wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Feb 19, 2018 – 06:18 pm GMT

PDB ID	:	3J96
EMDB ID:	:	EMD-6206
Title	:	Structure of 20S supercomplex determined by single particle cryoelectron microscopy (State I)
Authors	:	Zhao, M.; Wu, S.; Cheng, Y.; Brunger, A.T.
Deposited on	:	2014-12-05
Resolution	:	7.60 Å(reported)
Based on PDB ID	:	1QCS, 1NSF, 1N7S

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

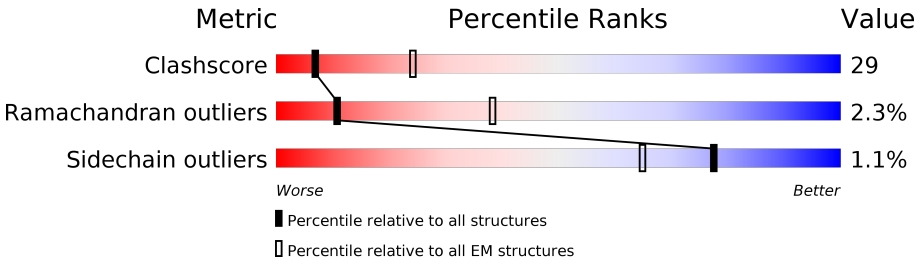
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30686

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 7.60 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.



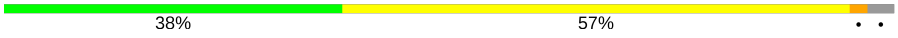
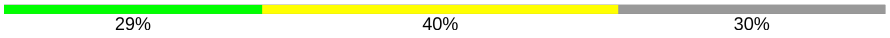
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	136279	1886
Ramachandran outliers	132675	1663
Sidechain outliers	132484	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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%

Mol	Chain	Length	Quality of chain
1	A	747	45% 42% • • 9%
1	B	747	46% 41% • 10%
1	C	747	47% 41% • 10%
1	D	747	45% 43% • 10%
1	E	747	45% 42% • 10%
1	F	747	44% 41% • 12%
2	G	297	49% 44% • •
2	H	297	54% 41% • •
2	I	297	57% 38% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	J	297	
3	K	63	
4	L	67	
5	M	188	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 41139 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Vesicle-fusing ATPase.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	678	Total	C	N	O	S	0	0
			5048	3203	876	946	23		
1	B	672	Total	C	N	O	S	0	0
			5037	3197	872	944	24		
1	C	676	Total	C	N	O	S	0	0
			5039	3196	872	948	23		
1	D	673	Total	C	N	O	S	0	0
			4994	3174	857	939	24		
1	E	670	Total	C	N	O	S	0	0
			5012	3183	866	939	24		
1	F	654	Total	C	N	O	S	0	0
			4926	3130	849	923	24		

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP P18708
A	-1	ALA	-	EXPRESSION TAG	UNP P18708
A	0	HIS	-	EXPRESSION TAG	UNP P18708
B	-2	GLY	-	EXPRESSION TAG	UNP P18708
B	-1	ALA	-	EXPRESSION TAG	UNP P18708
B	0	HIS	-	EXPRESSION TAG	UNP P18708
C	-2	GLY	-	EXPRESSION TAG	UNP P18708
C	-1	ALA	-	EXPRESSION TAG	UNP P18708
C	0	HIS	-	EXPRESSION TAG	UNP P18708
D	-2	GLY	-	EXPRESSION TAG	UNP P18708
D	-1	ALA	-	EXPRESSION TAG	UNP P18708
D	0	HIS	-	EXPRESSION TAG	UNP P18708
E	-2	GLY	-	EXPRESSION TAG	UNP P18708
E	-1	ALA	-	EXPRESSION TAG	UNP P18708
E	0	HIS	-	EXPRESSION TAG	UNP P18708
F	-2	GLY	-	EXPRESSION TAG	UNP P18708
F	-1	ALA	-	EXPRESSION TAG	UNP P18708
F	0	HIS	-	EXPRESSION TAG	UNP P18708

- Molecule 2 is a protein called Alpha-soluble NSF attachment protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	G	286	Total	C	N	O	S	0	0
			2255	1424	373	441	17		
2	H	286	Total	C	N	O	S	0	0
			2255	1424	373	441	17		
2	I	286	Total	C	N	O	S	0	0
			2251	1421	372	441	17		
2	J	286	Total	C	N	O	S	0	0
			2255	1424	373	441	17		

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-1	GLY	-	EXPRESSION TAG	UNP P54921
G	0	SER	-	EXPRESSION TAG	UNP P54921
H	-1	GLY	-	EXPRESSION TAG	UNP P54921
H	0	SER	-	EXPRESSION TAG	UNP P54921
I	-1	GLY	-	EXPRESSION TAG	UNP P54921
I	0	SER	-	EXPRESSION TAG	UNP P54921
J	-1	GLY	-	EXPRESSION TAG	UNP P54921
J	0	SER	-	EXPRESSION TAG	UNP P54921

- Molecule 3 is a protein called Vesicle-associated membrane protein 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	K	61	Total	C	N	O	S	0	0
			493	301	93	98	1		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	27	GLY	-	EXPRESSION TAG	UNP P63045

- Molecule 4 is a protein called Syntaxin-1A.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	L	66	Total	C	N	O	S	0	0
			536	331	91	109	5		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	190	MET	-	EXPRESSION TAG	UNP P32851

- Molecule 5 is a protein called Synaptosomal-associated protein 25.

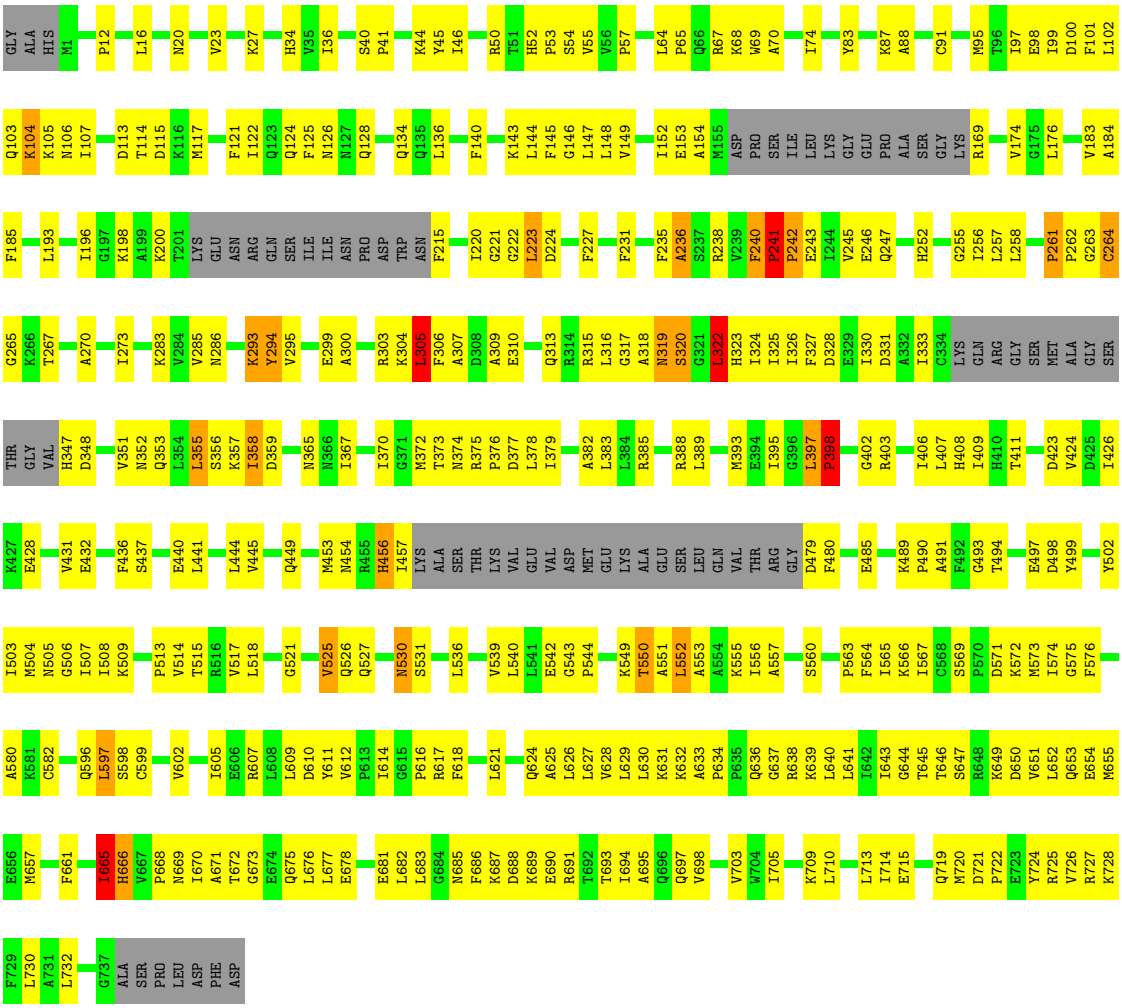
Mol	Chain	Residues	Atoms					AltConf	Trace
5	M	131	Total	C	N	O	S	0	0
			1038	614	194	221	9		

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. 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.

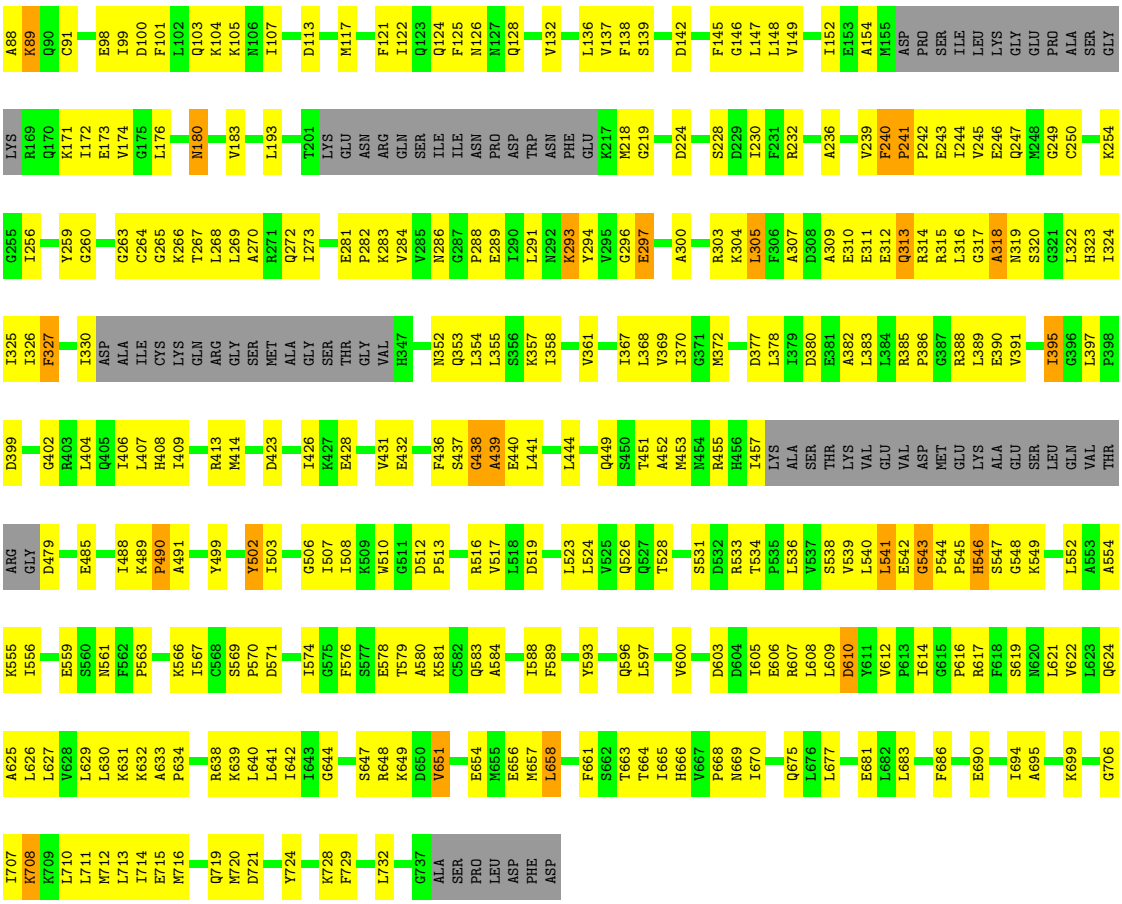
- Molecule 1: Vesicle-fusing ATPase

Chain A:



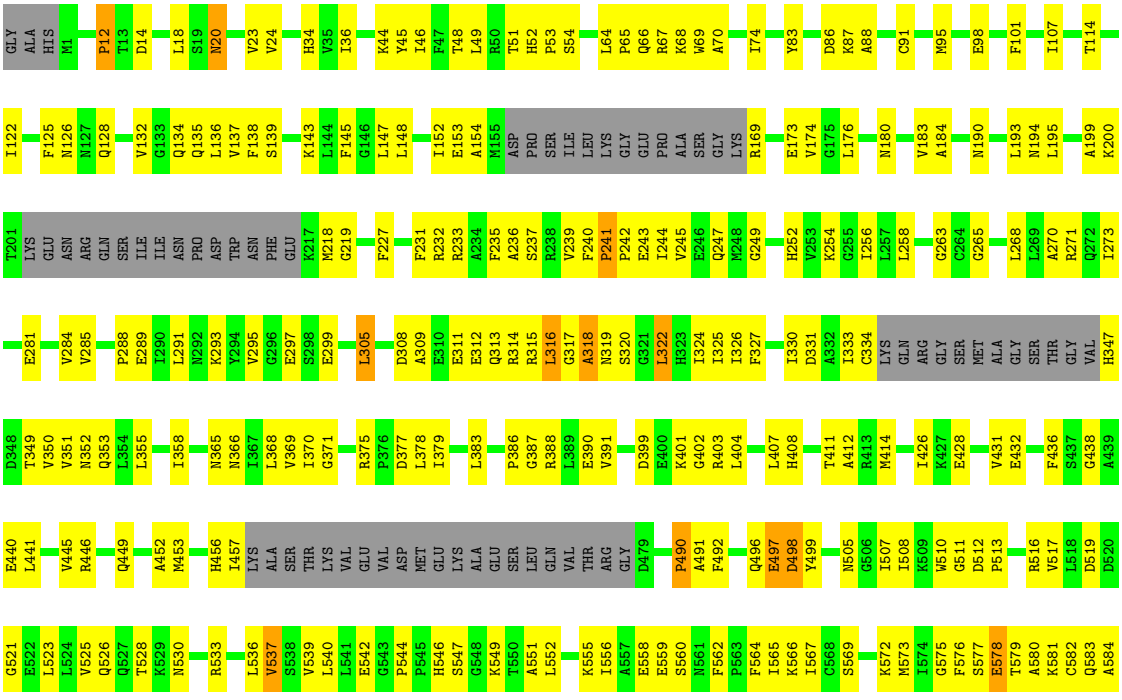
- Molecule 1: Vesicle-fusing ATPase

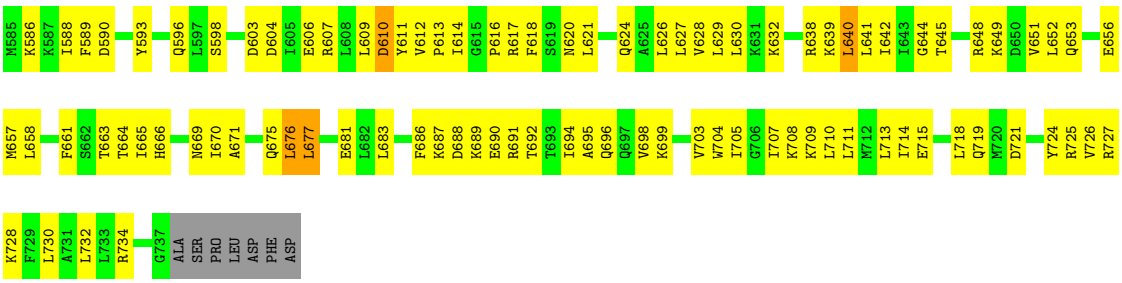
Chain B:



• Molecule 1: Vesicle-fusing ATPase

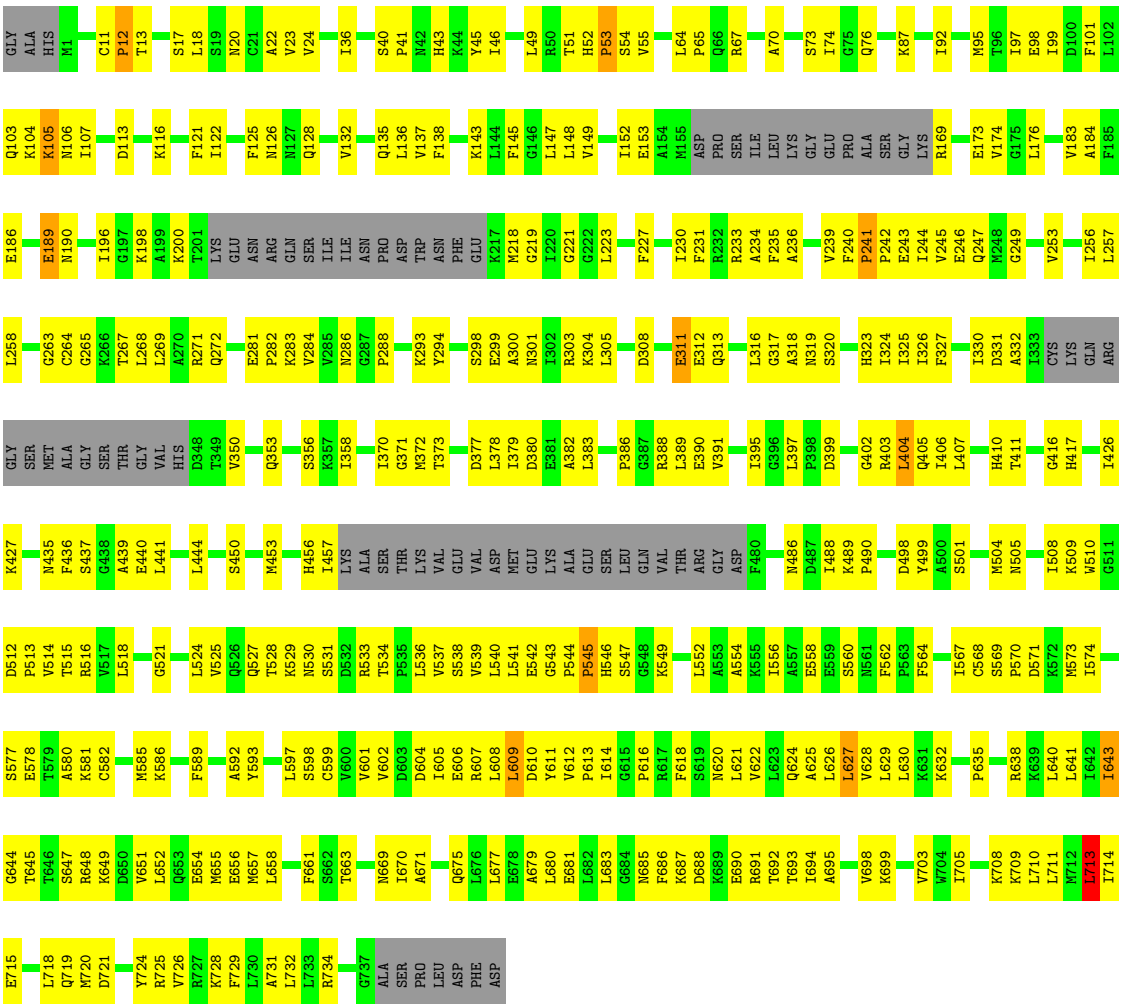
Chain C:





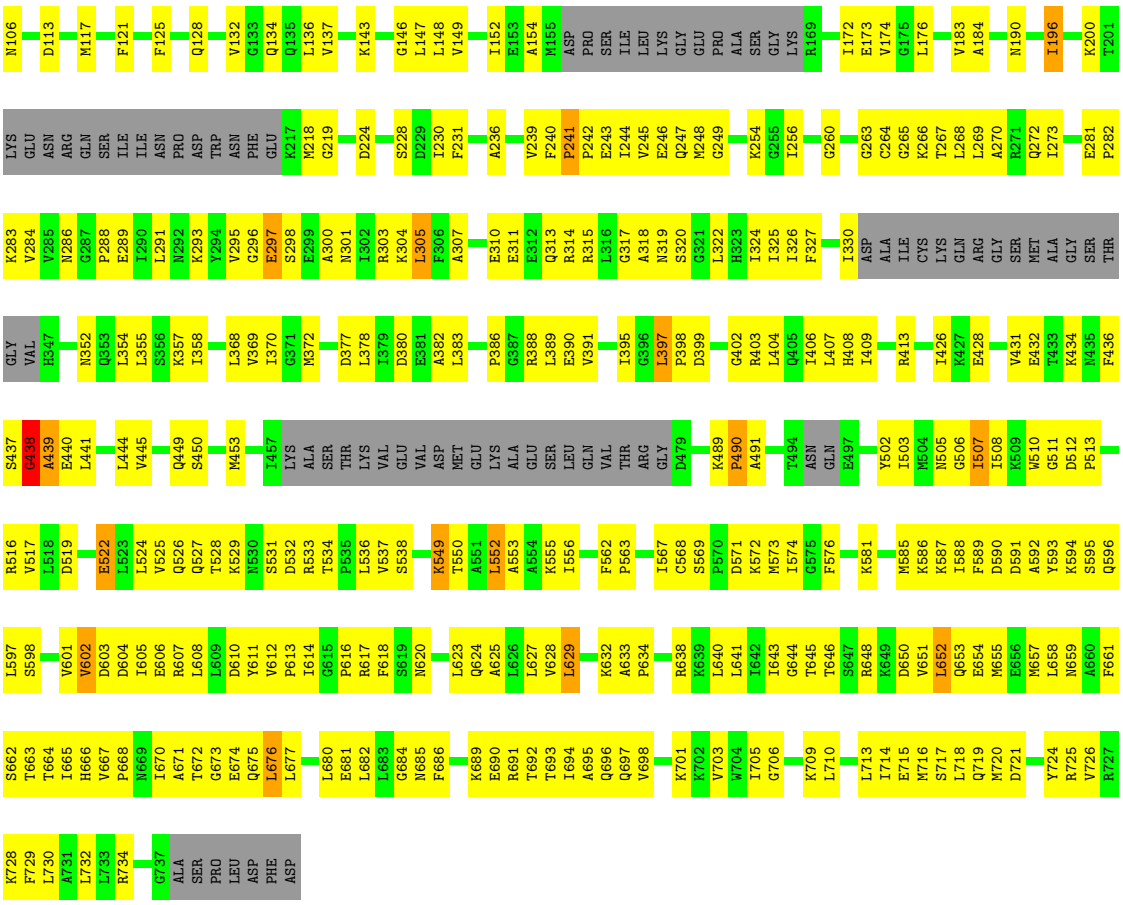
• Molecule 1: Vesicle-fusing ATPase

Chain D:



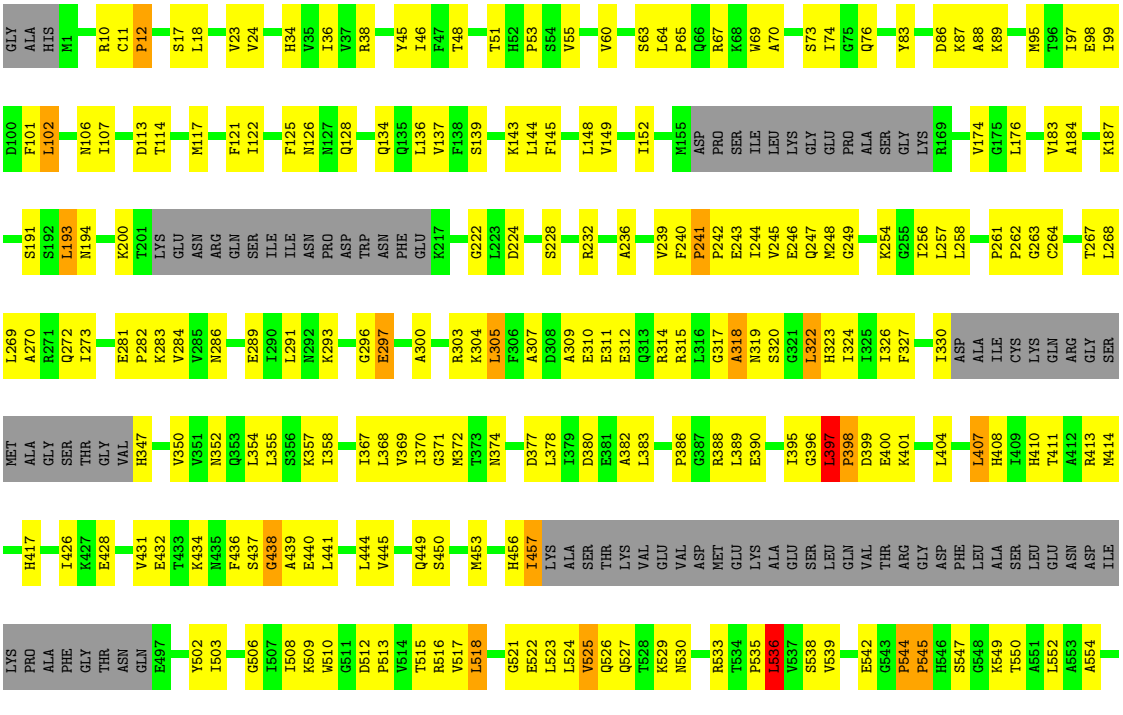
• Molecule 1: Vesicle-fusing ATPase

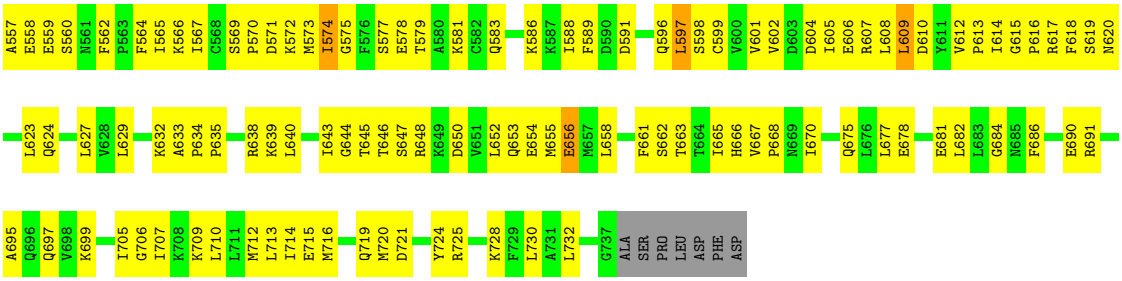
Chain E:



• Molecule 1: Vesicle-fusing ATPase

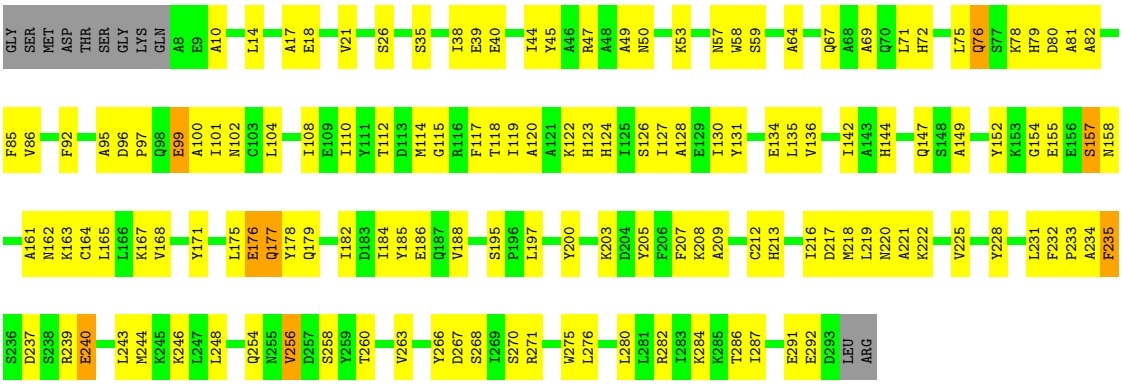
Chain F:





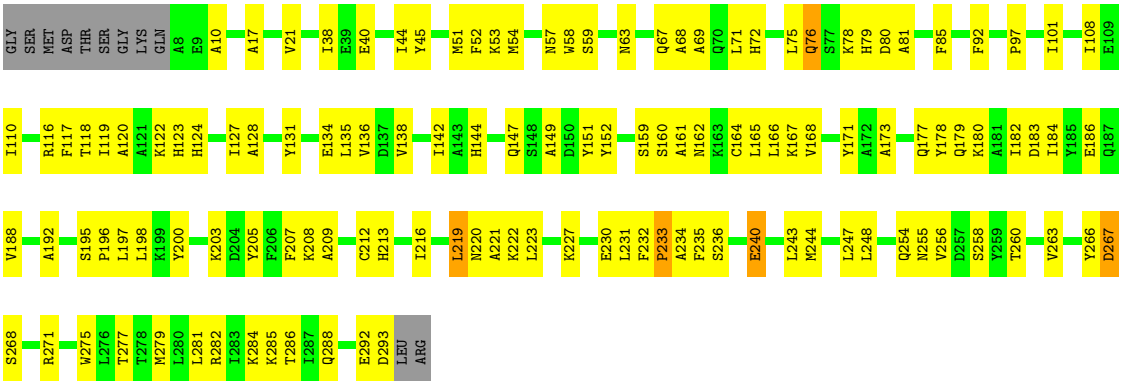
• Molecule 2: Alpha-soluble NSF attachment protein

Chain G:



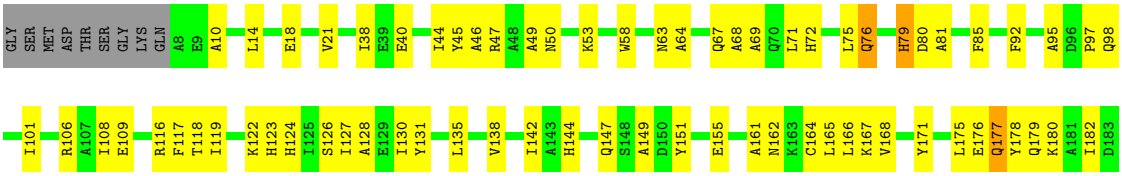
• Molecule 2: Alpha-soluble NSF attachment protein

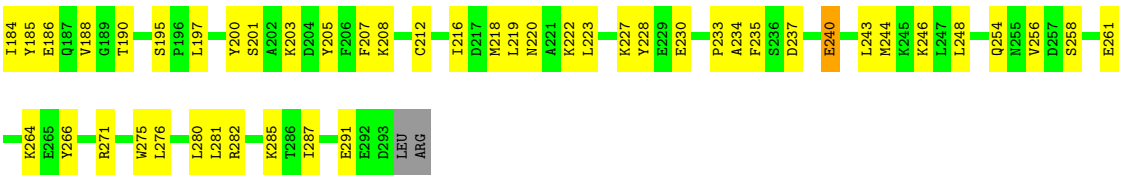
Chain H:



• Molecule 2: Alpha-soluble NSF attachment protein

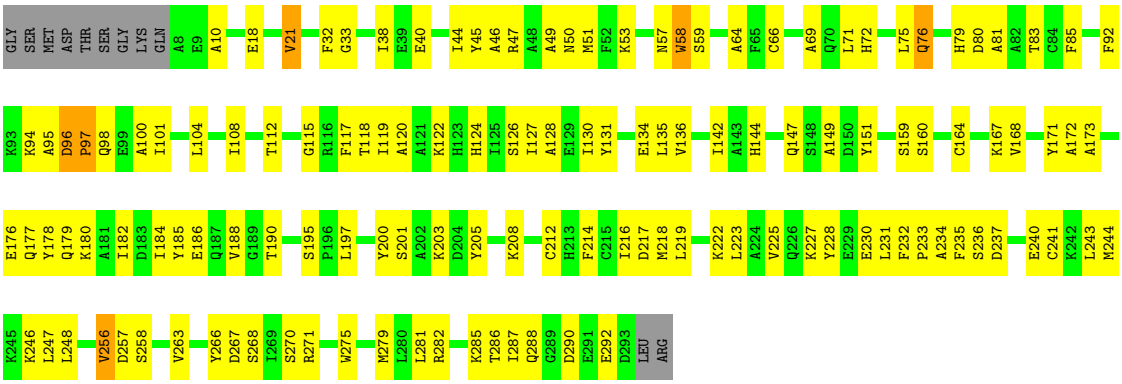
Chain I:





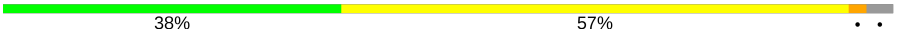
• Molecule 2: Alpha-soluble NSF attachment protein

Chain J:



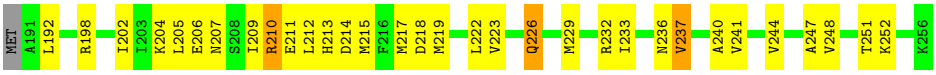
• Molecule 3: Vesicle-associated membrane protein 2

Chain K:



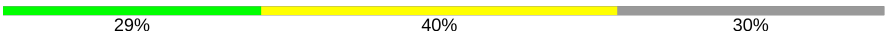
• Molecule 4: Syntaxin-1A

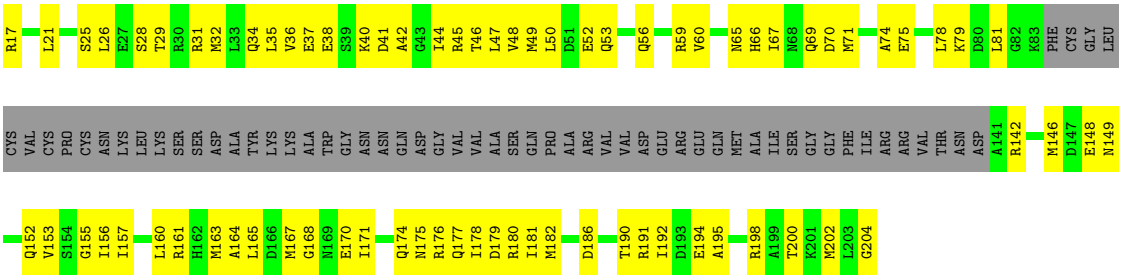
Chain L:



• Molecule 5: Synaptosomal-associated protein 25

Chain M:





4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	29717	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Each particle	Depositor
Microscope	FEI POLARA 300	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	44	Depositor
Minimum defocus (nm)	-1800	Depositor
Maximum defocus (nm)	-2800	Depositor
Magnification	31000	Depositor
Image detector	GATAN K2 (4k x 4k)	Depositor

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	#$\|Z\|$ >2	RMSZ	#$\|Z\|$ >2
1	A	0.46	1/5124 (0.0%)	0.85	13/6935 (0.2%)
1	B	0.41	1/5113 (0.0%)	0.77	7/6915 (0.1%)
1	C	0.39	0/5115	0.75	6/6922 (0.1%)
1	D	0.43	0/5069	0.77	8/6864 (0.1%)
1	E	0.44	0/5088	0.83	14/6881 (0.2%)
1	F	0.44	1/5001 (0.0%)	0.81	14/6760 (0.2%)
2	G	0.36	0/2295	0.65	0/3086
2	H	0.36	0/2295	0.62	2/3086 (0.1%)
2	I	0.35	0/2291	0.61	0/3082
2	J	0.35	0/2295	0.62	1/3086 (0.0%)
3	K	0.24	0/496	0.41	0/664
4	L	0.24	0/541	0.43	0/723
5	M	0.22	0/1038	0.41	0/1381
All	All	0.41	3/41761 (0.0%)	0.75	65/56385 (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	1
1	B	0	1
1	E	0	1
1	F	0	1
All	All	0	4

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	320	SER	CB-OG	-6.72	1.33	1.42
1	B	708	LYS	CE-NZ	5.60	1.63	1.49
1	F	545	PRO	N-CD	5.21	1.55	1.47

The worst 5 of 65 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	708	LYS	CD-CE-NZ	14.61	145.29	111.70
1	F	518	LEU	CB-CG-CD1	-10.49	93.17	111.00
1	A	597	LEU	CB-CG-CD2	-10.05	93.91	111.00
1	F	397	LEU	CA-CB-CG	9.14	136.32	115.30
1	C	322	LEU	CA-CB-CG	8.22	134.21	115.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	241	PRO	Peptide
1	B	438	GLY	Peptide
1	E	438	GLY	Peptide
1	F	438	GLY	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	5048	0	4974	346	0
1	B	5037	0	4996	313	0
1	C	5039	0	4965	309	0
1	D	4994	0	4923	331	0
1	E	5012	0	4954	336	0
1	F	4926	0	4896	311	0
2	G	2255	0	2199	142	0
2	H	2255	0	2199	113	0
2	I	2251	0	2188	117	0
2	J	2255	0	2199	125	0
3	K	493	0	491	61	0
4	L	536	0	527	58	0
5	M	1038	0	1011	117	0
All	All	41139	0	40522	2385	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 29.

The worst 5 of 2385 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:F:263:GLY:HA3	1:F:437:SER:HB2	1.29	1.14
1:A:549:LYS:NZ	1:A:647:SER:OG	1.93	1.01
2:H:271:ARG:HH11	2:I:234:ALA:HB2	1.28	0.98
2:I:200:TYR:HB3	5:M:161:ARG:HD2	1.43	0.97
2:H:219:LEU:HB2	2:H:222:LYS:HB3	1.46	0.97

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 PDB entries followed by that with respect to all EM entries.

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	668/747 (89%)	610 (91%)	43 (6%)	15 (2%)	7 42
1	B	662/747 (89%)	592 (89%)	55 (8%)	15 (2%)	7 41
1	C	666/747 (89%)	616 (92%)	37 (6%)	13 (2%)	8 45
1	D	663/747 (89%)	607 (92%)	44 (7%)	12 (2%)	9 47
1	E	658/747 (88%)	604 (92%)	43 (6%)	11 (2%)	10 49
1	F	644/747 (86%)	583 (90%)	43 (7%)	18 (3%)	5 37
2	G	284/297 (96%)	238 (84%)	34 (12%)	12 (4%)	3 28
2	H	284/297 (96%)	232 (82%)	44 (16%)	8 (3%)	5 37
2	I	284/297 (96%)	233 (82%)	44 (16%)	7 (2%)	6 39
2	J	284/297 (96%)	234 (82%)	41 (14%)	9 (3%)	4 34
3	K	59/63 (94%)	55 (93%)	2 (3%)	2 (3%)	4 33
4	L	64/67 (96%)	60 (94%)	3 (5%)	1 (2%)	11 50
5	M	127/188 (68%)	125 (98%)	2 (2%)	0	100 100
All	All	5347/5988 (89%)	4789 (90%)	435 (8%)	123 (2%)	11 41

5 of 123 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	223	LEU
1	A	293	LYS
1	A	294	TYR
1	A	318	ALA
1	A	333	ILE

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 PDB entries followed by that with respect to all EM entries.

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	515/638 (81%)	503 (98%)	12 (2%)	53	76
1	B	521/638 (82%)	511 (98%)	10 (2%)	60	80
1	C	516/638 (81%)	511 (99%)	5 (1%)	78	89
1	D	511/638 (80%)	504 (99%)	7 (1%)	69	85
1	E	516/638 (81%)	511 (99%)	5 (1%)	78	89
1	F	512/638 (80%)	509 (99%)	3 (1%)	87	93
2	G	235/244 (96%)	235 (100%)	0	100	100
2	H	235/244 (96%)	234 (100%)	1 (0%)	92	95
2	I	234/244 (96%)	234 (100%)	0	100	100
2	J	235/244 (96%)	235 (100%)	0	100	100
3	K	52/54 (96%)	52 (100%)	0	100	100
4	L	60/61 (98%)	58 (97%)	2 (3%)	41	67
5	M	113/161 (70%)	112 (99%)	1 (1%)	81	90
All	All	4255/5080 (84%)	4209 (99%)	46 (1%)	77	88

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	610	ASP
1	C	537	VAL
2	H	183	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	651	VAL
1	C	180	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	505	ASN
1	F	194	ASN
2	J	98	GLN
1	E	527	GLN
1	E	624	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.