Full wwPDB X-ray Structure Validation Report ⓘ

Feb 12, 2024 – 01:25 AM EST

PDB ID	:	3CW1
Title	:	Crystal Structure of Human Spliceosomal U1 snRNP
Authors	:	Pomeranz Krummel, D.A.; Oubridge, C.; Leung, A.K.; Li, J.; Nagai, K.
Deposited on	:	2008-04-21
Resolution	:	5.49 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ 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 ⓘ) were used in the production of this report:

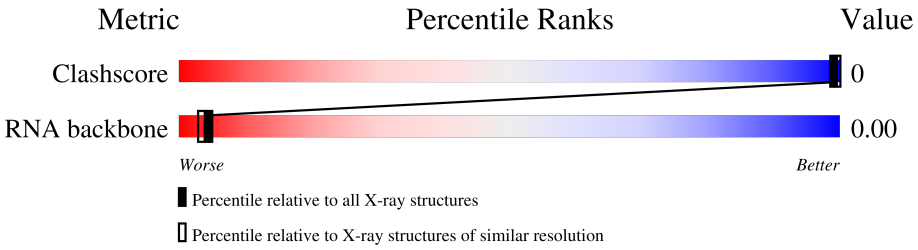
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 5.49 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))
Clashscore	141614	1010 (7.10-3.90)
RNA backbone	3102	1074 (7.80-3.00)

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%

Mol	Chain	Length	Quality of chain
1	V	138	100%
1	v	138	99% .
1	w	138	100%
1	x	138	100%
2	D	126	60% 40%
2	S	126	60% 40%
2	T	126	60% 40%
2	U	126	60% 40%
3	A	174	37% 63%
3	H	174	37% 63%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	I	174	36% 64%
3	J	174	36% 64%
4	B	119	65% 35%
4	M	119	65% 35%
4	N	119	64% 36%
4	O	119	65% 35%
5	C	118	75% 25%
5	P	118	73% 27%
5	Q	118	75% 25%
5	R	118	72% 28%
6	1	86	87% 13%
6	2	86	88% 12%
6	F	86	88% 12%
6	Z	86	81% 19%
7	E	92	82% 18%
7	W	92	82% 18%
7	X	92	82% 18%
7	Y	92	82% 18%
8	3	76	96% .
8	4	76	96% .
8	5	76	96% .
8	G	76	96% .
9	6	216	56% 44%
9	7	216	56% 44%
9	8	216	56% 44%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	K	216	55% 44%
10	0	77	70% 30%
10	9	77	75% 25%
10	L	77	74% 26%
10	l	77	75% 25%

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 3365 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 RNA chain called U1 snRNA.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
1	V	138	Total P 138 138	0	0	138
1	v	138	Total P 138 138	0	0	138
1	w	138	Total P 138 138	0	0	138
1	x	138	Total P 138 138	0	0	138

- Molecule 2 is a protein called Small nuclear ribonucleoprotein Sm D3.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	D	76	Total C 76 76	0	0	76
2	S	76	Total C 76 76	0	0	76
2	T	76	Total C 76 76	0	0	76
2	U	76	Total C 76 76	0	0	76

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	66	CYS	SER	conflict	UNP P62318
S	266	CYS	SER	conflict	UNP P62318
T	466	CYS	SER	conflict	UNP P62318
U	666	CYS	SER	conflict	UNP P62318

- Molecule 3 is a protein called Small nuclear ribonucleoprotein-associated proteins B and B'.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
3	A	64	Total C 64 64	0	0	64
3	H	64	Total C 64 64	0	0	64
3	I	63	Total C 63 63	0	0	63
3	J	63	Total C 63 63	0	0	63

- Molecule 4 is a protein called Small nuclear ribonucleoprotein Sm D1.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
4	B	77	Total C 77 77	0	0	77
4	M	77	Total C 77 77	0	0	77
4	N	76	Total C 76 76	0	0	76
4	O	77	Total C 77 77	0	0	77

- Molecule 5 is a protein called Small nuclear ribonucleoprotein Sm D2.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
5	C	88	Total C 88 88	0	0	88
5	P	86	Total C 86 86	0	0	86
5	Q	89	Total C 89 89	0	0	89
5	R	85	Total C 85 85	0	0	85

- Molecule 6 is a protein called Small nuclear ribonucleoprotein F.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
6	F	76	Total C 76 76	0	0	76
6	Z	70	Total C 70 70	0	0	70
6	1	75	Total C 75 75	0	0	75

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
6	2	76	Total C 76 76	0	0	76

- Molecule 7 is a protein called Small nuclear ribonucleoprotein E.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
7	E	75	Total C 75 75	0	0	75
7	W	75	Total C 75 75	0	0	75
7	X	75	Total C 75 75	0	0	75
7	Y	75	Total C 75 75	0	0	75

- Molecule 8 is a protein called Small nuclear ribonucleoprotein G.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
8	G	73	Total C 73 73	0	0	73
8	3	73	Total C 73 73	0	0	73
8	4	73	Total C 73 73	0	0	73
8	5	73	Total C 73 73	0	0	73

- Molecule 9 is a protein called U1 small nuclear ribonucleoprotein 70 kDa.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
9	K	120	Total C 120 120	0	0	120
9	6	120	Total C 120 120	0	0	120
9	7	120	Total C 120 120	0	0	120
9	8	120	Total C 120 120	0	0	120

- Molecule 10 is a protein called U1 small nuclear ribonucleoprotein C.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
10	L	57	Total C 57 57	0	0	57
10	9	58	Total C 58 58	0	0	58
10	0	54	Total C 54 54	0	0	54
10	l	58	Total C 58 58	0	0	58

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	39	CYS	GLN	conflict	UNP P09234
9	239	CYS	GLN	conflict	UNP P09234
0	439	CYS	GLN	conflict	UNP P09234
l	639	CYS	GLN	conflict	UNP P09234

- Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

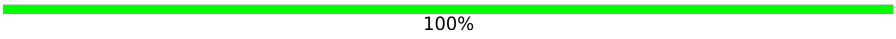
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
11	L	1	Total Zn 1 1	0	0
11	9	1	Total Zn 1 1	0	0
11	0	1	Total Zn 1 1	0	0
11	l	1	Total Zn 1 1	0	0

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. 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. 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: U1 snRNA

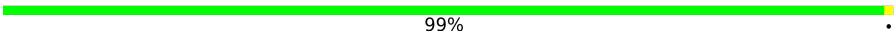
Chain V:



There are no outlier residues recorded for this chain.

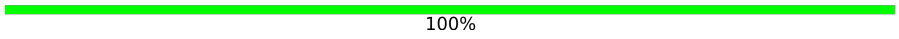
• Molecule 1: U1 snRNA

Chain v:



• Molecule 1: U1 snRNA

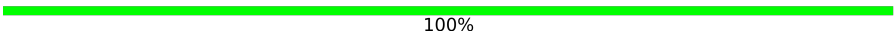
Chain w:



There are no outlier residues recorded for this chain.

• Molecule 1: U1 snRNA

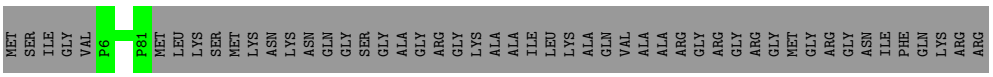
Chain x:



There are no outlier residues recorded for this chain.

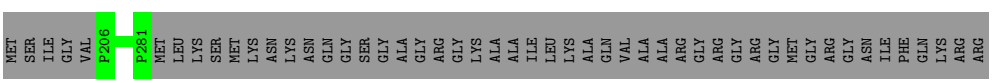
• Molecule 2: Small nuclear ribonucleoprotein Sm D3

Chain D:



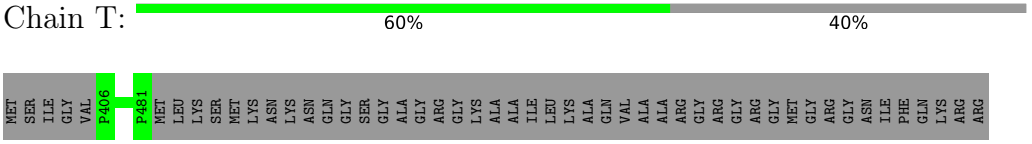
• Molecule 2: Small nuclear ribonucleoprotein Sm D3

Chain S:



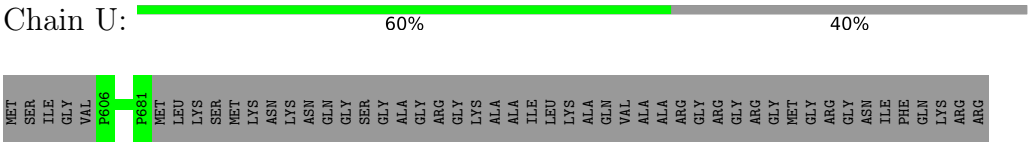
• Molecule 2: Small nuclear ribonucleoprotein Sm D3

Chain T:



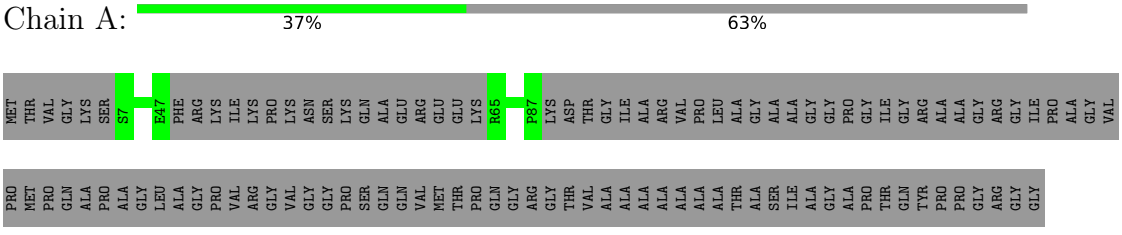
• Molecule 2: Small nuclear ribonucleoprotein Sm D3

Chain U:



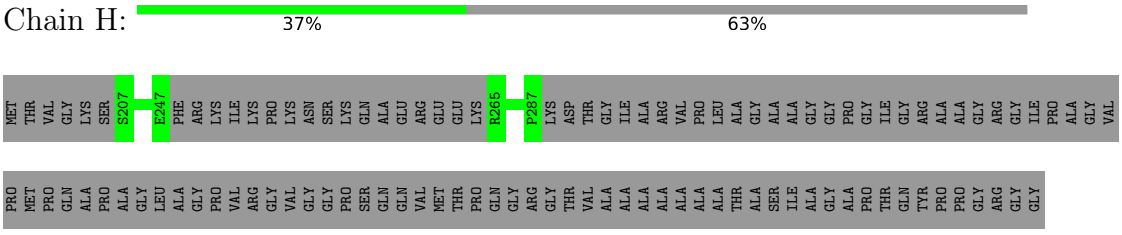
• Molecule 3: Small nuclear ribonucleoprotein-associated proteins B and B'

Chain A:



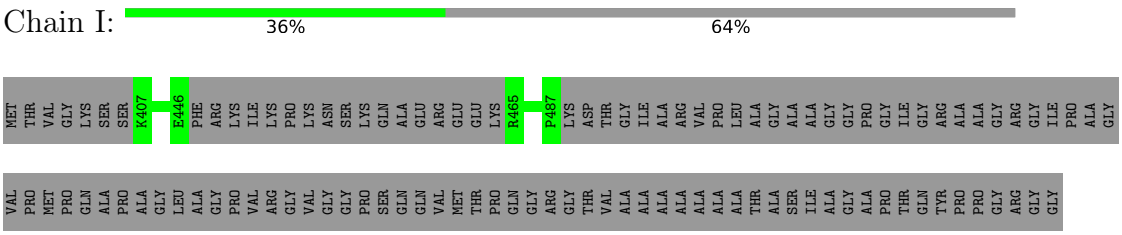
• Molecule 3: Small nuclear ribonucleoprotein-associated proteins B and B'

Chain H:



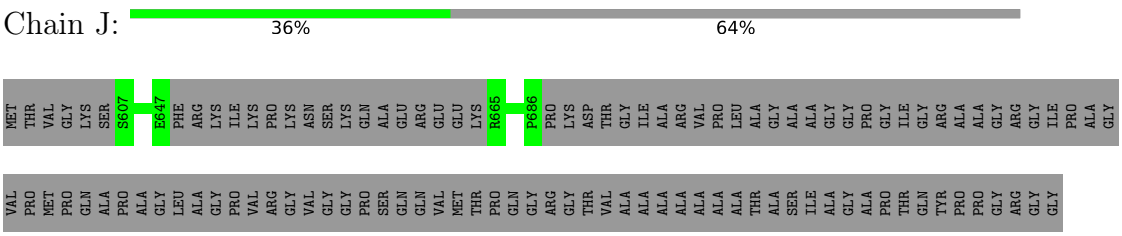
• Molecule 3: Small nuclear ribonucleoprotein-associated proteins B and B'

Chain I:



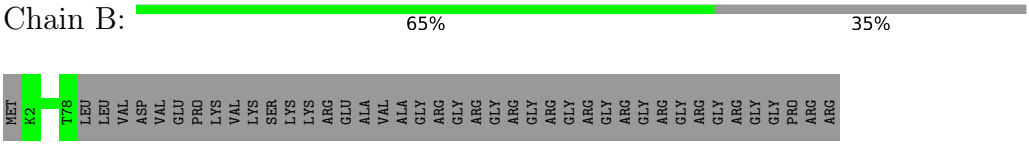
• Molecule 3: Small nuclear ribonucleoprotein-associated proteins B and B'

Chain J:



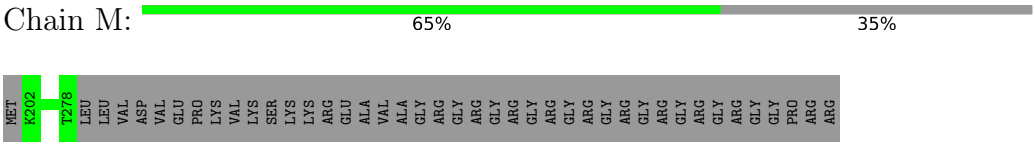
• Molecule 4: Small nuclear ribonucleoprotein Sm D1

Chain B:



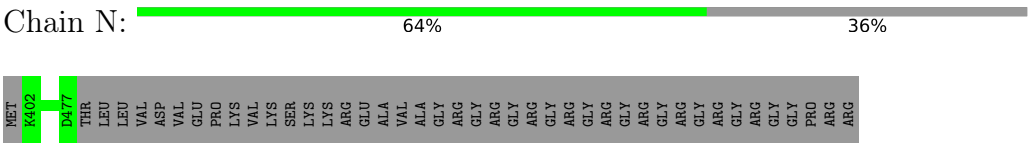
• Molecule 4: Small nuclear ribonucleoprotein Sm D1

Chain M:



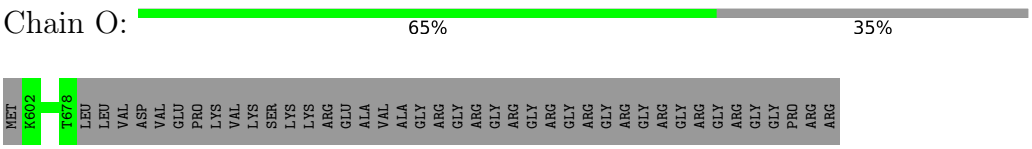
• Molecule 4: Small nuclear ribonucleoprotein Sm D1

Chain N:



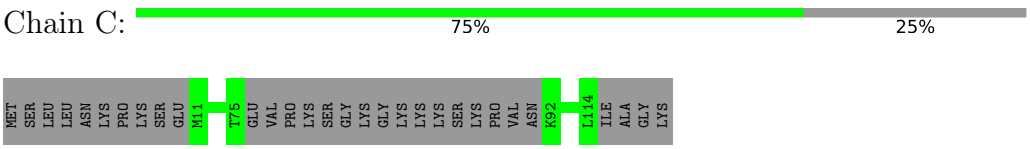
• Molecule 4: Small nuclear ribonucleoprotein Sm D1

Chain O:



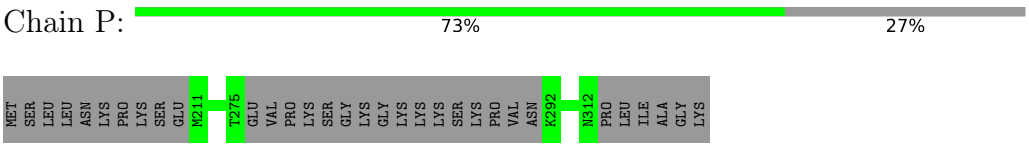
• Molecule 5: Small nuclear ribonucleoprotein Sm D2

Chain C:



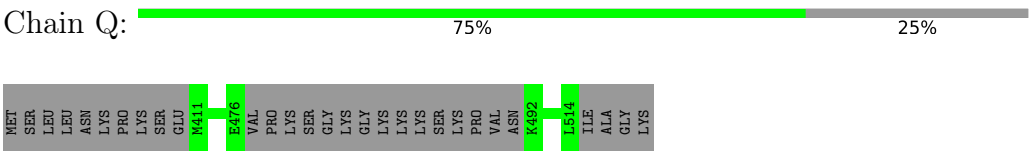
• Molecule 5: Small nuclear ribonucleoprotein Sm D2

Chain P:



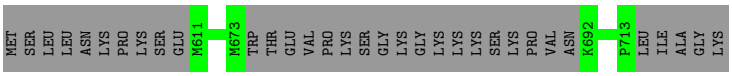
• Molecule 5: Small nuclear ribonucleoprotein Sm D2

Chain Q:



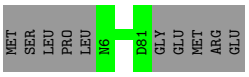
• Molecule 5: Small nuclear ribonucleoprotein Sm D2

Chain R:



• Molecule 6: Small nuclear ribonucleoprotein F

Chain F:

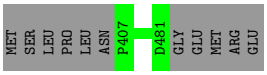


• Molecule 6: Small nuclear ribonucleoprotein F

Chain Z:

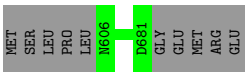
• Molecule 6: Small nuclear ribonucleoprotein F

Chain 1:



• Molecule 6: Small nuclear ribonucleoprotein F

Chain 2:



• Molecule 7: Small nuclear ribonucleoprotein E

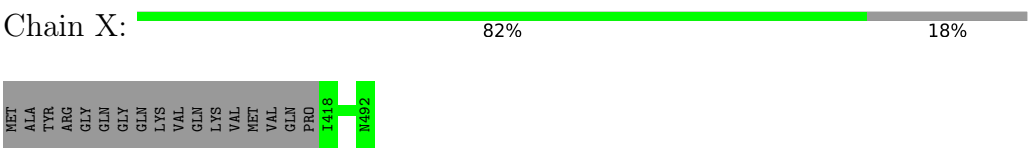
Chain E:

• Molecule 7: Small nuclear ribonucleoprotein E

Chain W:

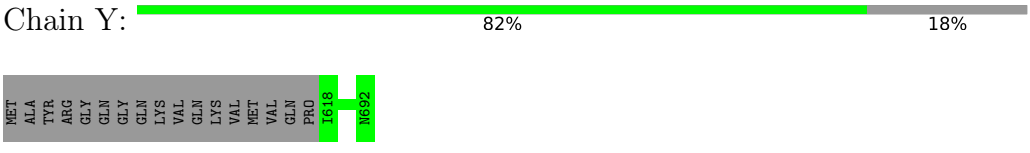
• Molecule 7: Small nuclear ribonucleoprotein E

Chain X:



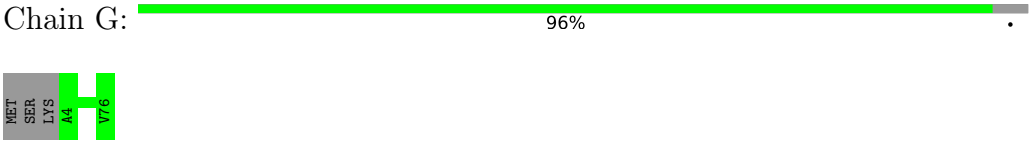
• Molecule 7: Small nuclear ribonucleoprotein E

Chain Y:



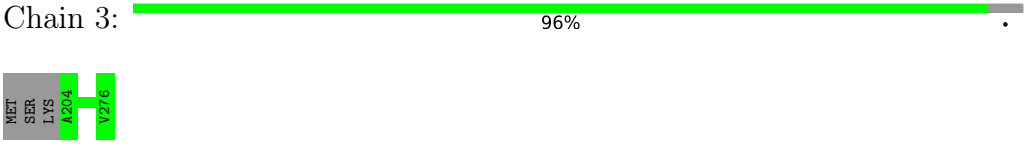
• Molecule 8: Small nuclear ribonucleoprotein G

Chain G:



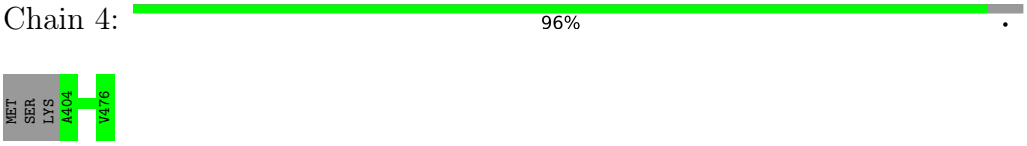
• Molecule 8: Small nuclear ribonucleoprotein G

Chain 3:



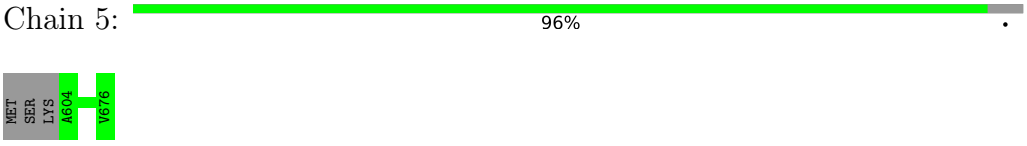
• Molecule 8: Small nuclear ribonucleoprotein G

Chain 4:



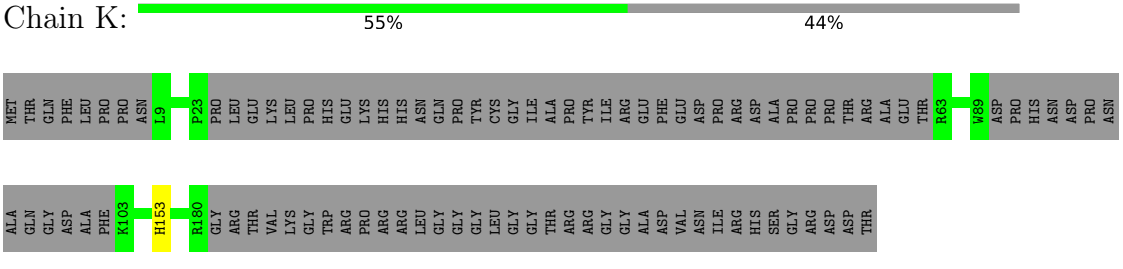
• Molecule 8: Small nuclear ribonucleoprotein G

Chain 5:



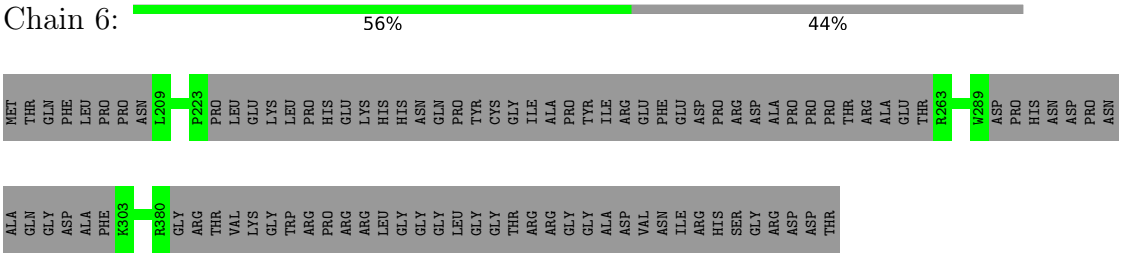
• Molecule 9: U1 small nuclear ribonucleoprotein 70 kDa

Chain K:



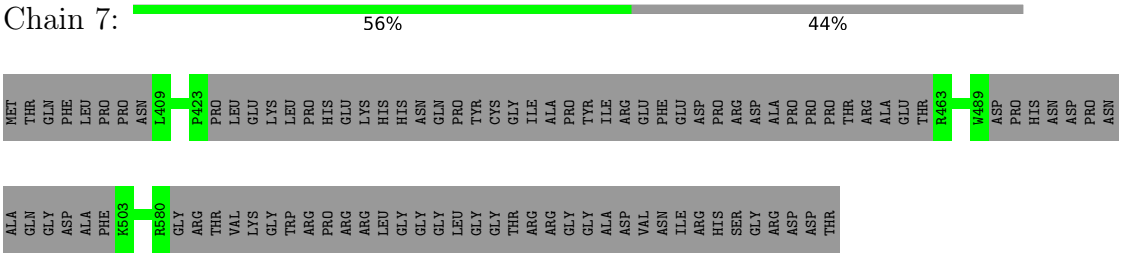
• Molecule 9: U1 small nuclear ribonucleoprotein 70 kDa

Chain 6:



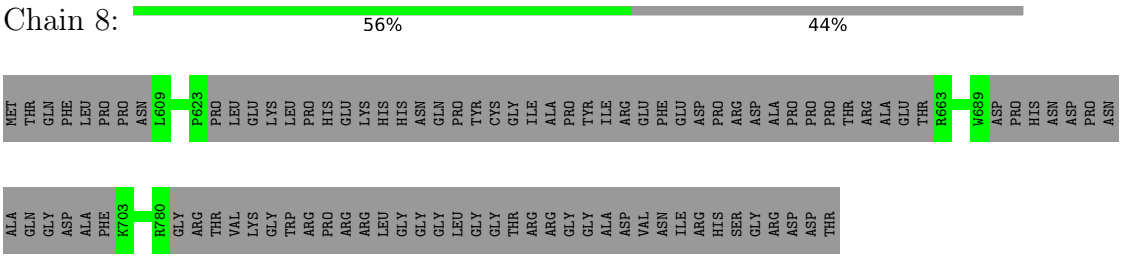
• Molecule 9: U1 small nuclear ribonucleoprotein 70 kDa

Chain 7:



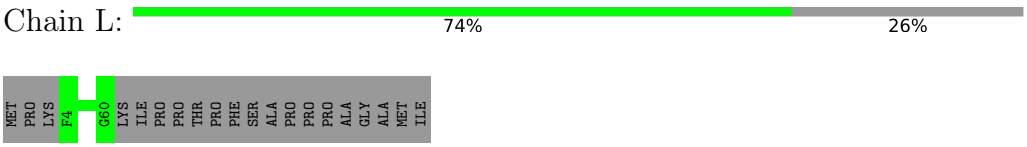
• Molecule 9: U1 small nuclear ribonucleoprotein 70 kDa

Chain 8:



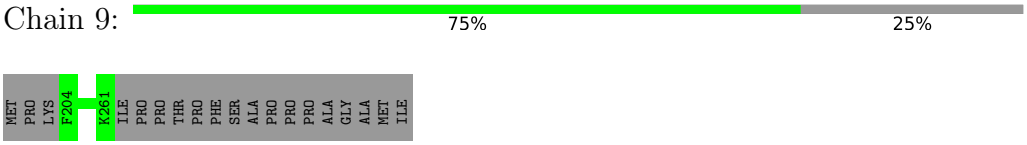
• Molecule 10: U1 small nuclear ribonucleoprotein C

Chain L:



• Molecule 10: U1 small nuclear ribonucleoprotein C

Chain 9:



• Molecule 10: U1 small nuclear ribonucleoprotein C

Chain 0:

• Molecule 10: U1 small nuclear ribonucleoprotein C

Chain l:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	126.47Å 127.08Å 152.02Å 95.42° 105.92° 101.80°	Depositor
Resolution (Å)	123.09 – 5.49 122.79 – 5.49	Depositor EDS
% Data completeness (in resolution range)	(Not available) (123.09-5.49) 99.0 (122.79-5.49)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.99 (at 5.42Å)	Xtriage
Refinement program		Depositor
R, R_{free}	(Not available) , (Not available) 0.469 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	251.7	Xtriage
Anisotropy	0.213	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.59 , -6.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.72	EDS
Total number of atoms	3365	wwPDB-VP
Average B, all atoms (Å^2)	231.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.10% 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

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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).

There are no protein, RNA or DNA chains available to summarize Z scores of covalent bonds and angles.

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	V	138	0	0	0	0
1	v	138	0	0	0	1
1	w	138	0	0	0	0
1	x	138	0	0	0	0
2	D	76	0	0	0	0
2	S	76	0	0	0	0
2	T	76	0	0	0	0
2	U	76	0	0	0	0
3	A	64	0	0	0	0
3	H	64	0	0	0	0
3	I	63	0	0	0	0
3	J	63	0	0	0	0
4	B	77	0	0	0	0
4	M	77	0	0	0	0
4	N	76	0	0	0	0
4	O	77	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	88	0	0	0	0
5	P	86	0	0	0	0
5	Q	89	0	0	0	0
5	R	85	0	0	0	0
6	1	75	0	0	0	0
6	2	76	0	0	0	0
6	F	76	0	0	0	0
6	Z	70	0	0	0	0
7	E	75	0	0	0	0
7	W	75	0	0	0	0
7	X	75	0	0	0	0
7	Y	75	0	0	0	0
8	3	73	0	0	0	0
8	4	73	0	0	0	0
8	5	73	0	0	0	0
8	G	73	0	0	0	0
9	6	120	0	0	0	0
9	7	120	0	0	0	0
9	8	120	0	0	0	0
9	K	120	0	0	0	1
10	0	54	0	0	0	0
10	9	58	0	0	0	0
10	L	57	0	0	0	0
10	l	58	0	0	0	0
11	0	1	0	0	0	0
11	9	1	0	0	0	0
11	L	1	0	0	0	0
11	l	1	0	0	0	0
All	All	3365	0	0	0	1

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 0.

There are no clashes within the asymmetric unit.

All (1) 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)
1:v:214:A:P	9:K:153:HIS:CA[1_545]	2.05	0.15

5.3 Torsion angles

5.3.1 Protein backbone

There are no protein backbone outliers to report in this entry.

5.3.2 Protein sidechains

There are no protein residues with a non-rotameric sidechain to report in this entry.

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	V	0/138	-	-
1	v	0/138	-	-
1	w	0/138	-	-
1	x	0/138	-	-
All	All	0/552	-	-

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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

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

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.