Full wwPDB X-ray Structure Validation Report (i)

Oct 10, 2023 – 10:12 AM EDT

PDB ID	:	7JJM
Title	:	Crystal structure of Importin alpha 2 in complex with LSD1 NLS
Authors	:	Tu, W.J.; McGuaig, R.; Tan, H.Y.A.; Hardy, C.; Seddiki, N.; Ali, S.; Dahlstrom, J.E.; Bean, E.G.; Dunn, J.; Forwood, J.K.; Tsimbalyuk, S.; Smith, K.M.; Yip, D.; Malik, L.; Prasana, T.; Milburn, P.; Rao, S.
Deposited on	:	2020-07-27
Resolution	:	2.06 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 (1)) were used in the production of this report:

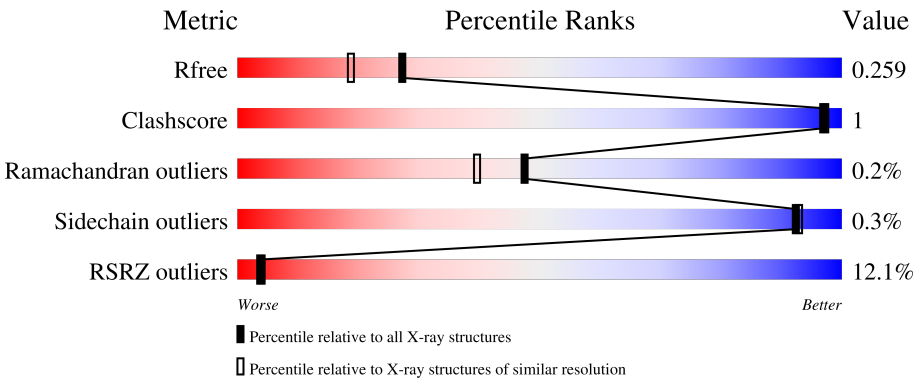
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.06 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	2684 (2.08-2.04)
Clashscore	141614	2801 (2.08-2.04)
Ramachandran outliers	138981	2768 (2.08-2.04)
Sidechain outliers	138945	2768 (2.08-2.04)
RSRZ outliers	127900	2646 (2.08-2.04)

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% 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	26	23% 77%
2	B	470	11% 87% • 10%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6793 atoms, of which 3397 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 Lysine-specific histone demethylase 1A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	H	N	O			
1	A	6	103	29	57	11	6	0	0	0

- Molecule 2 is a protein called Importin subunit alpha-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	422	6586	2069	3340	549	618	10	0	7	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	530	GLY	-	expression tag	UNP P52293
B	531	SER	-	expression tag	UNP P52293

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
3	B	1	1	1	0	0

- Molecule 4 is water.

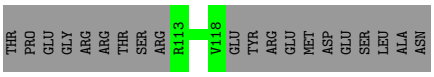
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	1	1	1	0	0
4	B	102	Total 102	O 102	0	0

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 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.

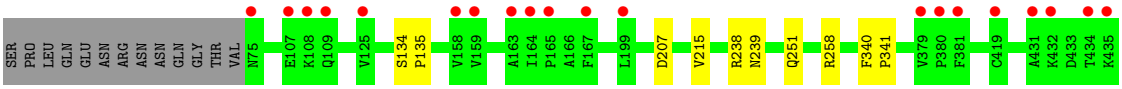
• Molecule 1: Lysine-specific histone demethylase 1A

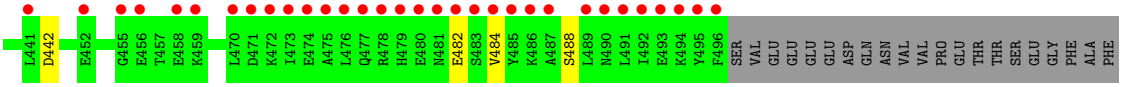
Chain A:



• Molecule 2: Importin subunit alpha-1

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	78.61Å 90.05Å 99.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.04 – 2.06 41.04 – 2.06	Depositor EDS
% Data completeness (in resolution range)	98.5 (41.04-2.06) 98.5 (41.04-2.06)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.44 (at 2.06Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.220 , 0.259 0.220 , 0.259	Depositor DCC
R_{free} test set	2183 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	36.6	Xtriage
Anisotropy	0.048	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.44 , 48.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6793	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.15% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CL

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	0.24	0/45	0.54	0/57
2	B	0.24	0/3325	0.40	0/4529
All	All	0.24	0/3370	0.40	0/4586

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	46	57	54	0	0
2	B	3246	3340	3339	6	0
3	B	1	0	0	0	0
4	A	1	0	0	0	0
4	B	102	0	0	0	0
All	All	3396	3397	3393	6	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.

All (6) 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)
2:B:484:VAL:O	2:B:488:SER:OG	2.06	0.73
2:B:340:PHE:N	2:B:341:PRO:CD	2.80	0.45
2:B:207:ASP:OD1	2:B:251:GLN:NE2	2.50	0.45
2:B:215:VAL:O	2:B:258[B]:ARG:NH2	2.46	0.43
2:B:238:ARG:O	2:B:239:ASN:HB2	2.18	0.43
2:B:134:SER:N	2:B:135:PRO:CD	2.85	0.40

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	4/26 (15%)	4 (100%)	0	0	100	100
2	B	427/470 (91%)	414 (97%)	12 (3%)	1 (0%)	47	39
All	All	431/496 (87%)	418 (97%)	12 (3%)	1 (0%)	47	39

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	482	GLU

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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	4/23 (17%)	4 (100%)	0	100	100
2	B	361/395 (91%)	360 (100%)	1 (0%)	92	93
All	All	365/418 (87%)	364 (100%)	1 (0%)	92	93

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	442	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

Of 1 ligands modelled in this entry, 1 is 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 (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)

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	6/26 (23%)	-0.22	0 100 100	33, 35, 41, 48	0
2	B	422/470 (89%)	0.72	52 (12%) 4 3	26, 40, 83, 111	0
All	All	428/496 (86%)	0.71	52 (12%) 4 4	26, 40, 83, 111	0

All (52) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	485	TYR	9.8
2	B	476	LEU	6.8
2	B	478	ARG	6.6
2	B	489	LEU	6.3
2	B	475	ALA	6.2
2	B	480	GLU	5.7
2	B	496	PHE	5.5
2	B	492	ILE	5.1
2	B	482	GLU	4.9
2	B	491	LEU	4.7
2	B	486	LYS	4.5
2	B	484	VAL	4.5
2	B	487	ALA	4.4
2	B	474	GLU	4.1
2	B	108	LYS	4.1
2	B	477	GLN	4.1
2	B	493	GLU	3.8
2	B	75	ASN	3.7
2	B	432	LYS	3.7
2	B	470	LEU	3.7
2	B	472	LYS	3.6
2	B	479	HIS	3.6
2	B	434	THR	3.5
2	B	481	ASN	3.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	473	ILE	3.4
2	B	419	CYS	3.1
2	B	455	GLY	3.0
2	B	158	VAL	2.9
2	B	471	ASP	2.8
2	B	159	VAL	2.8
2	B	495	TYR	2.8
2	B	458	GLU	2.7
2	B	431	ALA	2.7
2	B	483	SER	2.6
2	B	441	LEU	2.6
2	B	490	ASN	2.5
2	B	107	GLU	2.4
2	B	456	GLU	2.4
2	B	109	GLN	2.4
2	B	381	PHE	2.3
2	B	125	VAL	2.3
2	B	163	ALA	2.3
2	B	164	ILE	2.3
2	B	379	VAL	2.3
2	B	459	LYS	2.2
2	B	380	PRO	2.2
2	B	494	LYS	2.1
2	B	165	PRO	2.1
2	B	435	LYS	2.1
2	B	167	PHE	2.0
2	B	452	GLU	2.0
2	B	199	LEU	2.0

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 monosaccharides in this entry.

6.4 Ligands (i)

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. The B-factors column lists the minimum,

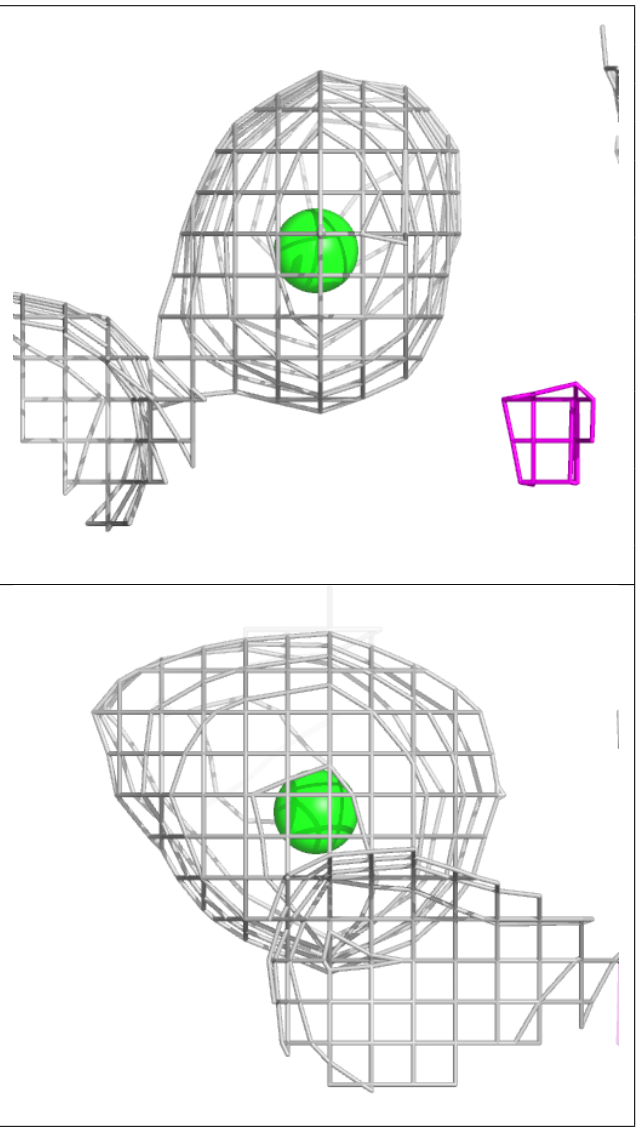
median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	CL	B	601	1/1	0.92	0.13	60,60,60,60	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around CL B 601:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.