wwPDB X-ray Structure Validation Summary Report (i)

Aug 8, 2023 – 09:47 PM EDT

PDB ID	:	4FHN
Title	:	Nup37-Nup120 full-length complex from Schizosaccharomyces pombe
Authors	:	Bilokapic, S.; Schwartz, T.U.
Deposited on	:	2012-06-06
Resolution	:	6.99 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 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:

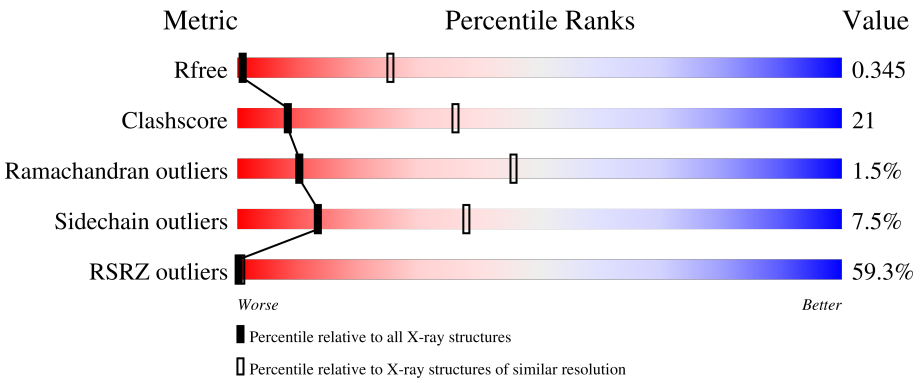
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 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 6.99 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	1004 (10.00-3.90)
Clashscore	141614	1069 (10.00-3.90)
Ramachandran outliers	138981	1002 (10.00-3.90)
Sidechain outliers	138945	1002 (10.00-3.86)
RSRZ outliers	127900	1004 (9.50-3.80)

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	394	68%; 56%, 28%, •, 13%
1	C	394	80%; 56%, 28%, •, 13%
2	B	1139	57%; 45%, 40%, 5%, 10%
2	D	1139	44%; 45%, 37%, •, 14%
3	X	450	27%; 59%, 38%, •, •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 24772 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 NUP37.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	343	2638	1676	447	500	15	0	0	0
1	C	344	2646	1680	449	502	15	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	PRO	-	expression tag	UNP O36030
A	-1	GLY	-	expression tag	UNP O36030
A	0	SER	-	expression tag	UNP O36030
C	-2	PRO	-	expression tag	UNP O36030
C	-1	GLY	-	expression tag	UNP O36030
C	0	SER	-	expression tag	UNP O36030

- Molecule 2 is a protein called Nucleoporin nup120.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1022	8251	5335	1322	1563	31	0	0	0
2	D	977	7871	5094	1258	1488	31	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	PRO	-	expression tag	UNP O43044
B	-1	GLY	-	expression tag	UNP O43044
B	0	SER	-	expression tag	UNP O43044
D	-2	PRO	-	expression tag	UNP O43044
D	-1	GLY	-	expression tag	UNP O43044
D	0	SER	-	expression tag	UNP O43044

- Molecule 3 is a protein called Glutamate dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	X	442	3366	2119	594	632	21	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

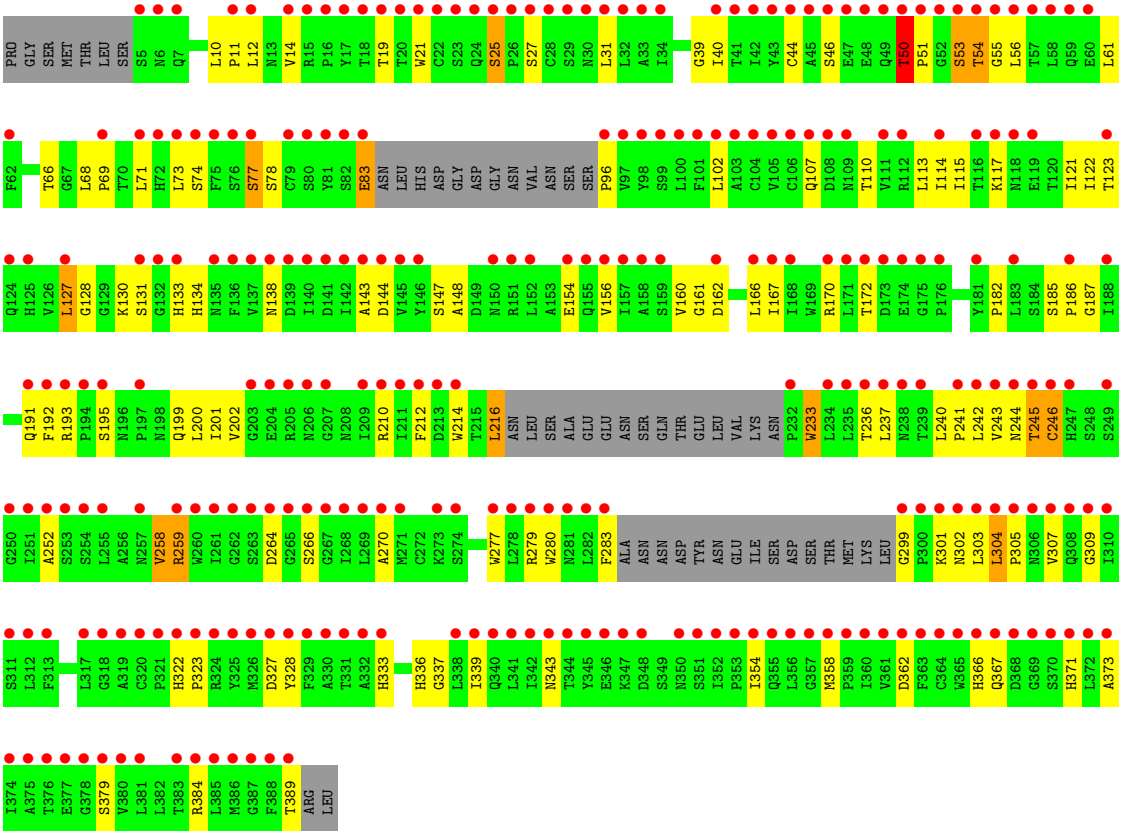
Chain	Residue	Modelled	Actual	Comment	Reference
X	-2	PRO	-	expression tag	UNP Q8XDW9
X	-1	GLY	-	expression tag	UNP Q8XDW9
X	0	SER	-	expression tag	UNP Q8XDW9

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.

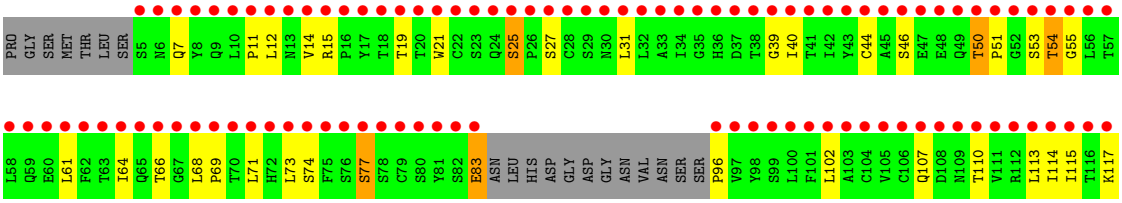
- Molecule 1: NUCLEOPORIN NUP37

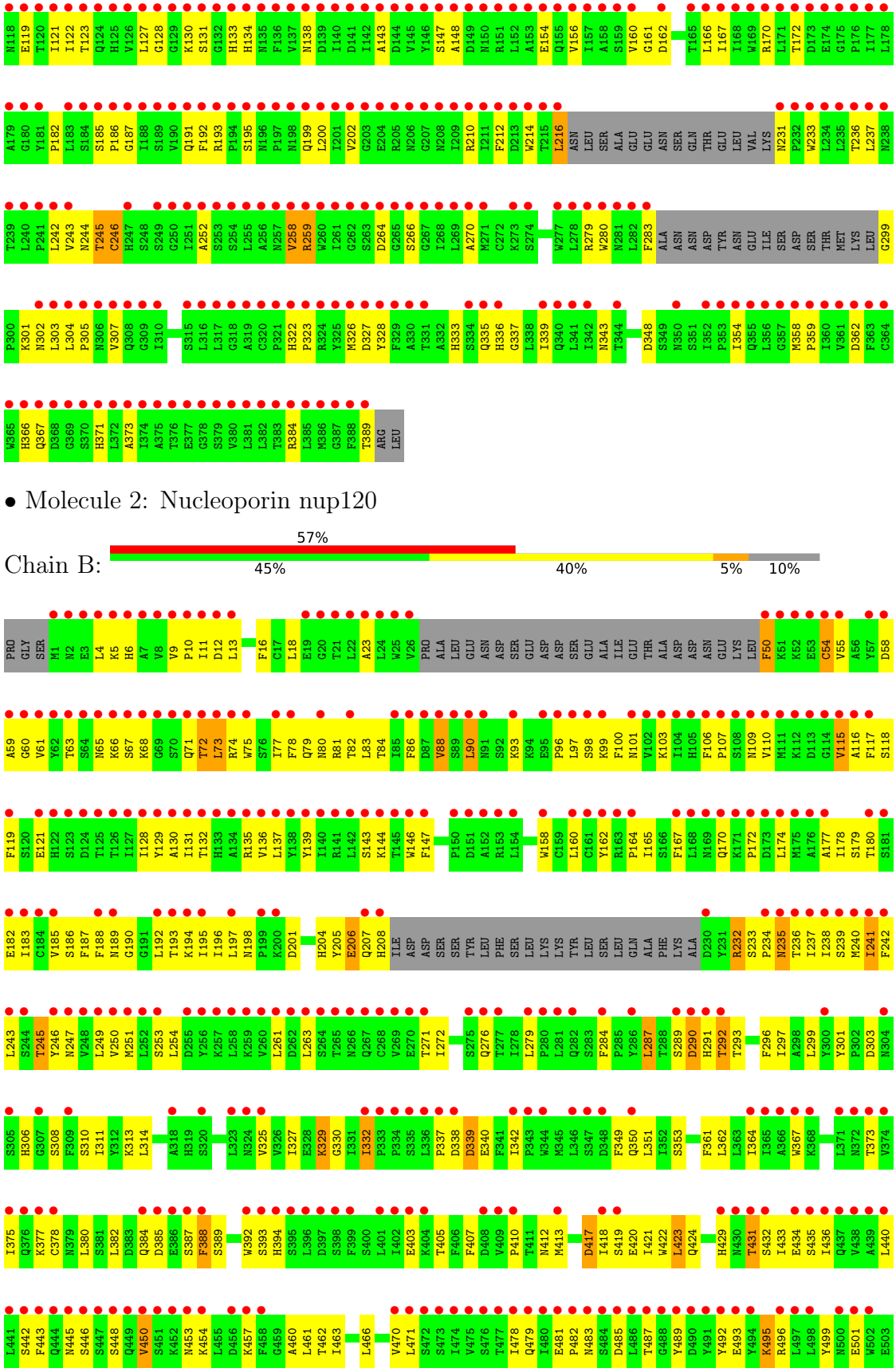
Chain A:



- Molecule 1: NUCLEOPORIN NUP37

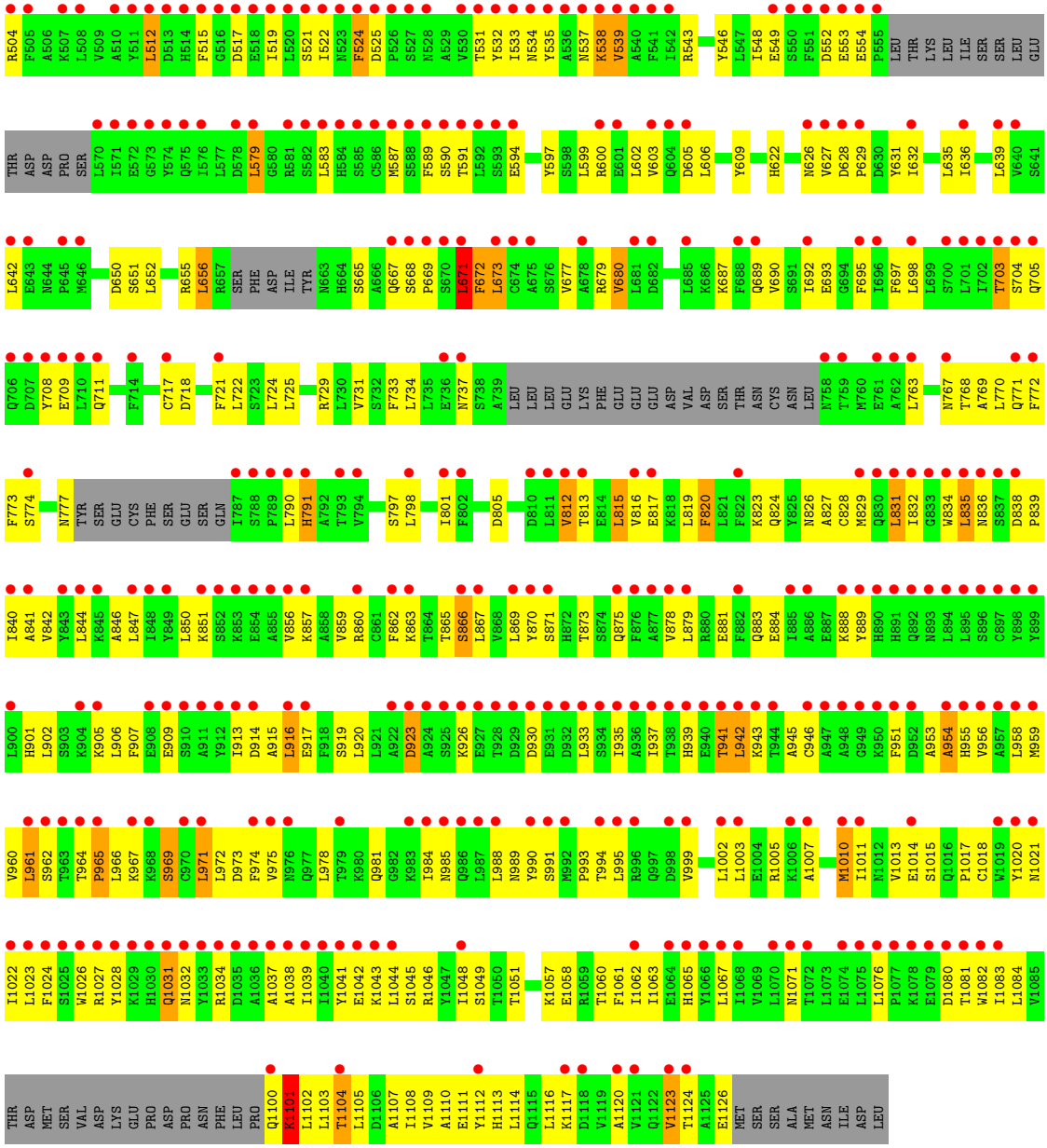
Chain C:





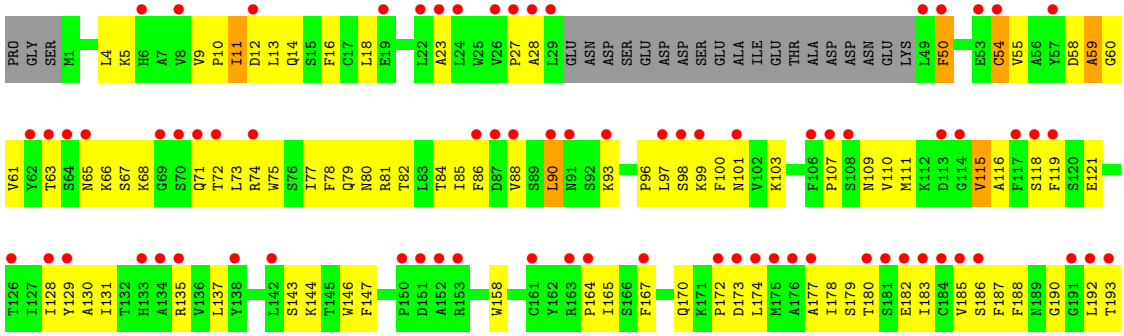
- Molecule 2: Nucleoporin nup120

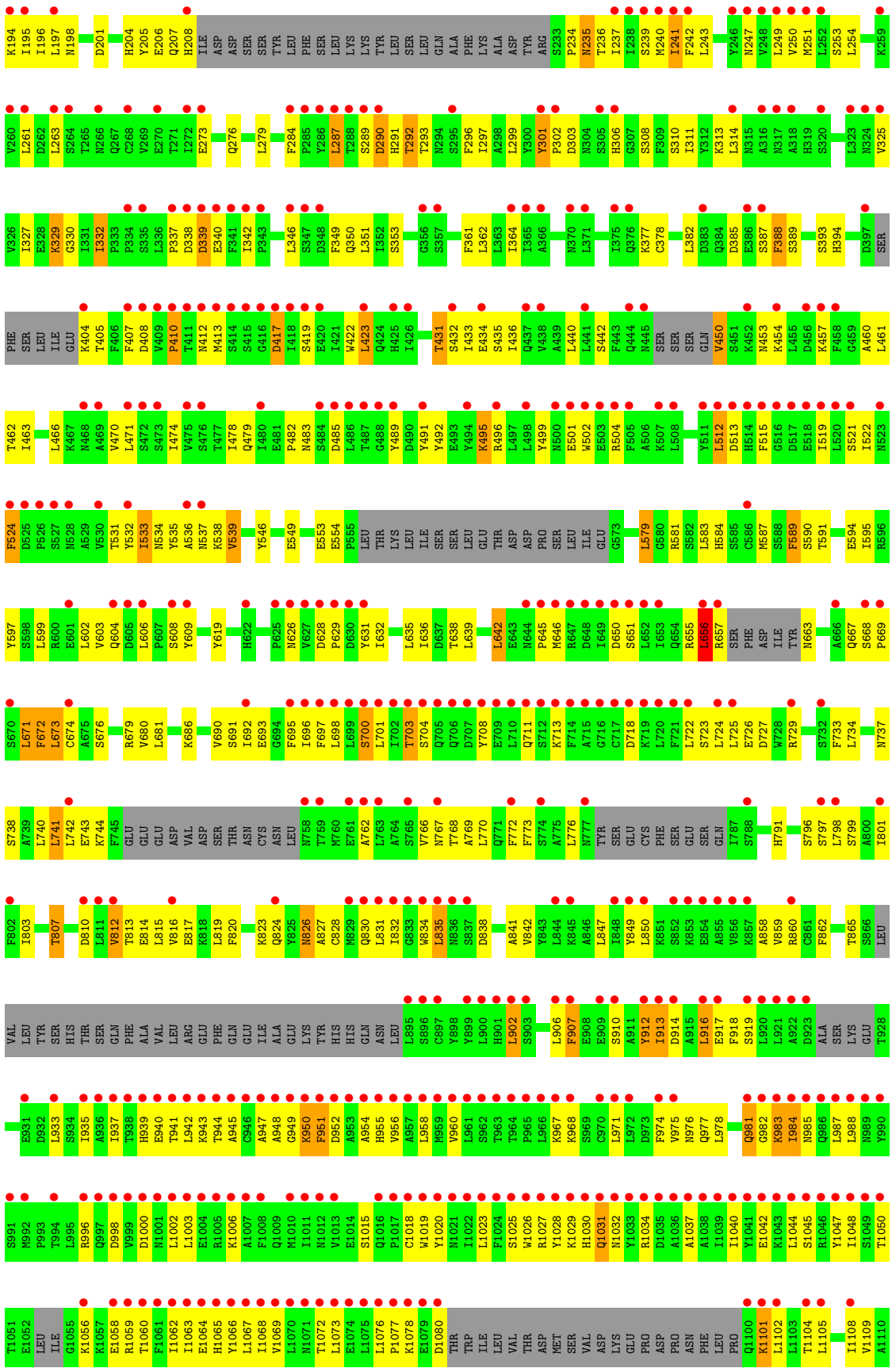
Chain B:

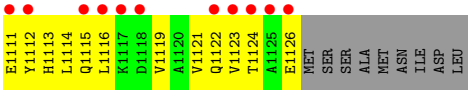


- Molecule 2: Nucleoporin nup120

Chain D:

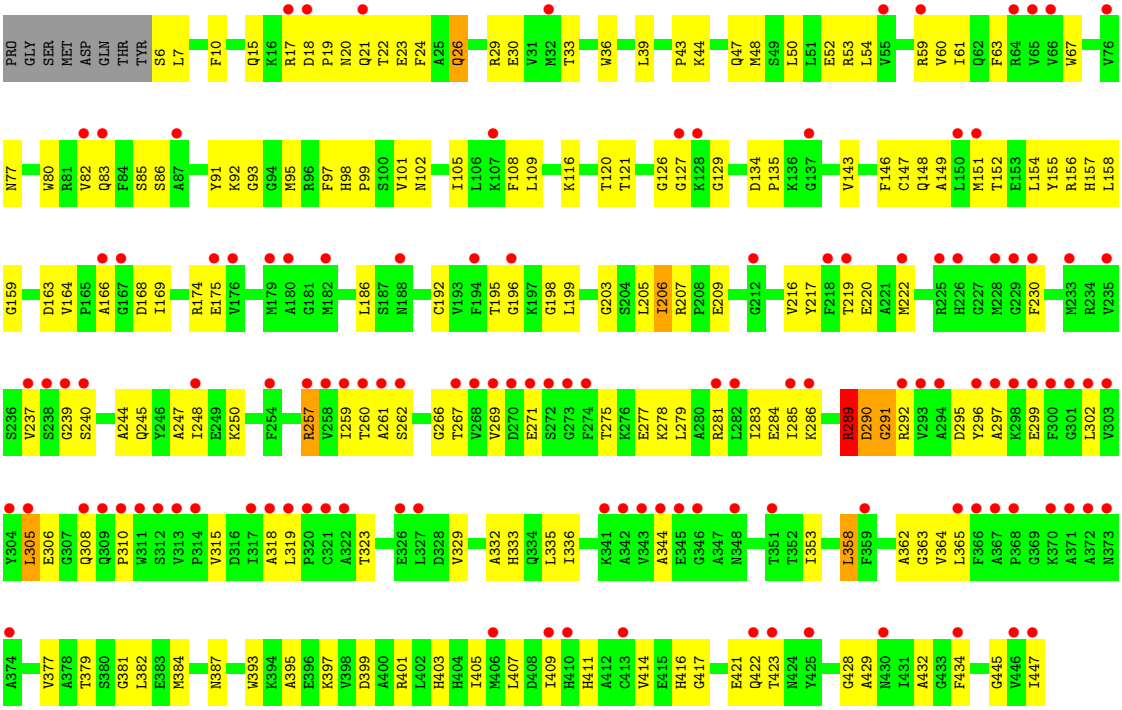






• Molecule 3: Glutamate dehydrogenase

Chain X:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, α, β, γ	330.00Å 330.00Å 350.26Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	95.12 – 6.99 108.02 – 6.99	Depositor EDS
% Data completeness (in resolution range)	99.5 (95.12-6.99) 99.5 (108.02-6.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.27	Depositor
$< I/\sigma(I) >$ [1]	1.29 (at 6.73Å)	Xtriage
Refinement program	PHENIX 1.7.3_928	Depositor
R, R_{free}	0.285 , 0.346 0.280 , 0.345	Depositor DCC
R_{free} test set	939 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å^2)	456.0	Xtriage
Anisotropy	0.094	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 684.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.78	EDS
Total number of atoms	24772	wwPDB-VP
Average B, all atoms (Å^2)	672.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.23% 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)

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	0.33	0/2702	0.65	0/3689
1	C	0.29	0/2710	0.54	0/3701
2	B	0.37	0/8433	0.71	8/11445 (0.1%)
2	D	0.32	0/8039	0.57	1/10903 (0.0%)
3	X	0.41	0/3431	0.59	0/4630
All	All	0.35	0/25315	0.63	9/34368 (0.0%)

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	3
2	B	0	1
2	D	0	1
All	All	0	5

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	245	THR	N-CA-C	-7.16	91.68	111.00
2	B	72	THR	N-CA-C	6.09	127.44	111.00
2	B	916	LEU	CA-CB-CG	5.79	128.62	115.30
2	B	671	LEU	CA-CB-CG	5.70	128.41	115.30
2	B	73	LEU	CA-CB-CG	5.65	128.31	115.30

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	233	TRP	Peptide
1	A	50	THR	Peptide
1	A	53	SER	Peptide
2	B	206	GLU	Peptide
2	D	206	GLU	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	2638	0	2588	98	0
1	C	2646	0	2593	93	0
2	B	8251	0	8211	424	2
2	D	7871	0	7847	365	2
3	X	3366	0	3324	118	2
All	All	24772	0	24563	1038	4

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 1038 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:828:CYS:HA	2:B:831:LEU:HB2	1.40	1.01
2:B:174:LEU:HD11	2:B:240:MET:H	1.31	0.95
2:D:744:LYS:H	2:D:823:LYS:HD3	1.32	0.95
1:A:252:ALA:HB2	2:B:826:ASN:HD22	1.39	0.88
2:B:942:LEU:HD11	2:B:961:LEU:HG	1.56	0.86

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:189:ASN:O	2:D:404:LYS:NZ[8_555]	2.07	0.13
3:X:156:ARG:NH1	3:X:186:LEU:O[12_544]	2.13	0.07
3:X:155:TYR:OH	3:X:186:LEU:O[12_544]	2.16	0.04

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:384:GLN:NE2	2:D:273:GLU:OE1[6_554]	2.17	0.03

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	335/394 (85%)	315 (94%)	19 (6%)	1 (0%)	41	77
1	C	336/394 (85%)	319 (95%)	16 (5%)	1 (0%)	41	77
2	B	1006/1139 (88%)	869 (86%)	117 (12%)	20 (2%)	7	38
2	D	951/1139 (84%)	821 (86%)	108 (11%)	22 (2%)	6	34
3	X	440/450 (98%)	420 (96%)	17 (4%)	3 (1%)	22	63
All	All	3068/3516 (87%)	2744 (89%)	277 (9%)	47 (2%)	10	46

5 of 47 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	965	PRO
2	D	951	PHE
3	X	206	ILE
3	X	289	ARG
1	A	50	THR

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	299/345 (87%)	278 (93%)	21 (7%)	15	40
1	C	300/345 (87%)	280 (93%)	20 (7%)	16	41
2	B	940/1050 (90%)	859 (91%)	81 (9%)	10	32
2	D	897/1050 (85%)	828 (92%)	69 (8%)	13	37
3	X	347/354 (98%)	330 (95%)	17 (5%)	25	50
All	All	2783/3144 (88%)	2575 (92%)	208 (8%)	13	38

5 of 208 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	216	LEU
2	D	417	ASP
3	X	305	LEU
1	C	259	ARG
2	D	235	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	528	ASN
3	X	117	ASN
2	B	989	ASN
2	B	1100	GLN
1	C	231	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 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 (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	343/394 (87%)	3.79	268 (78%)	0	0	358, 705, 903, 929	0
1	C	344/394 (87%)	7.37	315 (91%)	0	0	5, 817, 978, 997	0
2	B	1022/1139 (89%)	2.96	650 (63%)	0	0	273, 650, 901, 955	0
2	D	977/1139 (85%)	2.77	500 (51%)	0	1	298, 714, 927, 984	0
3	X	442/450 (98%)	1.53	121 (27%)	0	2	219, 451, 760, 881	0
All	All	3128/3516 (88%)	3.27	1854 (59%)	0	0	5, 667, 929, 997	0

The worst 5 of 1854 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	114	ILE	29.6
1	C	33	ALA	22.1
1	C	102	LEU	21.8
1	C	65	GLN	21.3
1	C	66	THR	20.2

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 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.