wwPDB X-ray Structure Validation Summary Report (i)

May 29, 2020 – 04:59 am BST

PDB ID	:	4HSD
Title	:	Crystal structure of a new form of plant lectin from Cicer arietinum at 2.45 Angstrom resolution
Authors	:	Kumar, S.; Singh, A.; Bhushan, A.; Dey, S.; Sharma, S.; Singh, T.P.
Deposited on	:	2012-10-30
Resolution	:	2.45 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 following versions of software and data (see references (1)) were used in the production of this report:

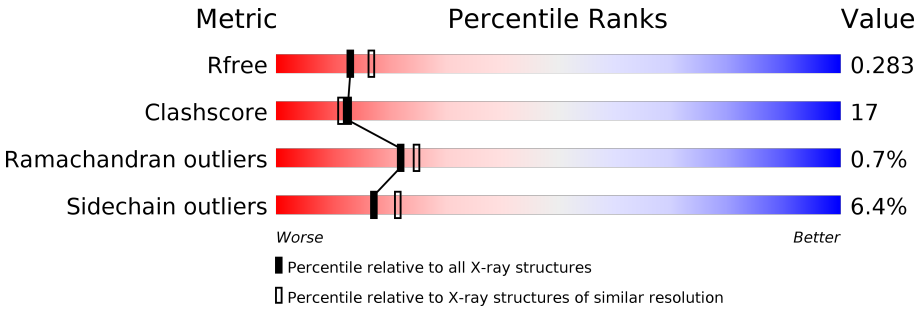
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.45 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	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)

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 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%

Mol	Chain	Length	Quality of chain
1	A	227	70% 25%
1	B	227	64% 31%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3816 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 Lectin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	224	Total	C	N	O	S	0	0	0
			1796	1162	291	337	6			
1	B	224	Total	C	N	O	S	0	0	0
			1796	1162	291	337	6			

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Na	0	0
			1	1		
2	A	1	Total	Na	0	0
			1	1		

- Molecule 3 is water.

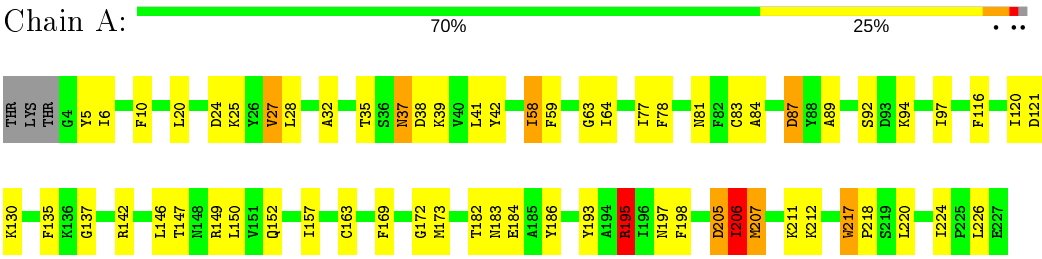
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	115	Total	O	0	0
			115	115		
3	B	107	Total	O	0	0
			107	107		

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. 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.

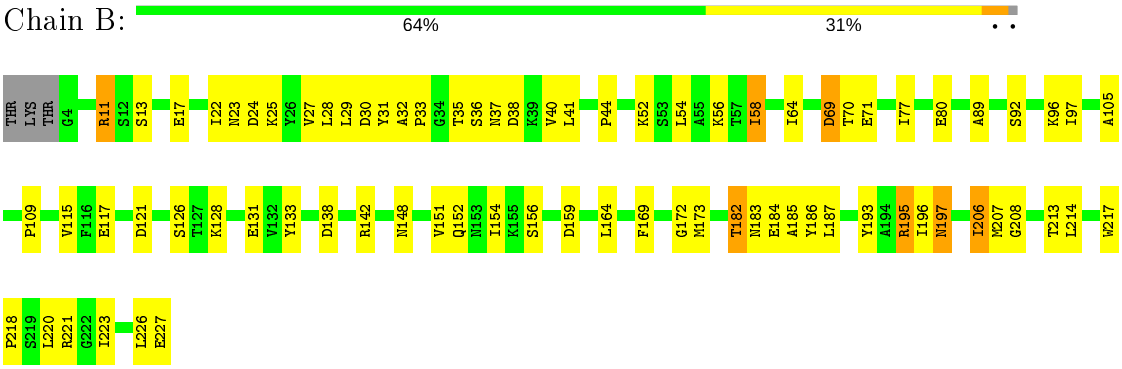
- Molecule 1: Lectin

Chain A:



- Molecule 1: Lectin

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, α, β, γ	80.48Å 80.48Å 69.18Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	69.70 – 2.45 69.70 – 2.45	Depositor EDS
% Data completeness (in resolution range)	97.0 (69.70-2.45) 97.0 (69.70-2.45)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	16.42 (at 2.45Å)	Xtriage
Refinement program	REFMAC 5.5.0110	Depositor
R, R_{free}	0.256 , 0.289 0.255 , 0.283	Depositor DCC
R_{free} test set	916 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	22.5	Xtriage
Anisotropy	0.137	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 17.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.39$, $< L^2 > = 0.21$	Xtriage
Estimated twinning fraction	0.120 for -h,-k,l 0.477 for h,-h-k,-l 0.125 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	3816	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.96% 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: NA

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	0.81	1/1844 (0.1%)	0.98	8/2487 (0.3%)
1	B	0.81	2/1844 (0.1%)	0.88	3/2487 (0.1%)
All	All	0.81	3/3688 (0.1%)	0.93	11/4974 (0.2%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	80	GLU	CD-OE2	-7.51	1.17	1.25
1	B	80	GLU	CD-OE1	-7.04	1.18	1.25
1	A	207	MET	CB-CG	-5.08	1.35	1.51

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	206	ILE	CB-CG1-CD1	13.31	151.16	113.90
1	A	58	ILE	CB-CG1-CD1	9.67	140.98	113.90
1	A	205	ASP	CB-CG-OD1	8.85	126.27	118.30
1	A	195	ARG	NE-CZ-NH2	8.31	124.46	120.30
1	A	206	ILE	CB-CA-C	6.15	123.91	111.60

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	1796	0	1735	61	0
1	B	1796	0	1735	57	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	115	0	0	3	0
3	B	107	0	0	2	0
All	All	3816	0	3470	117	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 17.

The worst 5 of 117 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:32:ALA:HB1	1:A:35:THR:CG2	2.00	0.91
1:B:207:MET:HB2	1:B:208:GLY:HA3	1.53	0.91
1:B:25:LYS:HE3	1:B:226:LEU:HD21	1.53	0.90
1:A:206:ILE:CG2	1:A:207:MET:HA	2.02	0.90
1:A:186:TYR:CE1	1:A:195:ARG:HG3	2.06	0.89

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	222/227 (98%)	205 (92%)	16 (7%)	1 (0%)	29 34
1	B	222/227 (98%)	194 (87%)	26 (12%)	2 (1%)	17 19
All	All	444/454 (98%)	399 (90%)	42 (10%)	3 (1%)	22 25

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	206	ILE
1	B	206	ILE
1	B	44	PRO

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	188/191 (98%)	183 (97%)	5 (3%)	44 57
1	B	188/191 (98%)	169 (90%)	19 (10%)	7 7
All	All	376/382 (98%)	352 (94%)	24 (6%)	17 21

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	37	ASN
1	B	69	ASP
1	B	223	ILE
1	B	52	LYS
1	B	58	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	37	ASN
1	A	197	ASN
1	B	16	ASN
1	B	73	ASN
1	B	197	ASN

5.3.3 RNA (i)

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

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.