



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 07:18 PM GMT

PDB ID : 4OIH  
Title : Importin Alpha in Complex with the Bipartite NLS of Prp20  
Authors : Roman, N.; Christie, M.; Swarbrick, C.M.D.; Kobe, B.; Forwood, J.K.  
Deposited on : 2014-01-19  
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](mailto: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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

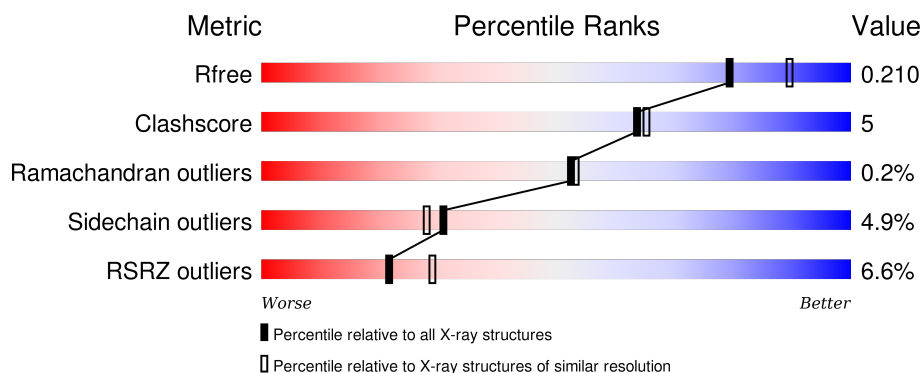
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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))
$R_{free}$	91344	3939 (2.10-2.10)
Clashscore	102246	4460 (2.10-2.10)
Ramachandran outliers	100387	4413 (2.10-2.10)
Sidechain outliers	100360	4414 (2.10-2.10)
RSRZ outliers	91569	3948 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	460	<div> <div>6%</div> <div>78%</div> <div>13%</div> <div>7%</div> </div>
2	B	27	<div> <div>11%</div> <div>37%</div> <div>11%</div> <div>52%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3480 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 Importin subunit alpha-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	426	Total	C	N	O	S	0	0	0
			3244	2066	550	618	10			

- Molecule 2 is a protein called Guanine nucleotide exchange factor SRM1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	13	Total	C	N	O	S	0	0	0
			111	67	27	16	1			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	GLY	-	EXPRESSION TAG	UNP P21827
B	0	SER	-	EXPRESSION TAG	UNP P21827

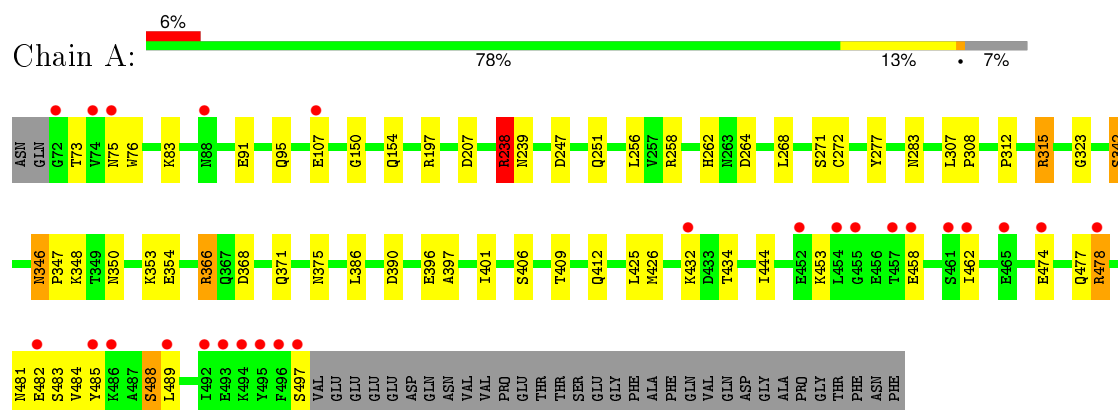
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	124	Total	O	0	0
			124	124		
3	B	1	Total	O	0	0
			1	1		

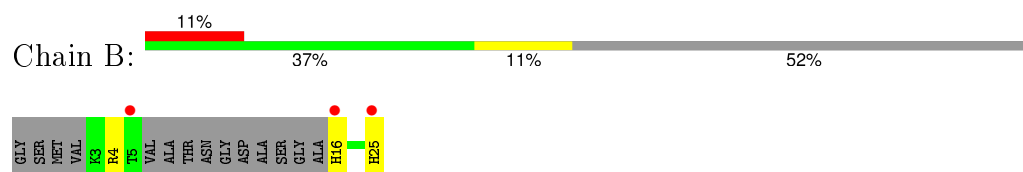
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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



#### • Molecule 2: Guanine nucleotide exchange factor SRM1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	78.91Å 89.92Å 99.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.37 – 2.10 36.37 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.8 (36.37-2.10) 99.9 (36.37-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.04 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.168 , 0.201 0.180 , 0.210	Depositor DCC
$R_{free}$ test set	2126 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	26.9	Xtriage
Anisotropy	0.255	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 47.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 42031 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3480	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.91% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	1.09	5/3302 (0.2%)	1.02	9/4500 (0.2%)
2	B	1.20	1/111 (0.9%)	1.33	1/141 (0.7%)
All	All	1.10	6/3413 (0.2%)	1.03	10/4641 (0.2%)

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	342	SER	CB-OG	-6.66	1.33	1.42
1	A	271	SER	CB-OG	-6.53	1.33	1.42
1	A	396	GLU	CD-OE2	-5.82	1.19	1.25
1	A	91	GLU	CD-OE2	-5.56	1.19	1.25
2	B	25	HIS	N-CA	5.55	1.57	1.46
1	A	406	SER	CB-OG	-5.49	1.35	1.42

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	197	ARG	NE-CZ-NH2	-8.01	116.29	120.30
1	A	366	ARG	NE-CZ-NH2	-7.79	116.40	120.30
2	B	4	ARG	NE-CZ-NH1	6.43	123.51	120.30
1	A	390	ASP	CB-CG-OD1	-6.31	112.62	118.30
1	A	247	ASP	CB-CG-OD2	-6.05	112.85	118.30
1	A	366	ARG	NE-CZ-NH1	5.83	123.22	120.30
1	A	264	ASP	CB-CG-OD2	5.57	123.31	118.30
1	A	258	ARG	CG-CD-NE	-5.50	100.26	111.80
1	A	238	ARG	NE-CZ-NH2	5.35	122.97	120.30
1	A	256	LEU	CA-CB-CG	5.31	127.51	115.30

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	3244	0	3320	35	0
2	B	111	0	123	0	0
3	A	124	0	0	0	0
3	B	1	0	0	0	0
All	All	3480	0	3443	35	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 5.

All (35) 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:346:ASN:HD22	1:A:348:LYS:H	1.32	0.74
1:A:238:ARG:HG3	1:A:277:TYR:CE2	2.26	0.71
1:A:386:LEU:HD21	1:A:425:LEU:HD13	1.75	0.68
1:A:346:ASN:ND2	1:A:348:LYS:H	1.95	0.64
1:A:426:MET:CE	1:A:444:ILE:HD11	2.30	0.62
1:A:482:GLU:HA	1:A:485:TYR:CE2	2.41	0.56
1:A:315:ARG:HH11	1:A:315:ARG:HB2	1.70	0.55
1:A:484:VAL:O	1:A:488:SER:OG	2.25	0.54
1:A:315:ARG:NH1	1:A:315:ARG:HB2	2.23	0.54
1:A:371:GLN:HE21	1:A:375:ASN:HD21	1.58	0.52
1:A:307:LEU:HB3	1:A:308:PRO:HD3	1.92	0.52
1:A:458:GLU:O	1:A:462:ILE:HD12	2.12	0.50
1:A:366:ARG:HD2	1:A:368:ASP:OD1	2.12	0.50
1:A:238:ARG:HG3	1:A:277:TYR:CZ	2.47	0.49
1:A:481:ASN:HD22	1:A:484:VAL:H	1.60	0.48
1:A:478:ARG:HB2	1:A:478:ARG:NH1	2.29	0.47
1:A:426:MET:HE3	1:A:444:ILE:HD11	1.96	0.47
1:A:477:GLN:HE21	1:A:489:LEU:HA	1.79	0.47
1:A:272:CYS:HB3	1:A:312:PRO:HB2	1.95	0.47
1:A:346:ASN:HD22	1:A:348:LYS:N	2.08	0.46
1:A:397:ALA:O	1:A:401:ILE:HG12	2.16	0.46
1:A:409:THR:H	1:A:412:GLN:NE2	2.14	0.45
1:A:477:GLN:O	1:A:485:TYR:HB2	2.16	0.45

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:315:ARG:NH1	1:A:354:GLU:OE1	2.50	0.45
1:A:283:ASN:ND2	1:A:323:GLY:HA2	2.32	0.45
1:A:353:LYS:HE3	1:A:353:LYS:HB3	1.79	0.44
1:A:238:ARG:HG3	1:A:277:TYR:CD2	2.53	0.44
1:A:409:THR:H	1:A:412:GLN:HE21	1.65	0.44
1:A:478:ARG:CZ	1:A:478:ARG:HB2	2.47	0.44
1:A:207:ASP:OD1	1:A:251:GLN:NE2	2.51	0.43
1:A:73:THR:HA	1:A:76:TRP:NE1	2.34	0.42
1:A:346:ASN:ND2	1:A:347:PRO:HD2	2.35	0.42
1:A:426:MET:CE	1:A:444:ILE:CD1	2.98	0.42
1:A:262:HIS:O	1:A:268:LEU:HD21	2.21	0.41
1:A:150:GLY:HA3	1:A:154:GLN:OE1	2.22	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	424/460 (92%)	418 (99%)	5 (1%)	1 (0%)	52	53
2	B	9/27 (33%)	9 (100%)	0	0	100	100
All	All	433/487 (89%)	427 (99%)	5 (1%)	1 (0%)	52	53

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	239	ASN



### 5.3.2 Protein sidechains ⓘ

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	357/386 (92%)	340 (95%)	17 (5%)	31	29
2	B	12/20 (60%)	11 (92%)	1 (8%)	14	9
All	All	369/406 (91%)	351 (95%)	18 (5%)	31	28

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	75	ASN
1	A	83	LYS
1	A	95	GLN
1	A	107	GLU
1	A	238	ARG
1	A	315	ARG
1	A	342	SER
1	A	346	ASN
1	A	350	ASN
1	A	432	LYS
1	A	434	THR
1	A	453	LYS
1	A	474	GLU
1	A	478	ARG
1	A	483	SER
1	A	488	SER
1	A	497	SER
2	B	16	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	86	ASN
1	A	283	ASN
1	A	346	ASN
1	A	352	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	375	ASN
1	A	412	GLN
1	A	438	GLN
1	A	477	GLN
1	A	479	HIS
1	A	481	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 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	426/460 (92%)	0.15	26 (6%) 25 33	17, 28, 67, 91	0
2	B	13/27 (48%)	0.91	3 (23%) 1 1	29, 44, 68, 76	0
All	All	439/487 (90%)	0.17	29 (6%) 22 29	17, 29, 68, 91	0

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	489	LEU	5.7
1	A	74	VAL	5.4
1	A	496	PHE	5.4
1	A	485	TYR	4.9
1	A	455	GLY	4.6
1	A	497	SER	4.4
2	B	25	HIS	4.0
2	B	16	HIS	3.7
1	A	482	GLU	3.4
1	A	432	LYS	3.4
1	A	495	TYR	3.4
1	A	458	GLU	3.2
2	B	5	THR	3.1
1	A	493	GLU	3.1
1	A	494	LYS	3.1
1	A	107	GLU	2.9
1	A	465	GLU	2.9
1	A	492	ILE	2.8
1	A	462	ILE	2.8
1	A	454	LEU	2.8
1	A	457	THR	2.7
1	A	478	ARG	2.6
1	A	452	GLU	2.5
1	A	474	GLU	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	88	ASN	2.4
1	A	75	ASN	2.4
1	A	461	SER	2.3
1	A	72	GLY	2.3
1	A	486	LYS	2.2

## 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.