Full wwPDB X-ray Structure Validation Report ⓘ

Feb 12, 2024 – 02:00 PM EST

PDB ID	:	8G8S
Title	:	bipartite p52 NLS in complex with Importin alpha 2
Authors	:	Tsimbslyuk, S.; Stewart, M.; Forwood, J.K.
Deposited on	:	2023-02-19
Resolution	:	2.10 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 ⓘ 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 ⓘ) were used in the production of this report:

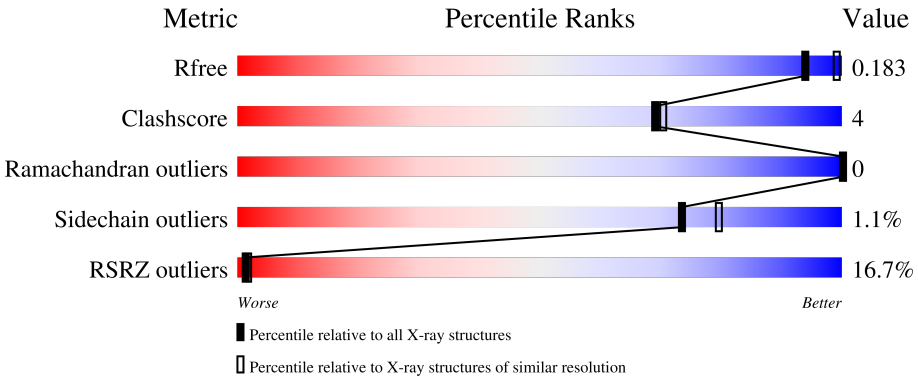
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	510	13% 77% 6% 16%
2	B	35	23% 34% 11% · 51%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7205 atoms, of which 3552 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 Importin subunit alpha-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	426	6600	2076	3341	553	620	10	0	2	0

There are 50 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	MET	-	expression tag	UNP P52293
A	21	HIS	-	expression tag	UNP P52293
A	22	HIS	-	expression tag	UNP P52293
A	23	HIS	-	expression tag	UNP P52293
A	24	HIS	-	expression tag	UNP P52293
A	25	HIS	-	expression tag	UNP P52293
A	26	HIS	-	expression tag	UNP P52293
A	27	SER	-	expression tag	UNP P52293
A	28	SER	-	expression tag	UNP P52293
A	29	GLY	-	expression tag	UNP P52293
A	30	LEU	-	expression tag	UNP P52293
A	31	VAL	-	expression tag	UNP P52293
A	32	PRO	-	expression tag	UNP P52293
A	33	ARG	-	expression tag	UNP P52293
A	34	GLY	-	expression tag	UNP P52293
A	35	SER	-	expression tag	UNP P52293
A	36	GLY	-	expression tag	UNP P52293
A	37	MET	-	expression tag	UNP P52293
A	38	LEU	-	expression tag	UNP P52293
A	39	GLU	-	expression tag	UNP P52293
A	40	THR	-	expression tag	UNP P52293
A	41	ALA	-	expression tag	UNP P52293
A	42	ALA	-	expression tag	UNP P52293
A	43	ALA	-	expression tag	UNP P52293
A	44	LEU	-	expression tag	UNP P52293
A	45	PHE	-	expression tag	UNP P52293
A	46	GLU	-	expression tag	UNP P52293

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	47	ARG	-	expression tag	UNP P52293
A	48	ASN	-	expression tag	UNP P52293
A	49	HIS	-	expression tag	UNP P52293
A	50	MET	-	expression tag	UNP P52293
A	51	ASP	-	expression tag	UNP P52293
A	52	SER	-	expression tag	UNP P52293
A	53	PRO	-	expression tag	UNP P52293
A	54	ASP	-	expression tag	UNP P52293
A	55	LEU	-	expression tag	UNP P52293
A	56	GLY	-	expression tag	UNP P52293
A	57	THR	-	expression tag	UNP P52293
A	58	ASP	-	expression tag	UNP P52293
A	59	ASP	-	expression tag	UNP P52293
A	60	ASP	-	expression tag	UNP P52293
A	61	ASP	-	expression tag	UNP P52293
A	62	LEU	-	expression tag	UNP P52293
A	63	ALA	-	expression tag	UNP P52293
A	64	MET	-	expression tag	UNP P52293
A	65	ALA	-	expression tag	UNP P52293
A	66	ASP	-	expression tag	UNP P52293
A	67	ILE	-	expression tag	UNP P52293
A	68	GLY	-	expression tag	UNP P52293
A	69	SER	-	expression tag	UNP P52293

- Molecule 2 is a protein called Nuclear factor NF-kappa-B p100 subunit isoform X1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	17	Total	C	H	N	O	0	3	0
			397	110	211	48	28			

- Molecule 3 is water.

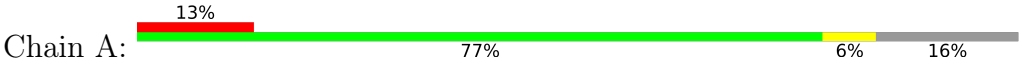
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	202	Total	O	0	0
			202	202		
3	B	6	Total	O	0	0
			6	6		

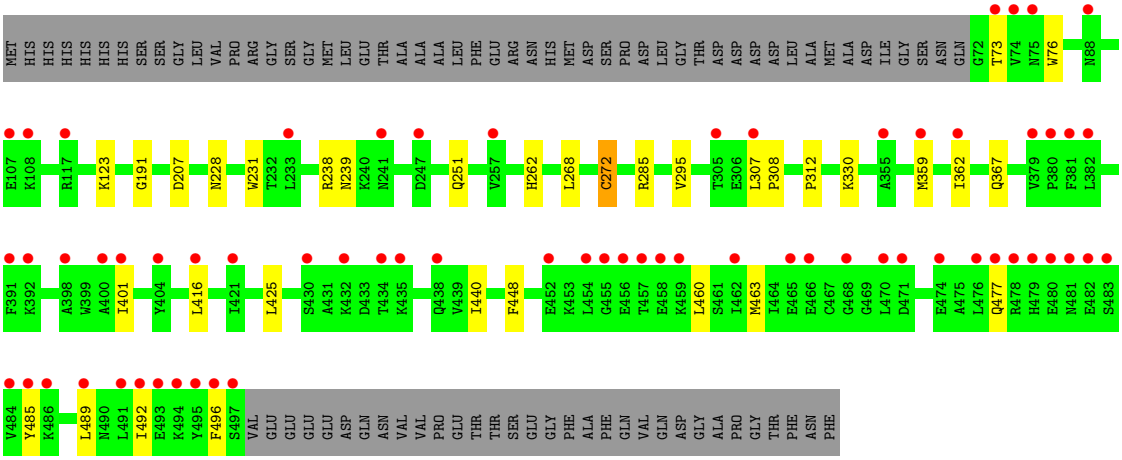
3 Residue-property plots (i)

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: Importin subunit alpha-1

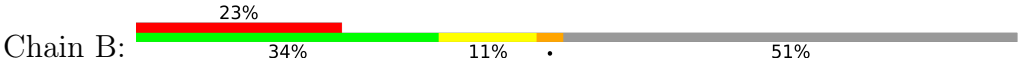
Chain A:

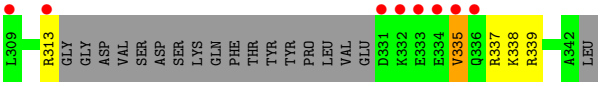




• Molecule 2: Nuclear factor NF-kappa-B p100 subunit isoform X1

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	77.07Å 88.58Å 96.67Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.73 – 2.10 29.73 – 2.10	Depositor EDS
% Data completeness (in resolution range)	100.0 (29.73-2.10) 100.0 (29.73-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.37 (at 2.10Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.169 , 0.186 0.166 , 0.183	Depositor DCC
R_{free} test set	1917 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å^2)	35.6	Xtriage
Anisotropy	0.490	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 48.6	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	0.96	EDS
Total number of atoms	7205	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.79% 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

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	0.39	1/3320 (0.0%)	0.55	0/4523
2	B	0.36	0/183	0.73	0/231
All	All	0.39	1/3503 (0.0%)	0.56	0/4754

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	272	CYS	CB-SG	-6.09	1.71	1.82

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	3259	3341	3340	23	0
2	B	186	211	209	8	0
3	A	202	0	0	3	0
3	B	6	0	0	0	0
All	All	3653	3552	3549	26	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 4.

All (26) 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:207:ASP:OD1	1:A:251:GLN:NE2	2.30	0.64
1:A:231:TRP:CE2	2:B:339[B]:ARG:HD3	2.37	0.59
1:A:272:CYS:HB3	1:A:312:PRO:HB2	1.85	0.58
1:A:425:LEU:HG	1:A:440:ILE:HG23	1.88	0.55
1:A:492:ILE:HA	1:A:496:PHE:HB2	1.87	0.55
2:B:335:VAL:HG11	2:B:339[A]:ARG:HG2	1.91	0.52
1:A:285:ARG:HD3	3:A:672:HOH:O	2.10	0.50
1:A:359:MET:HE1	1:A:362:ILE:HD12	1.94	0.47
1:A:231:TRP:CD2	2:B:339[B]:ARG:HD3	2.50	0.47
1:A:238:ARG:O	1:A:239:ASN:HB2	2.15	0.47
1:A:489:LEU:HD12	1:A:492:ILE:HD11	1.97	0.47
2:B:335:VAL:HG13	2:B:338:LYS:O	2.14	0.46
1:A:401:ILE:HD12	1:A:416:LEU:HD11	1.97	0.45
1:A:231:TRP:CZ2	2:B:339[B]:ARG:HD3	2.51	0.45
1:A:448:PHE:CD1	1:A:460:LEU:HD23	2.52	0.45
1:A:401:ILE:HG23	1:A:416:LEU:HD21	1.99	0.45
1:A:191:GLY:O	2:B:337:ARG:HG2	2.17	0.44
1:A:123[B]:LYS:NZ	3:A:611:HOH:O	2.50	0.44
2:B:335:VAL:CG1	2:B:339[A]:ARG:HG2	2.48	0.44
1:A:262:HIS:O	1:A:268:LEU:HD21	2.17	0.44
1:A:367:GLN:OE1	3:A:601:HOH:O	2.22	0.43
1:A:228:ASN:OD1	2:B:339[B]:ARG:NH2	2.52	0.42
1:A:307:LEU:N	1:A:308:PRO:HD2	2.35	0.42
1:A:477:GLN:CD	1:A:492:ILE:CD1	2.90	0.40
1:A:73:THR:HA	1:A:76:TRP:NE1	2.36	0.40
1:A:295:VAL:HB	1:A:330:LYS:HE3	2.03	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	426/510 (84%)	419 (98%)	7 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	B	15/35 (43%)	13 (87%)	2 (13%)	0	100 100
All	All	441/545 (81%)	432 (98%)	9 (2%)	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	359/426 (84%)	357 (99%)	2 (1%)	86 90
2	B	19/32 (59%)	16 (84%)	3 (16%)	2 1
All	All	378/458 (82%)	373 (99%)	5 (1%)	73 75

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

Mol	Chain	Res	Type
1	A	463	MET
1	A	485	TYR
2	B	313[A]	ARG
2	B	313[B]	ARG
2	B	335	VAL

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)

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, 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	426/510 (83%)	0.74	66 (15%)	2	2	26, 41, 97, 131	0
2	B	17/35 (48%)	2.37	8 (47%)	0	0	37, 65, 130, 130	0
All	All	443/545 (81%)	0.80	74 (16%)	1	2	26, 41, 100, 131	0

All (74) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	335	VAL	8.0
2	B	334	GLU	6.7
1	A	432	LYS	6.1
1	A	485	TYR	6.0
2	B	333	GLU	5.9
1	A	497	SER	5.9
2	B	309	LEU	5.1
2	B	332	LYS	4.9
1	A	454	LEU	4.8
1	A	486	LYS	4.5
1	A	478	ARG	4.5
1	A	75	ASN	4.3
1	A	493	GLU	4.3
1	A	480	GLU	4.0
1	A	482	GLU	4.0
1	A	73	THR	3.9
2	B	313[A]	ARG	3.9
1	A	476	LEU	3.9
1	A	474	GLU	3.9
1	A	452	GLU	3.9
1	A	496	PHE	3.8
1	A	455	GLY	3.8
1	A	479	HIS	3.8
1	A	74	VAL	3.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	495	TYR	3.7
1	A	307	LEU	3.7
1	A	400	ALA	3.6
1	A	435	LYS	3.5
1	A	88	ASN	3.4
1	A	477	GLN	3.4
1	A	462	ILE	3.4
1	A	382	LEU	3.3
1	A	457	THR	3.3
1	A	458	GLU	3.2
1	A	483	SER	3.1
1	A	456	GLU	3.1
1	A	484	VAL	3.1
1	A	391	PHE	3.0
1	A	494	LYS	2.9
1	A	404	TYR	2.9
1	A	355	ALA	2.9
1	A	359	MET	2.9
1	A	489	LEU	2.8
1	A	471	ASP	2.7
1	A	416	LEU	2.7
1	A	481	ASN	2.7
1	A	421	ILE	2.7
1	A	430	SER	2.6
1	A	241	ASN	2.6
1	A	492	ILE	2.6
1	A	257	VAL	2.6
1	A	305	THR	2.6
1	A	401	ILE	2.5
2	B	331	ASP	2.5
1	A	380	PRO	2.4
1	A	470	LEU	2.4
1	A	466	GLU	2.4
1	A	381	PHE	2.4
1	A	465	GLU	2.4
1	A	107	GLU	2.3
1	A	247	ASP	2.3
1	A	392	LYS	2.3
1	A	379	VAL	2.3
1	A	459	LYS	2.2
1	A	434	THR	2.2
1	A	108	LYS	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	491	LEU	2.2
2	B	336[A]	GLN	2.1
1	A	233	LEU	2.1
1	A	117	ARG	2.1
1	A	438	GLN	2.0
1	A	362	ILE	2.0
1	A	398	ALA	2.0
1	A	468	GLY	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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.