wwPDB X-ray Structure Validation Summary Report (i)

Aug 21, 2017 – 01:54 AM EDT

PDB ID	:	5MSU
Title	:	Structure of the R domain of carboxylic acid reductase (CAR) from Mycobacterium marinum in complex with NADP, P21 form
Authors	:	Gahloth, D.; Leys, D.
Deposited on	:	unknown
Resolution	:	1.74 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:

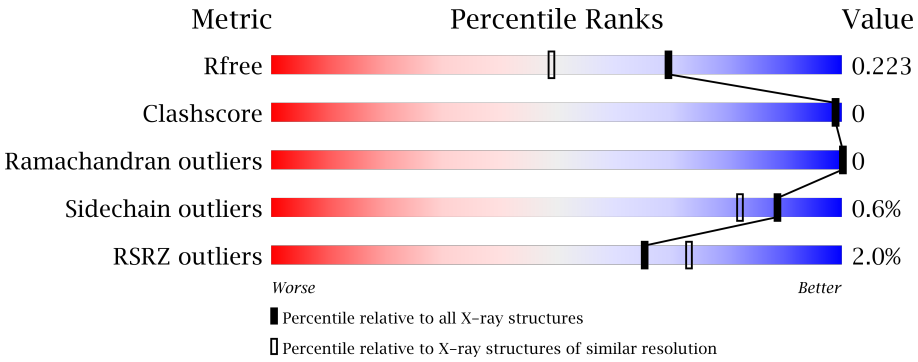
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20029824

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 1.74 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}	100719	2694 (1.76-1.72)
Clashscore	112137	2854 (1.76-1.72)
Ramachandran outliers	110173	2824 (1.76-1.72)
Sidechain outliers	110143	2824 (1.76-1.72)
RSRZ outliers	101464	2705 (1.76-1.72)

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	1174	.% 36% · 63%
1	B	1174	38% · 61%
1	C	1174	.% 36% · 63%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	NAP	B	1201	-	-	-	X

2 Entry composition

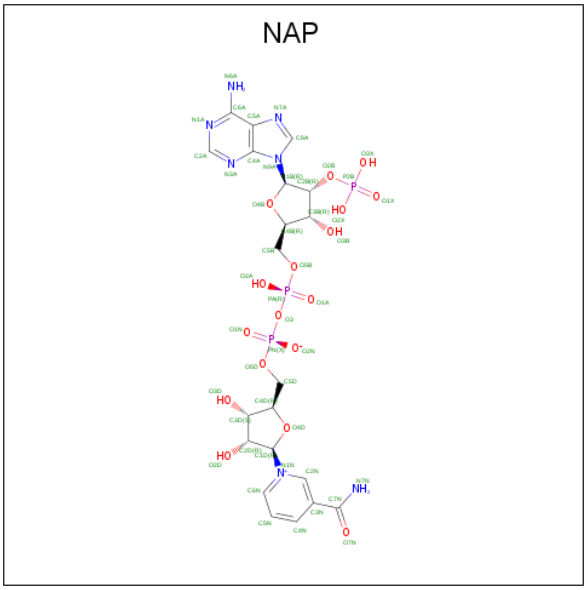
There are 3 unique types of molecules in this entry. The entry contains 11491 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Carboxylic acid reductase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	434	Total	C	N	O	S	0	0	0
			3354	2125	588	632	9			
1	A	436	Total	C	N	O	S	0	0	0
			3371	2134	590	638	9			
1	B	459	Total	C	N	O	S	0	0	0
			3527	2227	617	674	9			

- Molecule 2 is NADP NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE (three-letter code: NAP) (formula: $C_{21}H_{28}N_7O_{17}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	C	1	Total	C	N	O	P	0	0
			48	21	7	17	3		
2	A	1	Total	C	N	O	P	0	0
			48	21	7	17	3		
2	B	1	Total	C	N	O	P	0	0
			39	15	5	16	3		

- Molecule 3 is water.

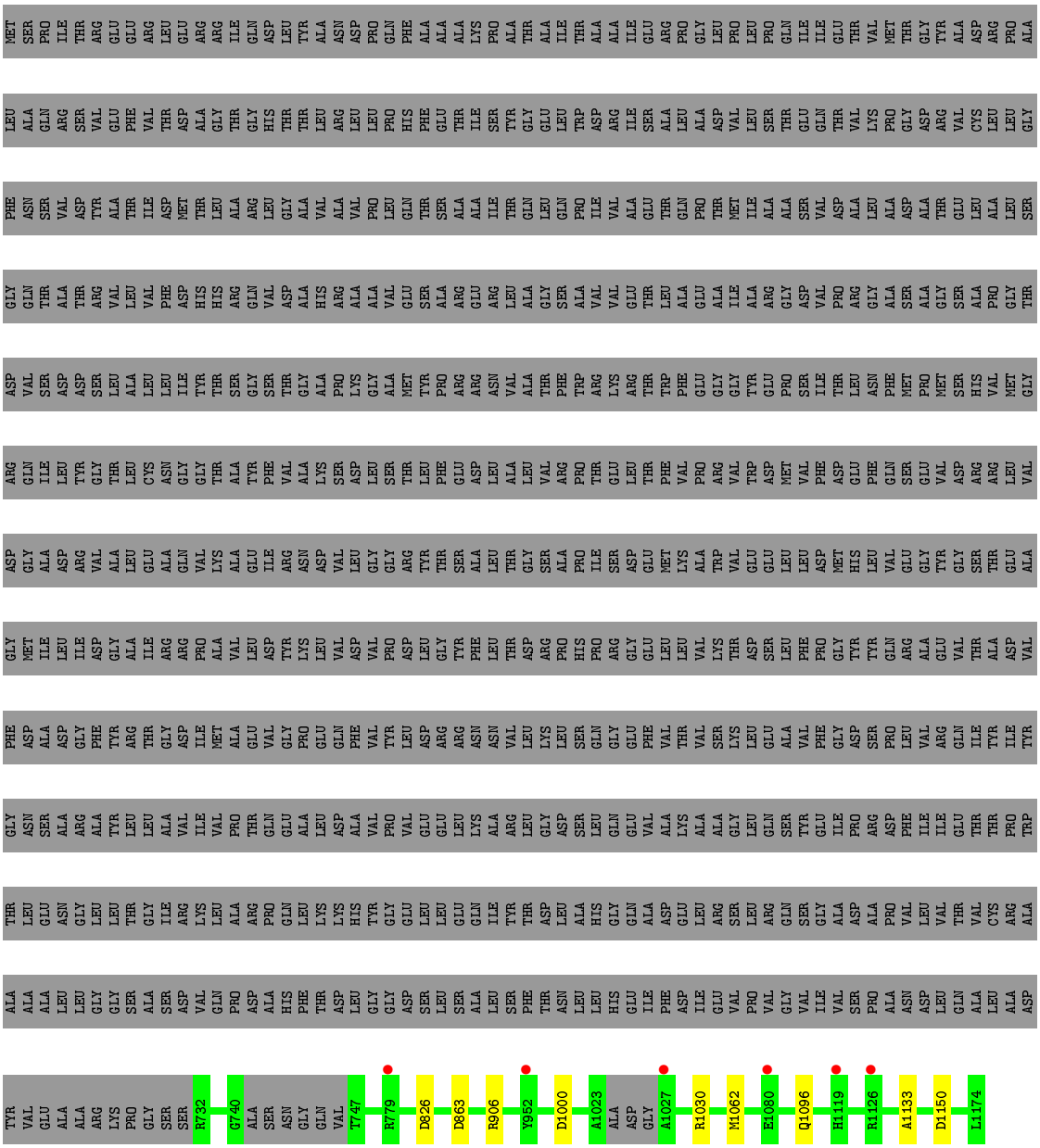
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	370	Total O 370 370	0	0
3	A	310	Total O 310 310	0	0
3	B	424	Total O 424 424	0	0

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 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.

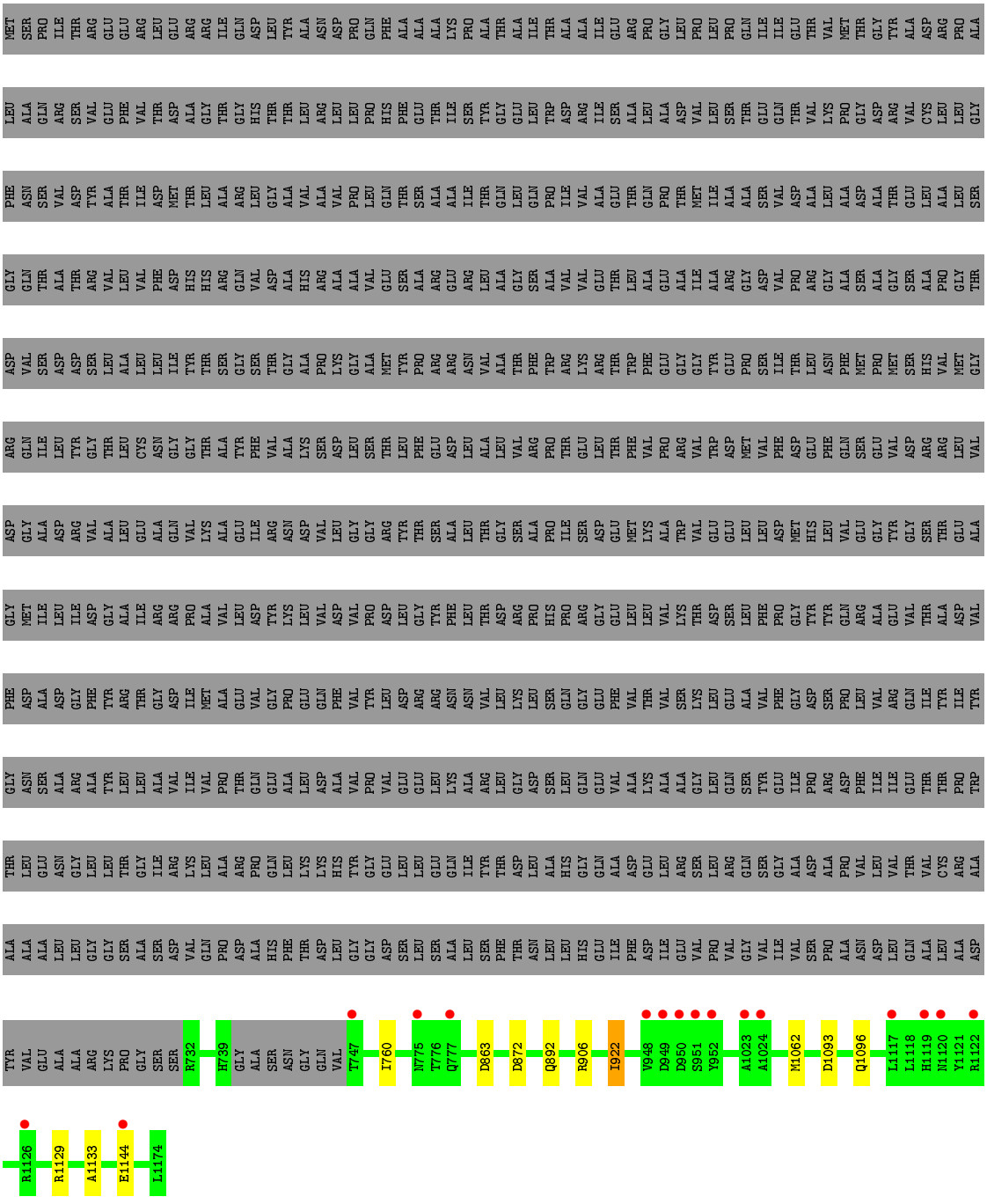
- Molecule 1: Carboxylic acid reductase

Chain C:



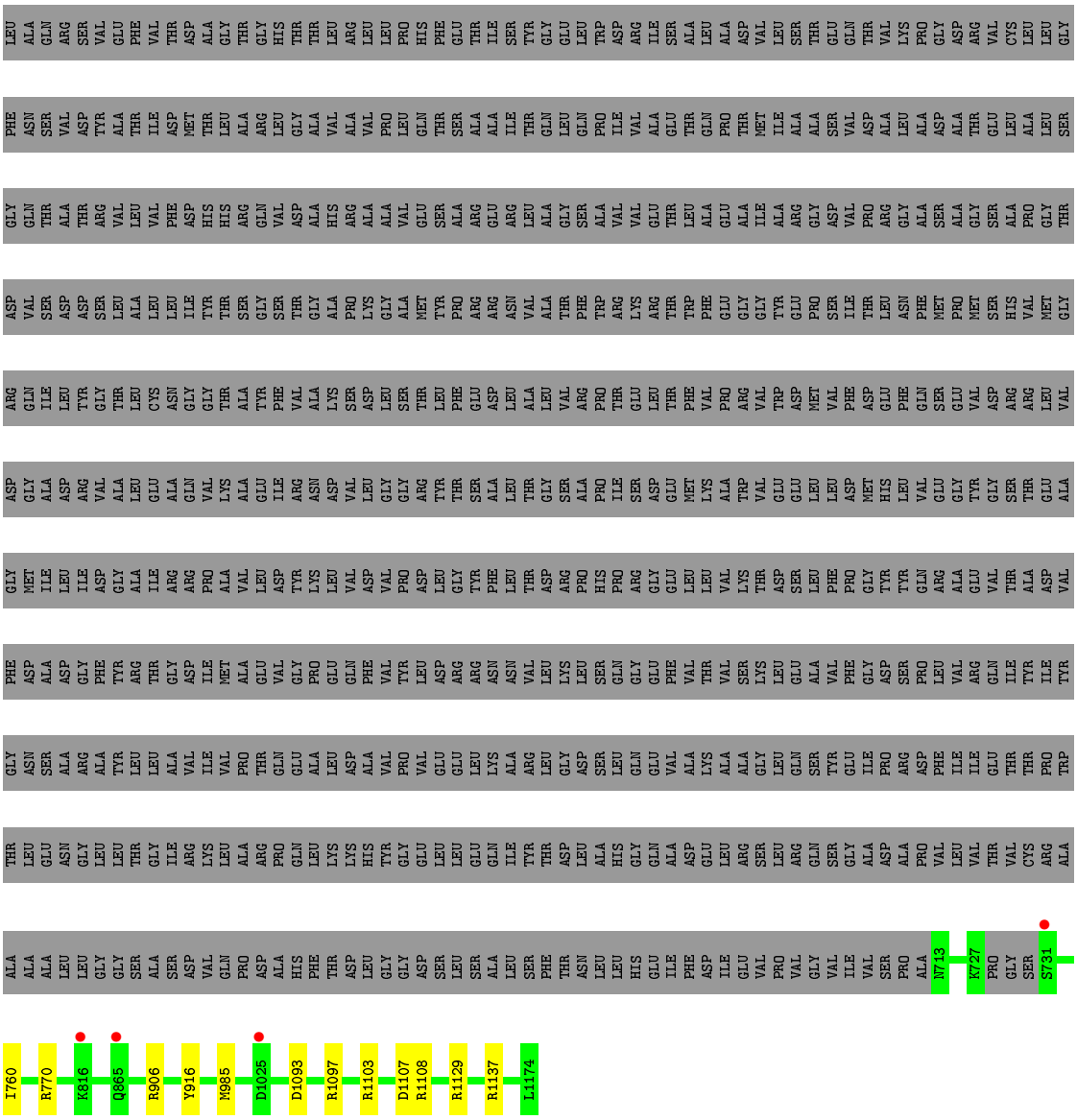
• Molecule 1: Carboxylic acid reductase

Chain A:



• Molecule 1: Carboxylic acid reductase

Chain B:


I760
R770
K816
Q865
R906
Y916
M985
D1025
D1093
R1097
R1103
D1107
R1108
R1129
R1137
L1174
N713
K727
S731

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	100.54Å 58.95Å 135.56Å 90.00° 105.92° 90.00°	Depositor
Resolution (Å)	130.36 – 1.74 50.33 – 1.74	Depositor EDS
% Data completeness (in resolution range)	99.7 (130.36-1.74) 99.8 (50.33-1.74)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.75 (at 1.74Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.188 , 0.213 0.202 , 0.223	Depositor DCC
R_{free} test set	7843 reflections (5.28%)	DCC
Wilson B-factor (Å^2)	24.8	Xtriage
Anisotropy	0.032	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 35.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	11491	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.67% 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: NAP

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	0.44	0/3445	0.72	1/4694 (0.0%)
1	B	0.47	0/3602	0.76	7/4909 (0.1%)
1	C	0.44	0/3427	0.75	7/4667 (0.1%)
All	All	0.45	0/10474	0.74	15/14270 (0.1%)

There are no bond length outliers.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	1108	ARG	NE-CZ-NH1	7.00	123.80	120.30
1	C	906	ARG	NE-CZ-NH1	6.62	123.61	120.30
1	C	906	ARG	NE-CZ-NH2	-6.57	117.02	120.30
1	B	985	MET	CG-SD-CE	6.20	110.12	100.20
1	C	1150	ASP	CB-CG-OD1	5.88	123.59	118.30

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	3371	0	3326	6	0
1	B	3527	0	3462	4	0
1	C	3354	0	3313	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	48	0	25	0	0
2	B	39	0	18	0	0
2	C	48	0	25	0	0
3	A	310	0	0	1	0
3	B	424	0	0	1	0
3	C	370	0	0	0	0
All	All	11491	0	10169	10	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 0.

The worst 5 of 10 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:1129:ARG:NH2	3:A:1701:HOH:O	1.97	0.81
1:A:922:ILE:HD13	1:A:922:ILE:O	2.02	0.59
1:A:872:ASP:OD2	1:B:770:ARG:HD2	2.04	0.57
1:B:1093:ASP:OD2	1:B:1097:ARG:NH2	2.44	0.50
1:A:1062:MET:SD	1:A:1133:ALA:HB3	2.56	0.46

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	A	432/1174 (37%)	426 (99%)	6 (1%)	0	100	100
1	B	455/1174 (39%)	447 (98%)	8 (2%)	0	100	100
1	C	428/1174 (36%)	422 (99%)	6 (1%)	0	100	100
All	All	1315/3522 (37%)	1295 (98%)	20 (2%)	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	352/946 (37%)	349 (99%)	3 (1%)	82 71
1	B	366/946 (39%)	365 (100%)	1 (0%)	94 91
1	C	350/946 (37%)	348 (99%)	2 (1%)	89 82
All	All	1068/2838 (38%)	1062 (99%)	6 (1%)	89 82

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	892	GLN
1	B	916	TYR
1	A	922	ILE
1	C	1096	GLN
1	A	1144	GLU

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

Mol	Chain	Res	Type
1	C	971	HIS
1	B	1143	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 (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	NAP	A	1601	-	44,52,52	1.52	3 (6%)	51,80,80	1.87	8 (15%)
2	NAP	B	1201	-	36,42,52	1.01	2 (5%)	40,65,80	1.50	4 (10%)
2	NAP	C	1201	-	44,52,52	1.34	2 (4%)	51,80,80	1.86	7 (13%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	NAP	A	1601	-	-	0/27/67/67	0/5/5/5
2	NAP	B	1201	-	-	0/23/56/67	0/4/4/5
2	NAP	C	1201	-	-	0/27/67/67	0/5/5/5

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1201	NAP	C2A-N3A	2.19	1.35	1.32
2	C	1201	NAP	C5N-C4N	2.81	1.44	1.38
2	A	1601	NAP	C5A-C4A	2.81	1.46	1.40
2	B	1201	NAP	C5A-C4A	3.23	1.47	1.40
2	A	1601	NAP	C5N-C4N	3.51	1.45	1.38

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	1201	NAP	N3A-C2A-N1A	-7.77	122.09	128.86
2	A	1601	NAP	N3A-C2A-N1A	-7.22	122.57	128.86
2	B	1201	NAP	N3A-C2A-N1A	-6.39	123.29	128.86
2	C	1201	NAP	C5N-C4N-C3N	-5.91	113.40	120.35
2	A	1601	NAP	C5N-C4N-C3N	-5.89	113.42	120.35

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

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	436/1174 (37%)	0.36	16 (3%) 42 49	20, 43, 75, 104	0
1	B	459/1174 (39%)	0.20	4 (0%) 84 89	19, 40, 67, 85	0
1	C	434/1174 (36%)	0.14	6 (1%) 75 82	18, 36, 64, 82	0
All	All	1329/3522 (37%)	0.23	26 (1%) 65 73	18, 39, 70, 104	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	952	TYR	5.4
1	A	950	ASP	4.1
1	A	1024	ALA	4.1
1	A	777	GLN	3.5
1	B	731	SER	3.4

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
2	NAP	B	1201	39/48	0.89	0.19	3.82	23,30,35,39	8
2	NAP	C	1201	48/48	0.98	0.06	-1.13	16,18,21,24	0
2	NAP	A	1601	48/48	0.97	0.07	-1.21	17,21,24,26	0

6.5 Other polymers (i)

There are no such residues in this entry.