1.10	1.23
1	B	2088	ARG	C-O	-6.82	1.10	1.23
1	C	2086	PHE	C-O	-6.36	1.11	1.23

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	2088	ARG	NE-CZ-NH1	-8.36	116.12	120.30
1	A	2032	TYR	O-C-N	5.09	130.84	122.70
1	C	2032	TYR	O-C-N	5.07	130.80	122.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	2032	TYR	O-C-N	5.06	130.79	122.70
1	B	2032	TYR	O-C-N	5.01	130.72	122.70

There are no chirality outliers.

5 of 19 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1161	TYR	Mainchain
1	A	2088	ARG	Mainchain
1	A	2096	GLY	Mainchain
1	A	2255	TYR	Mainchain
1	A	2257	VAL	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	A	6565	0	6623	25	0
1	B	5296	0	5340	14	0
1	C	6231	0	6281	20	0
1	D	6573	0	6644	16	0
All	All	24665	0	24888	73	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 1.

The worst 5 of 73 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:2184:GLU:OE1	1:A:2184:GLU:N	1.99	0.94
1:A:2246:ALA:O	1:A:2256:TYR:N	2.29	0.66
1:A:2246:ALA:N	1:A:2256:TYR:O	2.31	0.61
1:A:2064:CYS:N	1:A:2085:TYR:O	2.33	0.61
1:A:2183:PRO:HG2	1:A:2184:GLU:OE1	2.00	0.60

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	827/906 (91%)	796 (96%)	31 (4%)	0	100	100
1	B	667/906 (74%)	641 (96%)	26 (4%)	0	100	100
1	C	781/906 (86%)	753 (96%)	28 (4%)	0	100	100
1	D	827/906 (91%)	797 (96%)	30 (4%)	0	100	100
All	All	3102/3624 (86%)	2987 (96%)	115 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	721/777 (93%)	716 (99%)	5 (1%)	88	94
1	B	589/777 (76%)	586 (100%)	3 (0%)	92	95
1	C	686/777 (88%)	682 (99%)	4 (1%)	90	95
1	D	723/777 (93%)	719 (99%)	4 (1%)	90	95
All	All	2719/3108 (88%)	2703 (99%)	16 (1%)	90	95

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	2174	LEU
1	C	1104	PHE
1	D	1104	PHE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	268	TYR
1	D	268	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1132	ASN
1	D	1132	ASN

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 carbohydrates 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, 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	833/906 (91%)	0.23	37 (4%)	38	31	113, 153, 255, 341	0
1	B	673/906 (74%)	0.18	21 (3%)	52	46	85, 131, 194, 288	0
1	C	789/906 (87%)	0.24	41 (5%)	31	25	104, 155, 228, 267	0
1	D	833/906 (91%)	0.22	39 (4%)	35	29	101, 156, 244, 313	0
All	All	3128/3624 (86%)	0.22	138 (4%)	38	31	85, 149, 234, 341	0

The worst 5 of 138 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1029	ILE	10.6
1	A	1015	LEU	8.9
1	A	1017	ILE	8.5
1	A	1053	ASN	8.5
1	A	1055	ASN	8.1

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 carbohydrates 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.