wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 06:57 am BST

PDB ID	:	3JRO
Title	:	NUP84-NUP145C-SEC13 edge element of the NPC lattice
Authors	:	Brohawn, S.G.; Schwartz, T.U.
Deposited on	:	2009-09-08
Resolution	:	4.00 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:

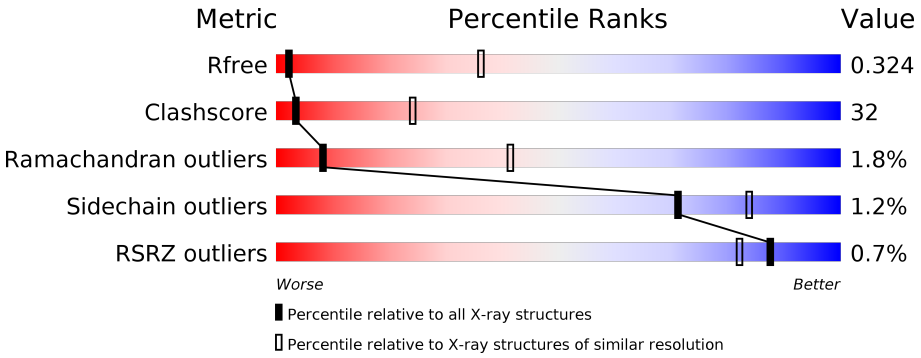
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 4.00 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	1087 (4.30-3.70)
Clashscore	141614	1148 (4.30-3.70)
Ramachandran outliers	138981	1108 (4.30-3.70)
Sidechain outliers	138945	1099 (4.30-3.70)
RSRZ outliers	127900	1028 (4.34-3.66)

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% 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	753	% 54% 38% 7%
2	C	426	35% 50% 11%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8671 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 Fusion Protein of Protein Transport Protein SEC13 and Nucleoporin NUP145.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	701	5623	3598	943	1068	9	5	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1100	GLY	-	LINKER	UNP P49687
A	1101	GLY	-	LINKER	UNP P49687
A	1102	GLY	-	LINKER	UNP P49687
A	1103	GLY	-	LINKER	UNP P49687
A	1104	SER	-	LINKER	UNP P49687
A	1105	GLY	-	LINKER	UNP P49687
A	1106	GLY	-	LINKER	UNP P49687
A	1107	GLY	-	LINKER	UNP P49687
A	1108	GLY	-	LINKER	UNP P49687

- Molecule 2 is a protein called Nucleoporin NUP84.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
2	C	379	3048	1949	504	584	4	7	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

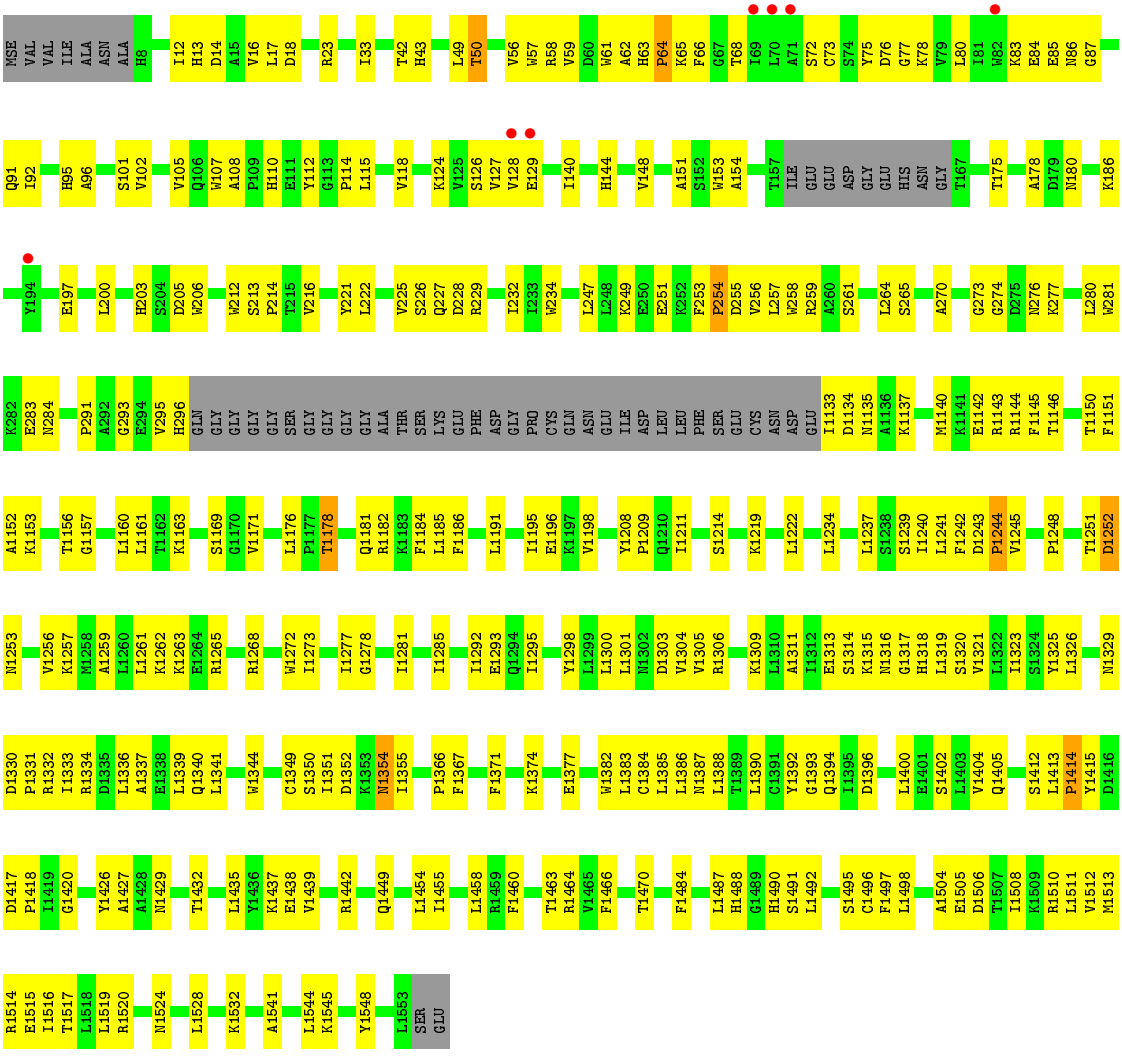
Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	GLY	-	EXPRESSION TAG	UNP P52891
C	0	SER	-	EXPRESSION TAG	UNP P52891

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 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.

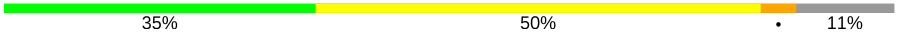
• Molecule 1: Fusion Protein of Protein Transport Protein SEC13 and Nucleoporin NUP145

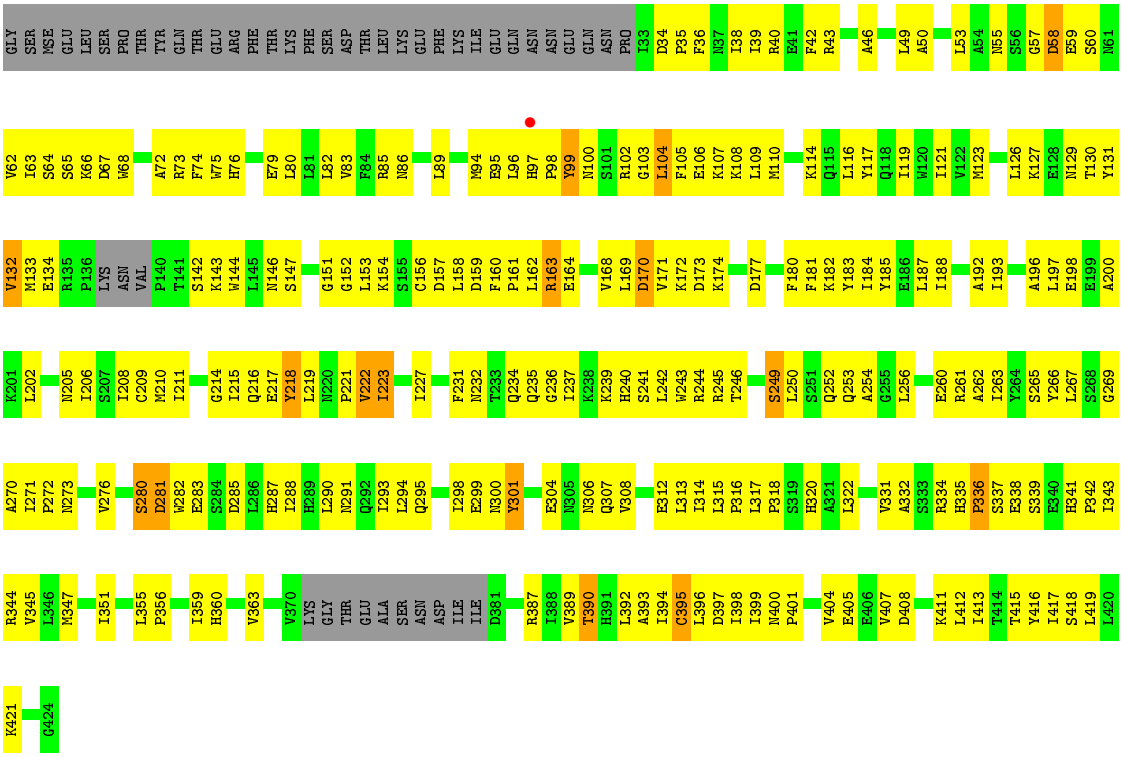
Chain A:



• Molecule 2: Nucleoporin NUP84

Chain C:




GLY SER MSE GLU LEU SER PRO THR TYR GLN THR GLU ARG PHE THR LYS PHE SER ASP THR LEU LYS GLU PHE LYS ILE GLU GLN ASN ASN GLU GLN ASN PRO I33 D34 P35 F36 N37 I38 I39 R40 E41 F42 R43 A46 L49 A50 L53 A54 N55 S56 G57 D58 E59 S60 N61
V62 I63 S64 S65 K66 D67 W68 A72 R73 F74 W75 H76 E79 L80 L81 L82 V83 F84 R85 N86 L89 M94 E95 L96 H97 P98 Y99 N100 S101 R102 G103 L104 F105 E106 K107 K108 L109 M110 K114 Q115 L116 Y117 Q118 I119 W120 I121 V122 M123 L126 K127 E128 N129 T130 Y131
V132 M133 E134 R135 P136 LYS ASN VAL P140 T141 S142 K143 W144 L145 N146 S147 G151 G152 L153 K154 S155 C156 D157 L158 D159 F160 P161 L162 R163 E164 V168 L169 D170 V171 K172 D173 K174 D177 F180 F181 K182 Y183 I184 Y185 E186 L187 I188 A192 I193 A196 L197 E198 E199 A200
K201 L202 N205 I206 S207 I208 C209 M210 I211 G214 I215 Q216 E217 Y218 L219 N220 P221 V222 I223 I227 F231 N232 T233 Q234 Q235 G236 I237 K238 K239 H240 S241 L242 W243 R244 R245 T246 S249 L250 S251 Q252 Q253 A254 G255 L256 E260 R261 A262 I263 Y264 S265 Y266 L267 S268 G269
A270 I271 P272 N273 V276 S280 D281 W282 E283 S284 D285 L286 H287 I288 H289 L290 N291 Q292 I293 L294 Q295 I298 E299 N300 Y301 E304 N305 N306 Q307 V308 E312 L313 I314 L315 P316 L317 P318 S319 H320 A321 L322 V331 A332 S333 R334 H335 P336 S337 E338 S339 E340 H341 P342 I343
R344 V345 L346 M347 I351 L355 P356 I359 H360 V363 V370 LYS GLY THR GLU ALA SER ASN ASP ILE ILE ILE D381 R387 I388 V389 T390 H391 L392 A393 I394 C395 L396 D397 I398 I399 N400 P401 V404 E405 E406 V407 D408 K411 L412 I413 T414 T415 Y416 I417 S418 L419 L420
K421 G424

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 62 2 2	Depositor
Cell constants a, b, c, α, β, γ	170.47Å 170.47Å 270.73Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.89 – 4.00 49.89 – 4.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (49.89-4.00) 99.9 (49.89-4.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.77 (at 4.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine)	Depositor
R, R_{free}	0.282 , 0.329 0.274 , 0.324	Depositor DCC
R_{free} test set	1586 reflections (7.83%)	wwPDB-VP
Wilson B-factor (Å^2)	133.7	Xtriage
Anisotropy	0.702	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 123.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	8671	wwPDB-VP
Average B, all atoms (Å^2)	188.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.09% 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	0.24	0/5744	0.41	0/7776
2	C	0.28	0/3101	0.49	0/4202
All	All	0.25	0/8845	0.44	0/11978

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

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	5623	0	5577	286	0
2	C	3048	0	3033	306	0
All	All	8671	0	8610	555	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 32.

The worst 5 of 555 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:C:104:LEU:HA	2:C:219:LEU:HD12	1.19	1.09
2:C:215:ILE:HG21	2:C:267:LEU:HD13	1.31	1.08

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1318:HIS:HB2	2:C:160:PHE:HB2	1.44	0.99
2:C:412:LEU:HB3	2:C:416:TYR:HE2	1.32	0.94
1:A:1251:THR:H	1:A:1257:LYS:HE2	1.35	0.91

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	695/753 (92%)	608 (88%)	78 (11%)	9 (1%)	12 48
2	C	373/426 (88%)	297 (80%)	66 (18%)	10 (3%)	5 34
All	All	1068/1179 (91%)	905 (85%)	144 (14%)	19 (2%)	8 41

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	218	TYR
2	C	281	ASP
2	C	55	ASN
2	C	171	VAL
2	C	222	VAL

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	621/655 (95%)	618 (100%)	3 (0%)	88 93
2	C	345/384 (90%)	336 (97%)	9 (3%)	46 67
All	All	966/1039 (93%)	954 (99%)	12 (1%)	71 84

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	163	ARG
2	C	170	ASP
2	C	301	TYR
2	C	99	TYR
2	C	280	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	234	GLN
2	C	360	HIS
2	C	273	ASN
2	C	118	GLN
2	C	240	HIS

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	696/753 (92%)	-0.23	7 (1%)	82	74	159, 188, 244, 325	0
2	C	372/426 (87%)	-0.37	1 (0%)	94	90	134, 171, 248, 370	0
All	All	1068/1179 (90%)	-0.28	8 (0%)	87	82	134, 183, 246, 370	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	70	LEU	3.0
1	A	71	ALA	2.7
1	A	194	TYR	2.6
1	A	82	TRP	2.6
1	A	69	ILE	2.3

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.