Full wwPDB X-ray Structure Validation Report (i)

Jun 12, 2024 – 10:13 PM EDT

PDB ID	:	1K4N
Title	:	Structural Genomics, Protein EC4020
Authors	:	Zhang, R.G.; Joachimiak, A.; Edwards, A.; Savchenko, A.; Skarina, T.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2001-10-08
Resolution	:	1.60 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
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 (i)) were used in the production of this report:

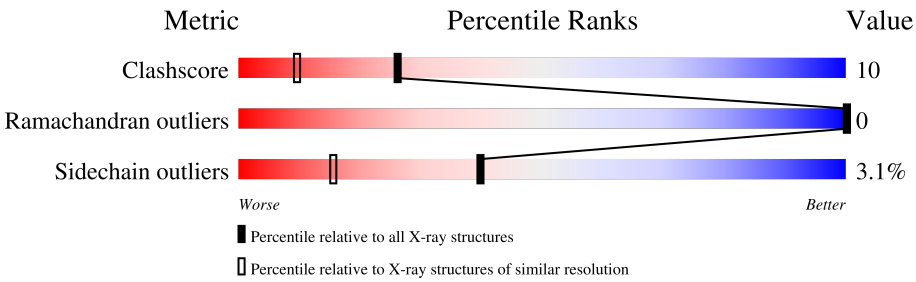
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 1.60 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))
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	192	76% 18% • 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1713 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 Protein EC4020.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	183	1460	926	259	269	3	3	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP P52007
A	2	HIS	-	CLONING ARTIFACT	UNP P52007
A	3	MSE	-	CLONING ARTIFACT	UNP P52007
A	4	ILE	-	CLONING ARTIFACT	UNP P52007
A	5	MSE	MET	MODIFIED RESIDUE	UNP P52007
A	75	MSE	MET	MODIFIED RESIDUE	UNP P52007

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	253	253	253	0	0

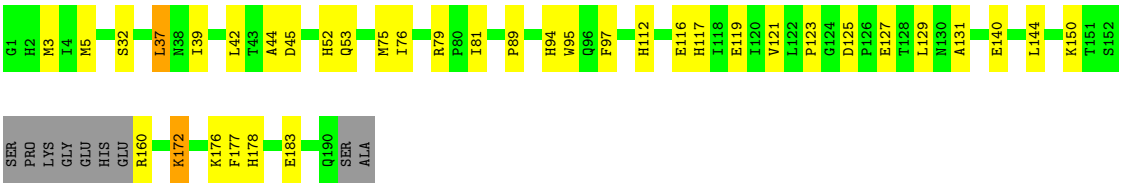
3 Residue-property plots (i)

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. 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. 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.

Note EDS was not executed.

- Molecule 1: Protein EC4020

Chain A:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	38.66Å 41.78Å 105.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.86 – 1.60	Depositor
% Data completeness (in resolution range)	89.8 (38.86-1.60)	Depositor
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS	Depositor
R, R_{free}	0.199 , 0.217	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1713	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

5 Model quality (i)

5.1 Standard geometry (i)

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	0.29	0/1495	0.59	0/2030

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 (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	1460	0	1437	30	0
2	A	253	0	0	5	0
All	All	1713	0	1437	30	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 10.

All (30) 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:3:MSE:HE3	1:A:5:MSE:SE	2.17	0.94
1:A:52:HIS:HD2	1:A:53:GLN:HE21	1.20	0.89
1:A:39:ILE:HG23	1:A:95:TRP:CZ2	2.19	0.77
1:A:116:GLU:OE1	1:A:117:HIS:HD2	1.71	0.74

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:45:ASP:HB2	1:A:121:VAL:HG23	1.68	0.74
1:A:140:GLU:O	1:A:144:LEU:HG	1.93	0.68
1:A:39:ILE:HG22	1:A:94:HIS:NE2	2.12	0.64
1:A:52:HIS:CD2	1:A:53:GLN:HE21	2.09	0.61
1:A:37:LEU:HD11	1:A:177:PHE:HZ	1.66	0.59
1:A:112:HIS:HD2	2:A:280:HOH:O	1.86	0.58
1:A:3:MSE:HG2	1:A:5:MSE:HG3	1.85	0.57
1:A:32:SER:OG	1:A:39:ILE:HD11	2.07	0.53
1:A:131:ALA:HB1	2:A:401:HOH:O	2.07	0.53
1:A:183:GLU:HG2	2:A:403:HOH:O	2.09	0.52
1:A:160:ARG:HD2	1:A:160:ARG:C	2.29	0.52
1:A:119:GLU:HG2	1:A:176:LYS:HB2	1.93	0.50
1:A:125:ASP:OD2	1:A:127:GLU:HB2	2.12	0.49
1:A:44:ALA:HB1	1:A:97:PHE:CE1	2.47	0.49
1:A:52:HIS:CD2	1:A:53:GLN:HG2	2.49	0.48
1:A:76:ILE:HG13	1:A:81:ILE:HD12	1.96	0.48
1:A:129:LEU:C	1:A:129:LEU:HD13	2.35	0.46
1:A:117:HIS:HE1	1:A:119:GLU:OE2	1.99	0.46
1:A:150:LYS:HD3	2:A:428:HOH:O	2.16	0.45
1:A:75:MSE:HE2	1:A:75:MSE:HA	1.99	0.45
1:A:39:ILE:HG23	1:A:95:TRP:CH2	2.51	0.45
1:A:172:LYS:HE2	1:A:172:LYS:HB3	1.70	0.44
1:A:123:PRO:HG3	2:A:343:HOH:O	2.18	0.43
1:A:116:GLU:OE1	1:A:117:HIS:CD2	2.61	0.42
1:A:37:LEU:HD22	1:A:42:LEU:HD11	2.01	0.42
1:A:44:ALA:HB2	1:A:95:TRP:CD1	2.54	0.42

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	179/192 (93%)	174 (97%)	5 (3%)	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	160/164 (98%)	155 (97%)	5 (3%)	40	15

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	37	LEU
1	A	79	ARG
1	A	89	PRO
1	A	172	LYS
1	A	178	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	46	HIS
1	A	52	HIS
1	A	77	ASN
1	A	91	GLN
1	A	112	HIS
1	A	117	HIS
1	A	163	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 (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.