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 (i)) were used in the production of this report:

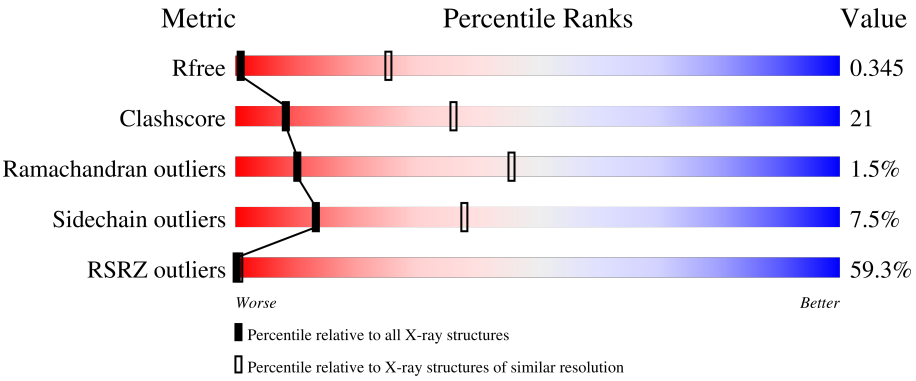
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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
1	A	343	Total	C	N	O	S	0	0	0
			2638	1676	447	500	15			
1	C	344	Total	C	N	O	S	0	0	0
			2646	1680	449	502	15			

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
2	B	1022	Total	C	N	O	S	0	0	0
			8251	5335	1322	1563	31			
2	D	977	Total	C	N	O	S	0	0	0
			7871	5094	1258	1488	31			

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
3	X	442	Total	C	N	O	S	0	0	0
			3366	2119	594	632	21			

There are 3 discrepancies between the modelled and reference sequences:

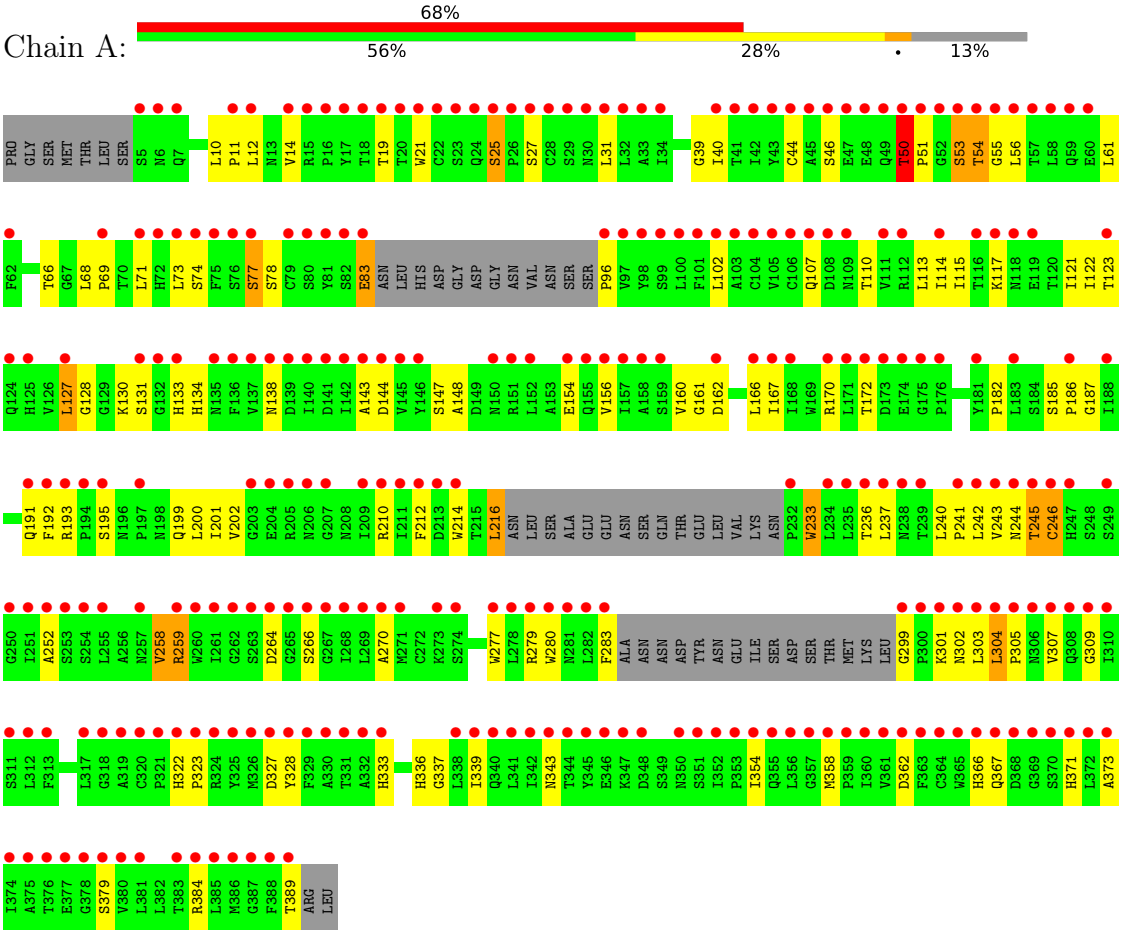
Chain	Residue	Modelled	Actual	Comment	Reference
X	-2	PRO	-	expression tag	UNP Q8XDW9
X	-1	GLY	-	expression tag	UNP Q8XDW9
X	0	SER	-	expression tag	UNP Q8XDW9

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.

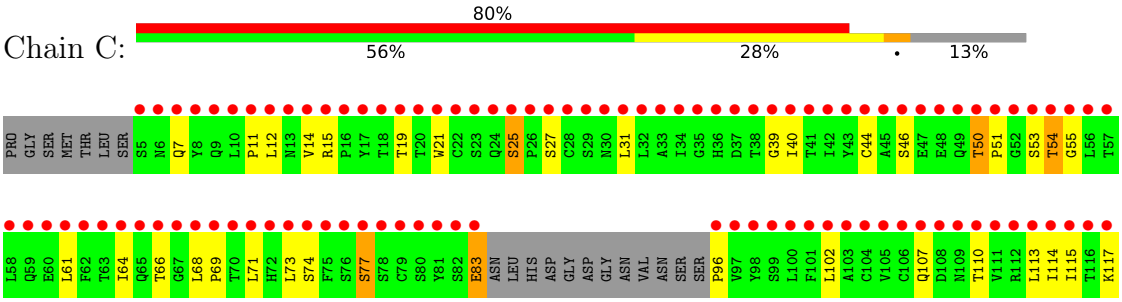
• Molecule 1: NUCLEOPORIN NUP37

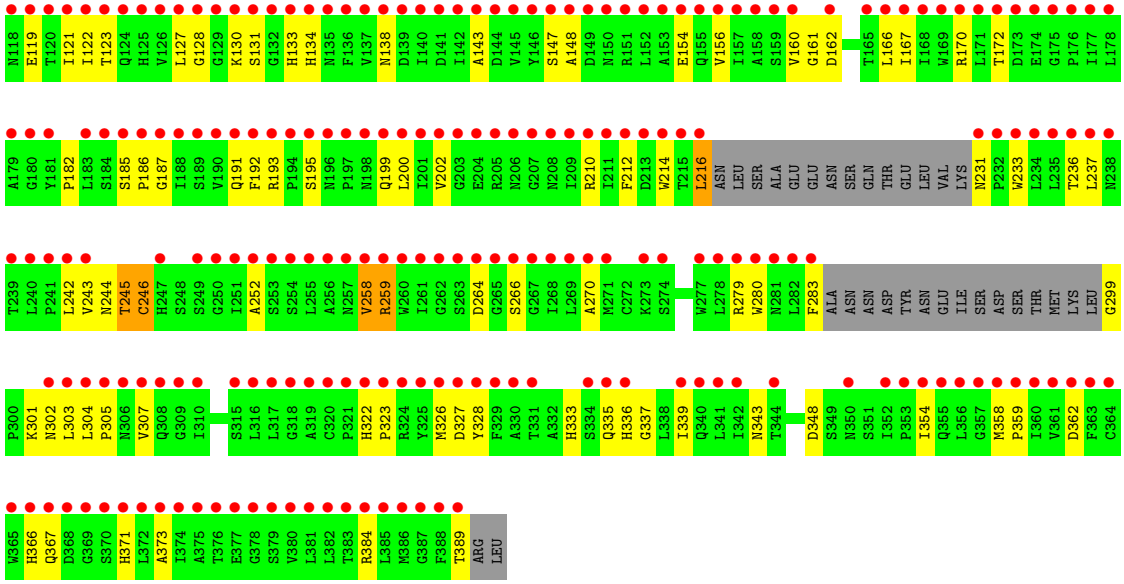
Chain A:



• Molecule 1: NUCLEOPORIN NUP37

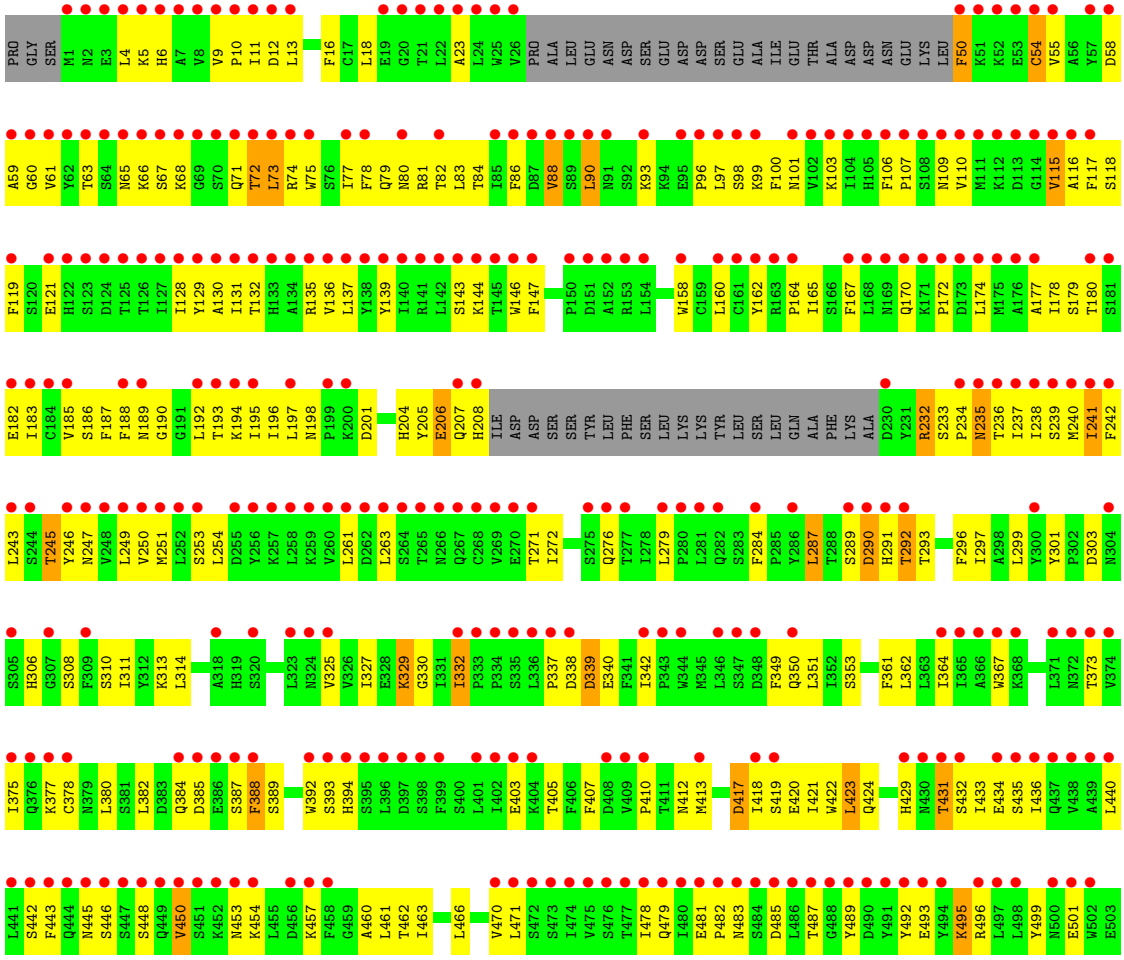
Chain C:

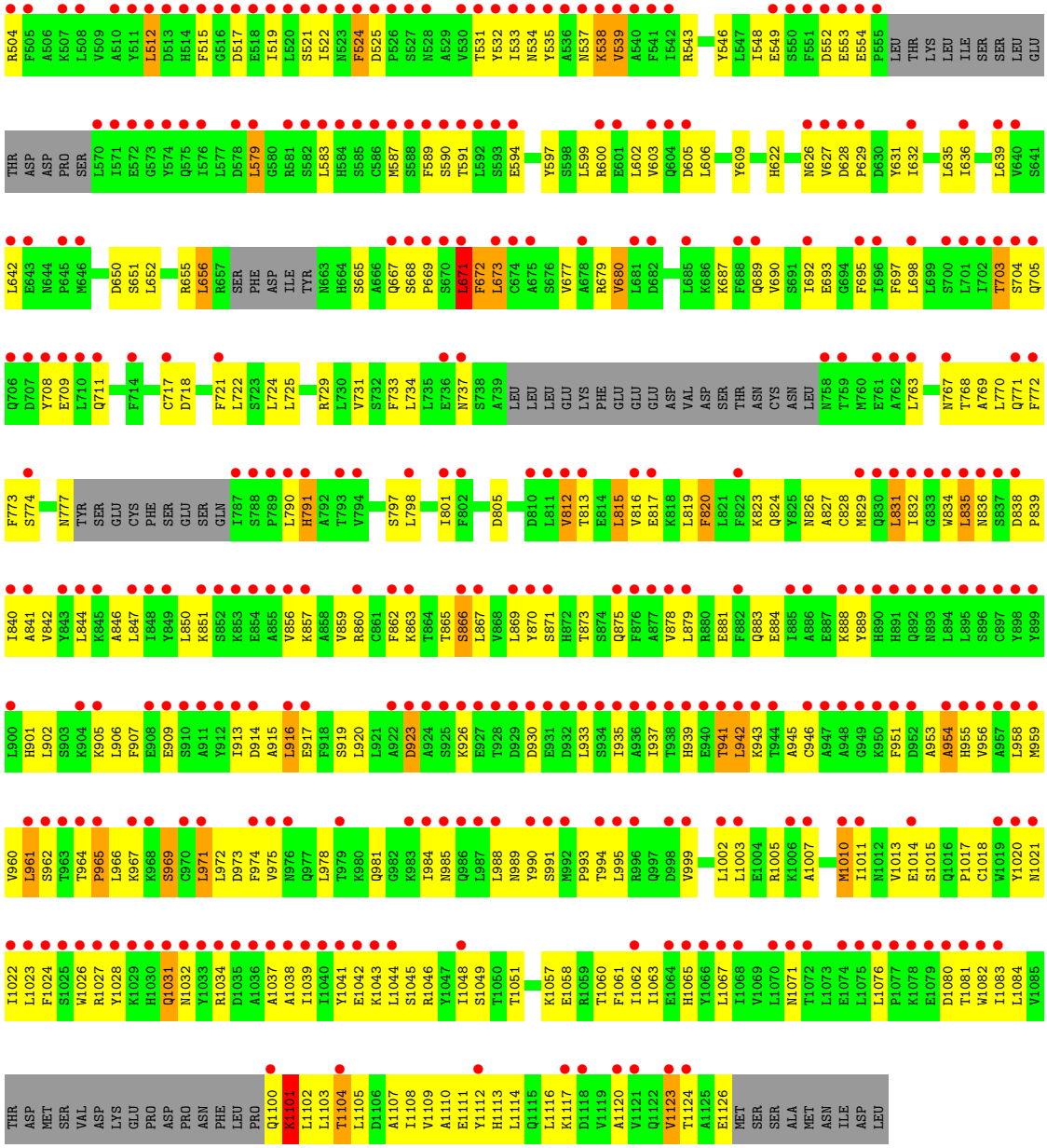




- Molecule 2: Nucleoporin nup120

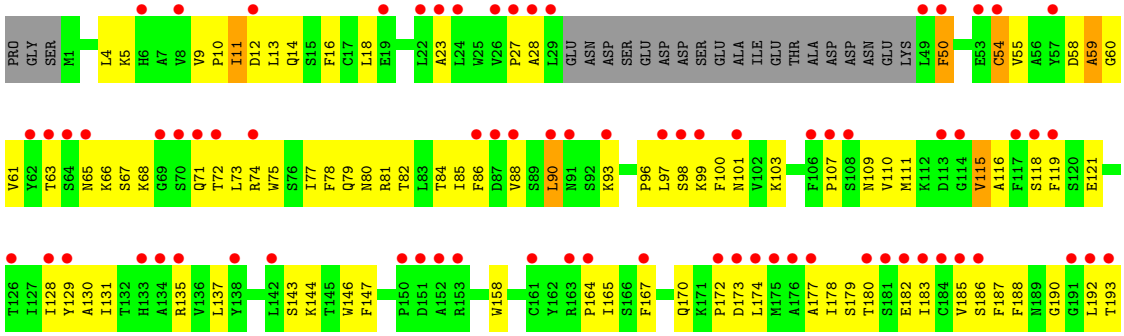
Chain B:

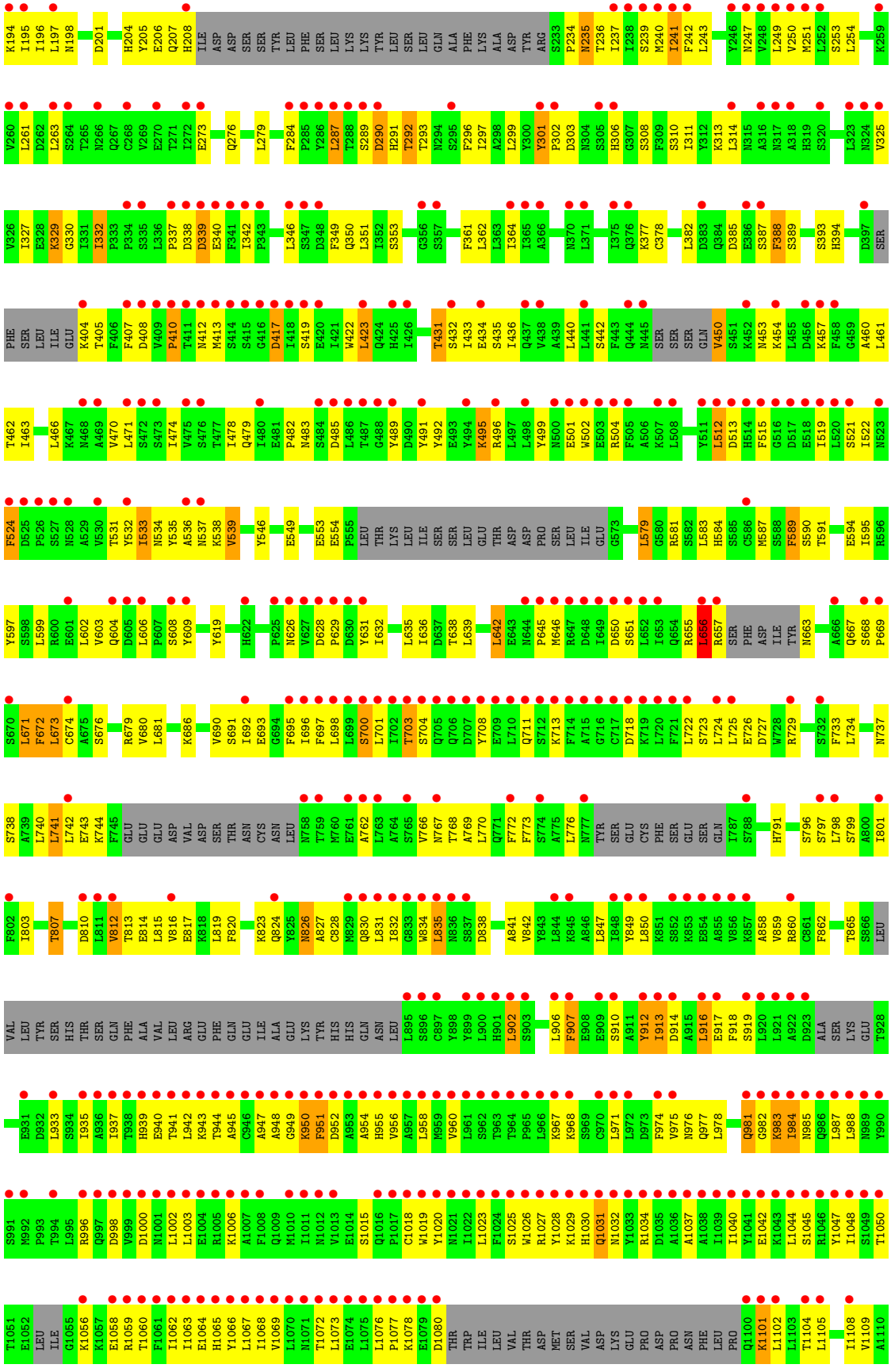


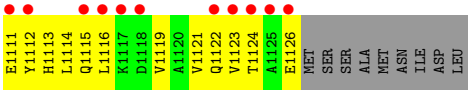


• Molecule 2: Nucleoporin nup120

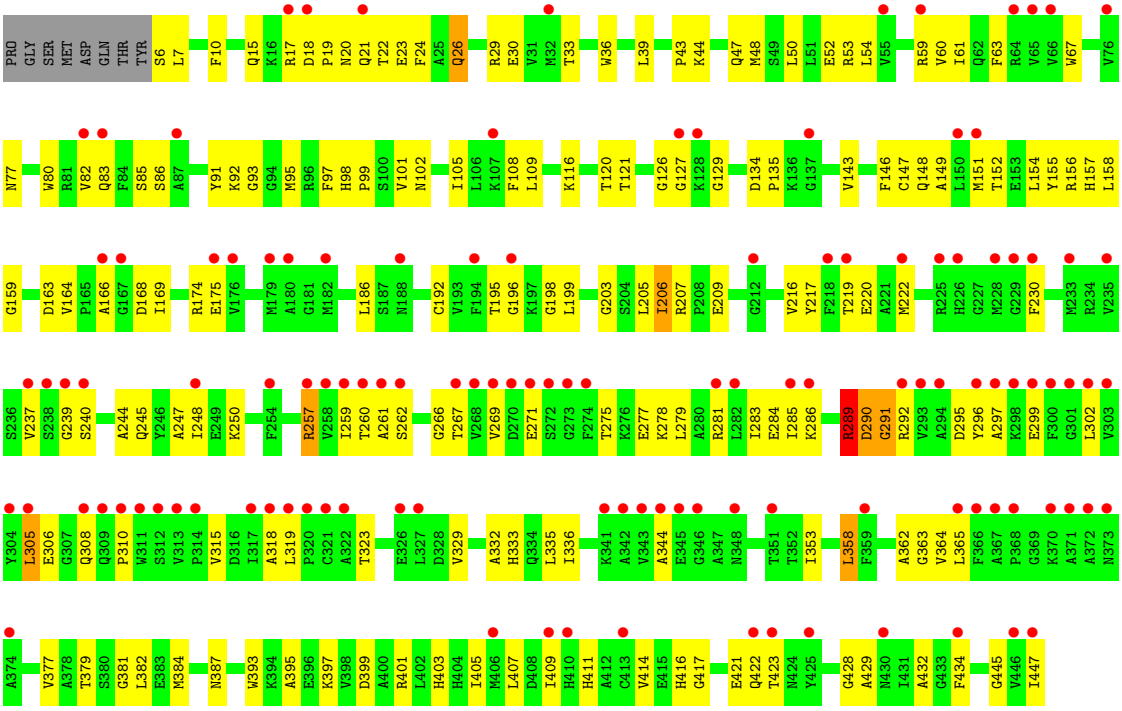
Chain D:




K194 I195 I196 L197 N198 D201 H204 Y205 E206 Q207 H208 ILE ASP ASP SER SER TYR LEU PHE SER LEU LYS LYS TYR LEU SER LEU GLN ALA PHE LYS ALA ASP TYR ARG S233 P234 N235 T236 I237 I238 S239 M240 I241 F242 L243 Y246 N247 V248 L249 V250 M251 L252 S253 L254 K259
V260 L261 D262 L263 S264 T265 N266 Q267 C268 V269 E270 T271 I272 E273 Q276 L279 F284 P285 Y286 L287 T288 S289 D290 H291 T292 T293 N294 S295 F296 I297 A298 L299 Y300 Y301 P302 D303 N304 S305 H306 G307 S308 F309 S310 I311 Y312 K313 L314 N315 A316 N317 A318 H319 S320 L323 N324 V325
V326 I327 E328 K329 G330 I331 I332 P333 P334 S335 L336 P337 D338 D339 E340 F341 I342 P343 L346 S347 D348 F349 Q350 L351 I352 S353 G356 S357 F361 L362 L363 I364 I365 A366 N370 L371 I375 Q376 K377 C378 L382 D383 Q384 D385 E386 S387 F388 S389 S393 H394 D397 SER
PHE SER LEU ILE GLU K404 T405 F406 F407 D408 V409 P410 T411 N412 M413 S414 S415 G416 D417 I418 S419 E420 I421 W422 L423 Q424 H425 I426 T431 S432 I433 E434 S435 I436 Q437 V438 A439 L440 L441 S442 F443 Q444 N445 SER SER SER GLN V450 S451 K452 N453 K454 L455 D456 K457 F458 G459 A460 L461
T462 I463 L466 K467 N468 A469 V470 Q471 S472 S473 I474 V475 S476 T477 I478 Q479 I480 E481 P482 N483 S484 D485 L486 T487 G488 Y489 D490 Y491 Y492 E493 Y494 K495 R496 L497 L498 Y499 N500 E501 W502 E503 R504 F505 A506 K507 L508 Y511 L512 D513 H514 F515 G516 D517 E518 I519 L520 S521 I522 N523
F524 D525 P526 S527 N528 A529 V530 T531 Y532 I533 N534 Y535 A536 N537 K538 V539 Y546 E549 E553 E554 P555 LEU THR LYS LEU ILE SER SER LEU GLU THR ASP ASP PRO SER LEU ILE GLU G573 L579 G580 R581 S582 L583 H584 S585 C586 M587 S588 P589 S590 T591 E594 I595 R596
Y597 S598 L599 R600 E601 L602 V603 Q604 D605 L606 P607 S608 Y609 Y619 H622 P625 N626 V627 D628 P629 D630 Y631 I632 L635 I636 D637 T638 L639 L642 E643 N644 P645 M646 R647 D648 I649 D650 S651 L652 I653 Q654 R655 L656 R657 SER PHE ASP ILE TYR N663 A666 Q667 S668 P669
S670 L671 F672 L673 C674 A675 S676 R679 V680 L681 K686 V690 S691 I692 E693 G694 F695 I696 F697 L698 L699 S700 L701 I702 T703 S704 Q705 Q706 D707 Y708 E709 L710 Q711 S712 K713 F714 A715 G716 C717 D718 K719 L720 F721 L722 S723 L724 L725 E726 D727 W728 R729 S732 F733 L734 N737
S738 A739 L740 L741 L742 E743 K744 F745 GLU GLU GLU ASP VAL ASP SER THR ASN CYS ASN LEU N758 T759 M760 E761 A762 L763 A764 S765 V766 N767 T768 A769 L770 Q771 F772 F773 S774 A775 L776 N777 TYR SER GLU CYS PHE SER GLU SER GLN I787 S788 H791 S796 S797 L798 S799 A800 I801
F802 I803 T807 D810 L811 V812 T813 E814 L815 V816 E817 K818 L819 F820 K823 Q824 Y825 N826 A827 C828 M829 Q830 L831 I832 G833 W834 L835 N836 S837 D838 A841 V842 Y843 L844 K845 A846 L847 I848 Y849 L850 K851 S852 K853 E854 A855 V856 K857 A858 V859 R860 C861 F862 T865 S866 LEU
VAL LEU TYR SER HIS THR SER GLN PHE ALA VAL LEU ARG GLU PHE GLN GLU ILE ALA GLU LYS TYR HIS HIS GLN ASN LEU L895 S896 C897 Y898 Y899 L900 H901 L902 S903 L906 F907 E908 E909 S910 A911 Y912 I913 D914 A915 L916 E917 F918 S919 L920 L921 A922 D923 ALA SER LYS GLU T928
E931 D932 L933 S934 I935 A936 I937 T938 H939 E940 T941 L942 K943 T944 A945 C946 A947 A948 G949 K950 F951 D952 A953 A954 H955 V956 A957 L958 M959 V960 L961 S962 T963 T964 P965 L966 K967 K968 S969 C970 L971 L972 D973 F974 V975 N976 Q977 L978 Q981 G982 K983 I984 N985 Q986 L987 L988 N989 Y990
S991 M992 P993 T994 L995 R996 Q997 D998 Y999 D1000 N1001 L1002 L1003 E1004 R1005 K1006 A1007 F1008 Q1009 M1010 I1011 N1012 V1013 E1014 S1015 Q1016 P1017 C1018 W1019 Y1020 N1021 I1022 L1023 F1024 S1025 W1026 R1027 K1028 K1029 H1030 Q1031 N1032 Y1033 R1034 D1035 A1036 A1037 A1038 I1039 I1040 Y1041 E1042 K1043 L1044 S1045 R1046 Y1047 I1048 S1049 T1050
T1051 E1052 LEU ILE G1055 K1056 R1057 E1058 R1059 T1060 F1061 I1062 I1063 E1064 H1065 Y1066 L1067 I1068 V1069 L1070 N1071 T1072 L1073 E1074 L1075 L1076 P1077 K1078 E1079 D1080 THR TRP ILE LEU VAL THR ASP MET SER VAL ASP LYS GLU PRO ASP PRO ASN PHE LEU PRO Q1100 K1101 L1102 L1103 T1104 L1105 I1108 V1109 A1110



• Molecule 3: Glutamate dehydrogenase



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

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	0.33	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

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	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

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.