wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:36 AM GMT

PDB ID	:	2RFO
Title	:	Crystral Structure of the nucleoporin Nic96
Authors	:	Schrader, N.; Stelter, P.; Flemming, D.; Kunze, K.; Hurt, E.; Vetter, I.R.
Deposited on	:	2007-10-01
Resolution	:	2.60 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
http://wwpdb.org/validation/2016/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:

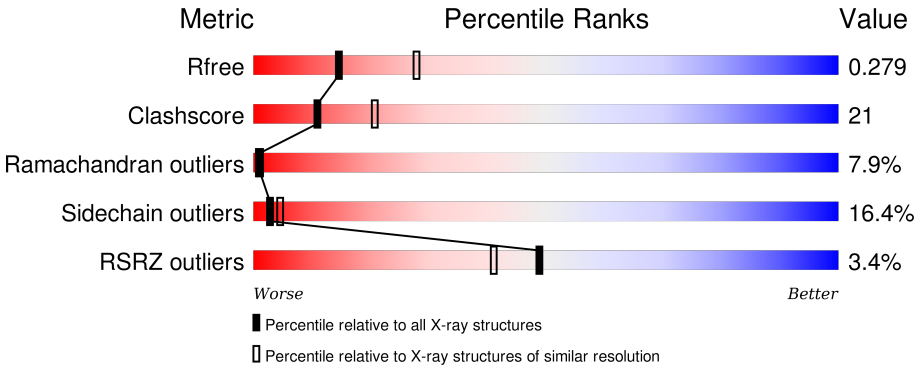
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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))
R_{free}	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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 on the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	651	4% 55% 27% 10% • 6%
1	B	651	3% 50% 29% 12% • 6%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9984 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 Nucleoporin NIC96.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	612	Total	C	N	O	S	0	0	0
			4961	3171	836	937	17			
1	B	609	Total	C	N	O	S	0	0	0
			4935	3156	831	931	17			

- Molecule 2 is water.

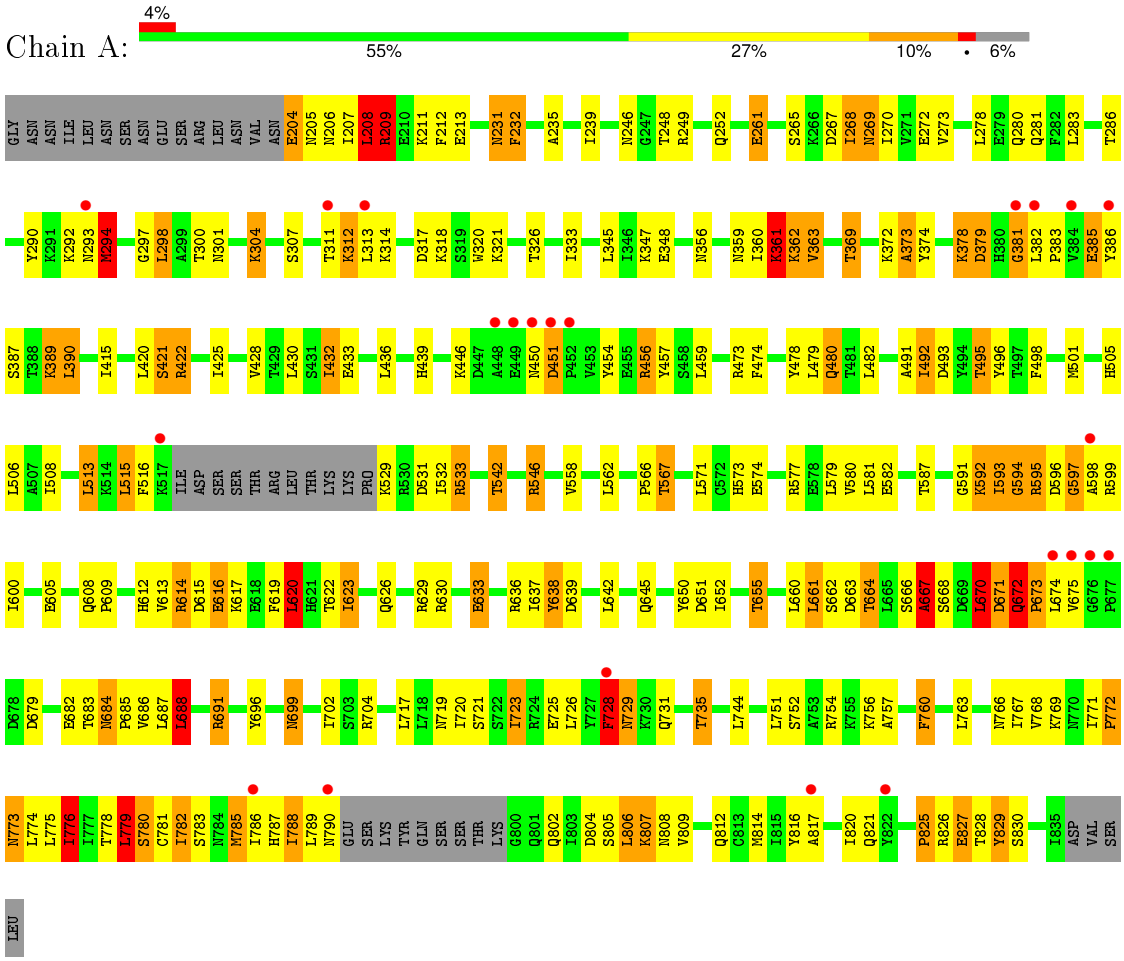
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	37	Total	O	0	0
			37	37		
2	B	51	Total	O	0	0
			51	51		

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 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.

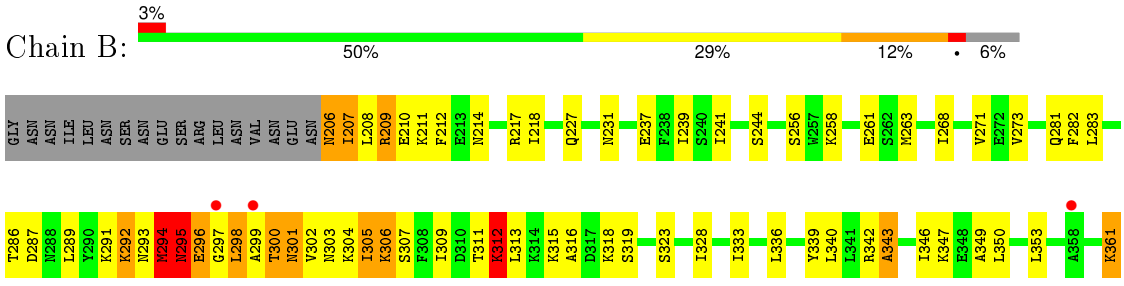
- Molecule 1: Nucleoporin NIC96

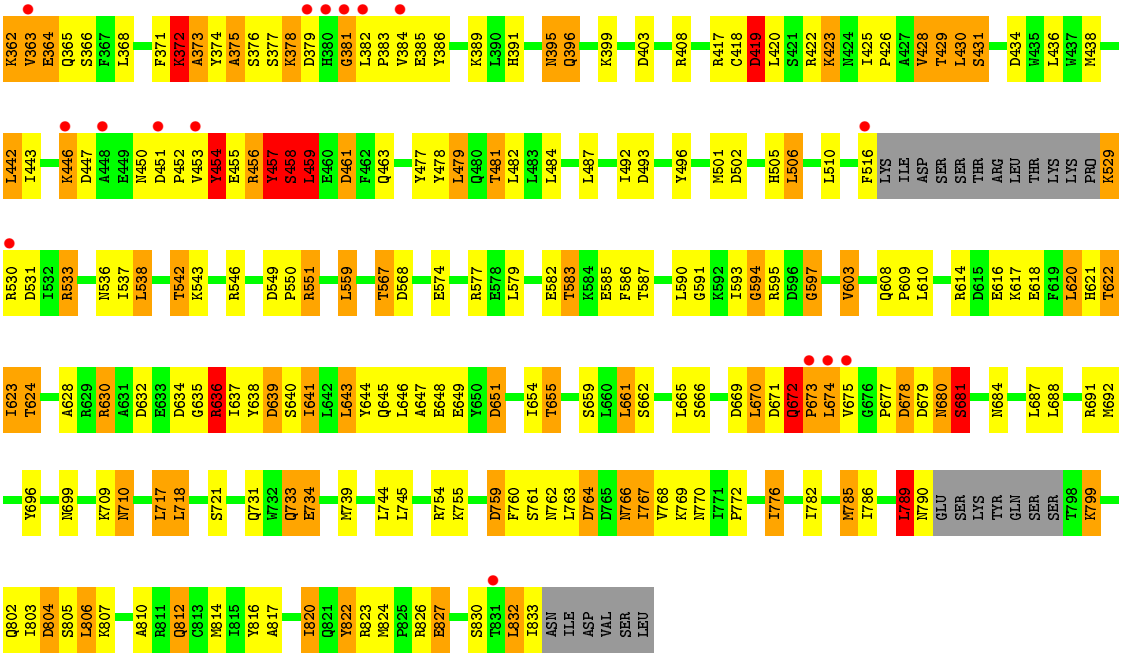
Chain A:



- Molecule 1: Nucleoporin NIC96

Chain B:




K362 V363 E364 Q365 S366 F367 L368 F371 K372 A373 Y374 A375 S376 S377 K378 D379 H380 G381 L382 P383 V384 E385 Y386 K389 L390 H391 N395 Q396 K399 D403 R408 R417 C418 D419 L420 S421 R422 K423 M424 I425 P426 A427 V428 T429 L430 S431 D434 W435 L436 M437 N438
L442 I443 K446 D447 A448 E449 N450 D451 P452 V453 Y454 E455 R456 Y457 S458 L459 E460 D461 F462 Q463 Y477 Y478 L479 Q480 T481 L482 L483 L484 L487 I492 D493 Y496 M501 D502 H505 L506 L510 F516 LYS ILE ASP SER SER THR ARG LEU THR LYS LYS PRO K529
R530 D531 I532 R533 N536 I537 L538 T542 K543 R546 D549 P550 R551 L569 T567 D568 E574 R577 E578 L579 E582 T583 K584 E585 F586 T587 L590 G591 K592 I593 G594 R595 D596 G597 V603 Q608 P609 L610 R614 D615 E616 K617 E618 F619 L620 H621 T622
I623 T624 A628 R629 R630 A631 D632 E633 D634 G635 R636 I637 Y638 D639 S640 I641 L642 L643 Y644 Q645 L646 A647 E648 E649 Y650 D651 I654 T655 S659 L660 L661 S662 L665 S666 D669 L670 D671 Q672 P673 L674 V675 G676 P677 D678 D679 N680 S681 N684 L687 L688 R691 M692
Y696 N699 K709 N710 L717 L718 S721 Q731 W732 Q733 E734 M739 L744 L745 R754 K755 D759 F760 S761 N762 L763 D764 D765 N766 I767 V768 K769 N770 I771 P772 I776 I782 N785 I786 L789 N790 GLU SER LYS TYR GLN SER SER T798 K799
Q802 I803 D804 S805 L806 K807 A810 R811 Q812 C813 M814 I815 Y816 A817 I820 Q821 Y822 R823 M824 P825 R826 E827 S830 T831 L832 I833 ASN ILE ASP VAL SER LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	62.00Å 100.00Å 104.00Å 62.00° 82.00° 86.00°	Depositor
Resolution (Å)	19.78 – 2.60 19.78 – 2.60	Depositor EDS
% Data completeness (in resolution range)	97.8 (19.78-2.60) 81.3 (19.78-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.48 (at 2.59Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.227 , 0.285 0.225 , 0.279	Depositor DCC
R_{free} test set	3313 reflections (5.33%)	DCC
Wilson B-factor (Å^2)	55.9	Xtriage
Anisotropy	0.592	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 48.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.30$	Xtriage
Outliers	0 of 65513 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	9984	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.90% 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, 0.375 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	0.63	3/5048 (0.1%)	0.77	6/6825 (0.1%)
1	B	1.31	19/5022 (0.4%)	0.86	8/6790 (0.1%)
All	All	1.03	22/10070 (0.2%)	0.81	14/13615 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	5
1	B	0	4
All	All	0	9

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	296	GLU	CD-OE1	53.18	1.84	1.25
1	B	597	GLY	CA-C	34.73	2.07	1.51
1	B	296	GLU	CD-OE2	-26.74	0.96	1.25
1	B	296	GLU	CG-CD	19.24	1.80	1.51
1	B	636	ARG	CZ-NH1	17.23	1.55	1.33

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	296	GLU	CG-CD-OE2	13.23	144.77	118.30
1	B	296	GLU	CG-CD-OE1	-11.73	94.85	118.30
1	B	636	ARG	NE-CZ-NH2	-11.17	114.72	120.30
1	B	419	ASP	CB-CG-OD1	-9.25	109.98	118.30
1	A	208	LEU	CA-CB-CG	8.16	134.06	115.30

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	294	MET	Peptide
1	A	361	LYS	Peptide
1	A	599	ARG	Peptide
1	A	672	GLN	Peptide
1	A	728	PHE	Peptide

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	4961	0	5011	196	0
1	B	4935	0	4989	227	0
2	A	37	0	0	0	0
2	B	51	0	0	4	0
All	All	9984	0	10000	418	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 21.

The worst 5 of 418 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:B:296:GLU:CB	1:B:296:GLU:CG	1.74	1.61
1:B:789:LEU:CD1	1:B:789:LEU:CG	1.77	1.61
1:B:296:GLU:CD	1:B:296:GLU:CG	1.80	1.49
1:B:294:MET:CG	1:B:294:MET:SD	2.05	1.44
1:B:294:MET:SD	1:B:294:MET:CE	2.09	1.40

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	606/651 (93%)	503 (83%)	56 (9%)	47 (8%)	1 1
1	B	603/651 (93%)	486 (81%)	68 (11%)	49 (8%)	1 1
All	All	1209/1302 (93%)	989 (82%)	124 (10%)	96 (8%)	1 1

5 of 96 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	209	ARG
1	A	317	ASP
1	A	421	SER
1	A	598	ALA
1	A	614	ARG

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	552/590 (94%)	465 (84%)	87 (16%)	3 5
1	B	549/590 (93%)	455 (83%)	94 (17%)	2 4
All	All	1101/1180 (93%)	920 (84%)	181 (16%)	3 4

5 of 181 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	788	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	305	ILE
1	B	766	ASN
1	A	820	ILE
1	B	261	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	672	GLN
1	A	787	HIS
1	B	731	GLN
1	A	699	ASN
1	A	733	GLN

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^{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	612/651 (94%)	-0.13	23 (3%)	44	36	36, 49, 68, 80	0
1	B	609/651 (93%)	-0.12	19 (3%)	52	45	38, 49, 66, 76	0
All	All	1221/1302 (93%)	-0.13	42 (3%)	49	41	36, 49, 67, 80	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	831	THR	7.8
1	A	382	LEU	6.0
1	A	675	VAL	6.0
1	A	674	LEU	5.0
1	A	676	GLY	4.8

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.