



Full wwPDB X-ray Structure Validation Report ⓘ

Jul 18, 2017 – 07:36 PM EDT

PDB ID : 3MLG
Title : 2ouf-2x, a designed knotted protein
Authors : King, N.P.; Sawaya, M.R.; Jacobitz, A.W.; Yeates, T.O.
Deposited on : unknown
Resolution : 2.29 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ symbol.

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

MolProbity : 4.02b-467
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

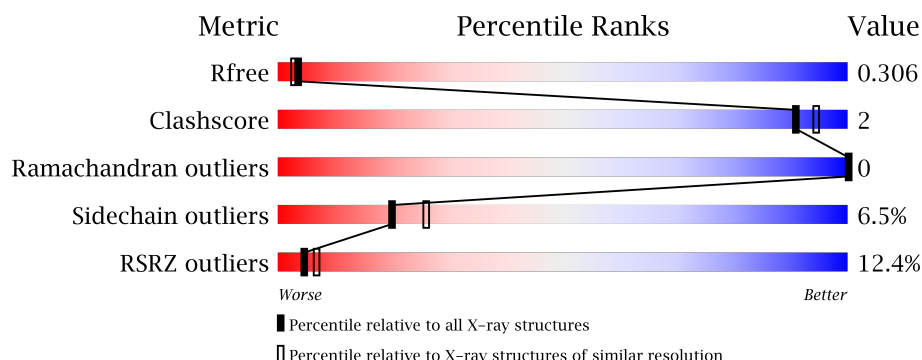
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.29 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	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)

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 $\leq 5\%$. 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	189	<div> <div>10%</div> <div> <div></div> <div>77%</div> <div>8%</div> <div>15%</div> </div> </div>
1	B	189	<div> <div>11%</div> <div> <div></div> <div>75%</div> <div>10%</div> <div>15%</div> </div> </div>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 2570 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 2X chimera of Helicobacter pylori protein HP0242.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	161	Total	C	N	O	S	0	0	0
			1266	811	205	248	2			
1	B	161	Total	C	N	O	S	0	0	0
			1254	803	202	247	2			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	93	SER	-	linker	UNP O25025
A	94	GLY	-	linker	UNP O25025
A	95	SER	-	linker	UNP O25025
A	96	GLY	-	linker	UNP O25025
A	97	SER	-	linker	UNP O25025
A	98	GLY	-	linker	UNP O25025
A	99	SER	-	linker	UNP O25025
A	100	SER	-	linker	UNP O25025
A	182	LEU	-	expression tag	UNP O25025
A	183	GLU	-	expression tag	UNP O25025
A	184	HIS	-	expression tag	UNP O25025
A	185	HIS	-	expression tag	UNP O25025
A	186	HIS	-	expression tag	UNP O25025
A	187	HIS	-	expression tag	UNP O25025
A	188	HIS	-	expression tag	UNP O25025
A	189	HIS	-	expression tag	UNP O25025
B	93	SER	-	linker	UNP O25025
B	94	GLY	-	linker	UNP O25025
B	95	SER	-	linker	UNP O25025
B	96	GLY	-	linker	UNP O25025
B	97	SER	-	linker	UNP O25025
B	98	GLY	-	linker	UNP O25025
B	99	SER	-	linker	UNP O25025
B	100	SER	-	linker	UNP O25025
B	182	LEU	-	expression tag	UNP O25025

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	183	GLU	-	expression tag	UNP O25025
B	184	HIS	-	expression tag	UNP O25025
B	185	HIS	-	expression tag	UNP O25025
B	186	HIS	-	expression tag	UNP O25025
B	187	HIS	-	expression tag	UNP O25025
B	188	HIS	-	expression tag	UNP O25025
B	189	HIS	-	expression tag	UNP O25025

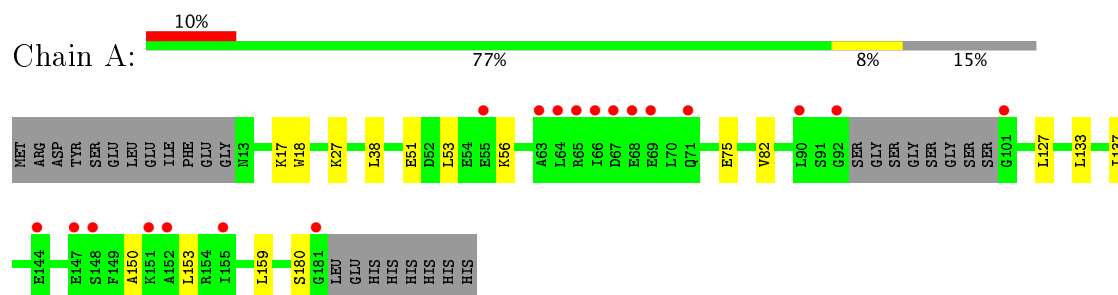
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	26	Total O 26 26	0	0
2	B	24	Total O 24 24	0	0

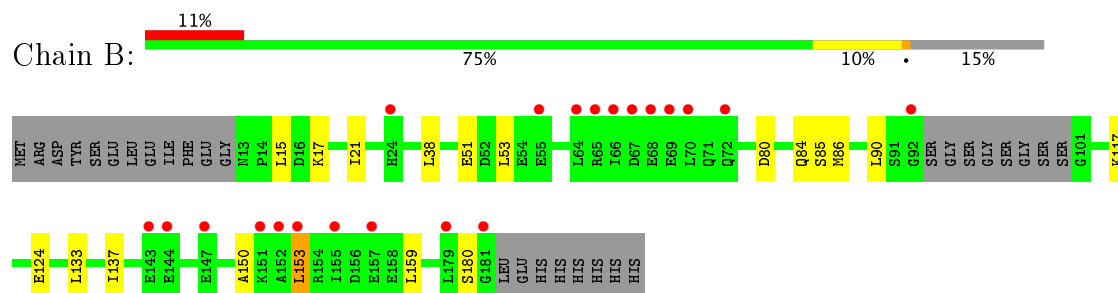
3 Residue-property plots [i](#)

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: 2X chimera of *Helicobacter pylori* protein HP0242



- Molecule 1: 2X chimera of *Helicobacter pylori* protein HP0242



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	44.94Å 75.39Å 118.57Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	21.99 – 2.29 21.93 – 2.29	Depositor EDS
% Data completeness (in resolution range)	(Not available) (21.99-2.29) 92.0 (21.93-2.29)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.46 (at 2.28Å)	Xtriage
Refinement program	BUSTER-TNT	Depositor
R, R_{free}	0.248 , 0.278 0.261 , 0.306	Depositor DCC
R_{free} test set	873 reflections (5.33%)	DCC
Wilson B-factor (Å ²)	26.2	Xtriage
Anisotropy	0.042	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 47.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.52$, $\langle L^2 \rangle = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	2570	wwPDB-VP
Average B, all atoms (Å ²)	51.0	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 43.83 % 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 1.6673e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.50	0/1280	0.65	0/1717
1	B	0.49	0/1268	0.64	0/1708
All	All	0.49	0/2548	0.65	0/3425

There are no bond length outliers.

There are no bond angle outliers.

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	1266	0	1255	5	0
1	B	1254	0	1222	5	0
2	A	26	0	0	0	0
2	B	24	0	0	0	0
All	All	2570	0	2477	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 2.

All (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:B:15:LEU:HD22	1:B:124:GLU:HG2	1.71	0.70

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:51:GLU:HB3	1:B:53:LEU:HD12	1.81	0.62
1:A:51:GLU:HB3	1:A:53:LEU:HD12	1.81	0.61
1:B:133:LEU:O	1:B:137:ILE:HG12	2.12	0.49
1:A:133:LEU:O	1:A:137:ILE:HG12	2.14	0.47
1:B:21:ILE:HG21	1:B:86:MET:HG2	1.97	0.46
1:A:18:TRP:HB2	1:A:82:VAL:HG13	1.98	0.46
1:A:150:ALA:HA	1:A:153:LEU:HB2	1.98	0.46
1:A:82:VAL:HG23	1:A:127:LEU:HD21	1.97	0.46
1:B:150:ALA:HA	1:B:153:LEU:HB2	1.99	0.45

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	157/189 (83%)	156 (99%)	1 (1%)	0	100	100
1	B	157/189 (83%)	156 (99%)	1 (1%)	0	100	100
All	All	314/378 (83%)	312 (99%)	2 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	133/172 (77%)	126 (95%)	7 (5%)	26	35
1	B	130/172 (76%)	120 (92%)	10 (8%)	15	18
All	All	263/344 (76%)	246 (94%)	17 (6%)	20	26

All (17) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	17	LYS
1	A	27	LYS
1	A	38	LEU
1	A	56	LYS
1	A	75	GLU
1	A	159	LEU
1	A	180	SER
1	B	17	LYS
1	B	38	LEU
1	B	80	ASP
1	B	84	GLN
1	B	85	SER
1	B	90	LEU
1	B	117	LYS
1	B	153	LEU
1	B	159	LEU
1	B	180	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	84	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](#)

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, 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	161/189 (85%)	0.54	19 (11%) 5 7	15, 45, 94, 111	0
1	B	161/189 (85%)	0.58	21 (13%) 4 5	17, 47, 108, 126	0
All	All	322/378 (85%)	0.56	40 (12%) 4 7	15, 47, 100, 126	0

All (40) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	152	ALA	6.7
1	A	66	ILE	6.3
1	A	152	ALA	5.9
1	B	64	LEU	5.3
1	A	64	LEU	5.2
1	A	92	GLY	5.1
1	A	144	GLU	4.5
1	A	55	GLU	4.4
1	B	66	ILE	4.2
1	A	69	GLU	3.9
1	B	143	GLU	3.8
1	A	68	GLU	3.7
1	B	155	ILE	3.7
1	B	144	GLU	3.6
1	A	65	ARG	3.2
1	A	101	GLY	3.2
1	A	147	GLU	3.1
1	B	147	GLU	3.1
1	B	151	LYS	3.1
1	B	65	ARG	3.1
1	B	67	ASP	3.0
1	A	151	LYS	2.9
1	A	71	GLN	2.8
1	B	92	GLY	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	70	LEU	2.6
1	A	155	ILE	2.6
1	A	181	GLY	2.5
1	B	69	GLU	2.4
1	B	157	GLU	2.4
1	B	153	LEU	2.4
1	A	67	ASP	2.4
1	B	55	GLU	2.3
1	A	90	LEU	2.3
1	B	24	HIS	2.3
1	A	148	SER	2.3
1	B	68	GLU	2.2
1	B	72	GLN	2.2
1	B	179	LEU	2.1
1	A	63	ALA	2.1
1	B	181	GLY	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.