wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 01:55 PM BST

PDB ID	:	3J98
EMDB ID:	:	EMD-6208
Title	:	Structure of 20S supercomplex determined by single particle cryoelectron microscopy (State IIIa)
Authors	:	Zhao, M.; Wu, S.; Cheng, Y.; Brunger, A.T.
Deposited on	:	2014-12-05
Resolution	:	8.40 Å(reported)
Based on PDB ID	:	1QCS, 1NSF, 1N7S

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

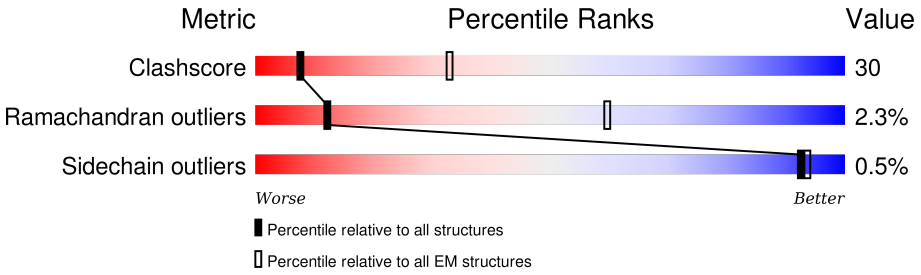
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

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 8.40 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.



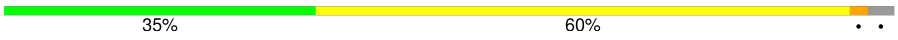
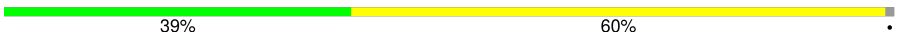
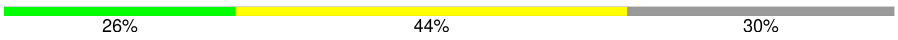
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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	46% 41% • • 10%
1	B	747	48% 40% • 10%
1	C	747	46% 43% • 10%
1	D	747	48% 41% • 10%
1	E	747	44% 44% • 10%
1	F	747	44% 41% • 12%
2	G	297	53% 41% • •
2	H	297	49% 45% • •
2	I	297	53% 41% • •

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

There are 5 unique types of molecules in this entry. The entry contains 40956 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	676	Total	C	N	O	S	0	0
			5022	3185	872	943	22		
1	B	671	Total	C	N	O	S	0	0
			4988	3163	863	938	24		
1	C	676	Total	C	N	O	S	0	0
			5019	3181	870	945	23		
1	D	673	Total	C	N	O	S	0	0
			4971	3156	854	937	24		
1	E	670	Total	C	N	O	S	0	0
			4983	3162	860	938	23		
1	F	654	Total	C	N	O	S	0	0
			4923	3120	856	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	H	286	Total	C	N	O	S	0	0
			2241	1414	372	438	17		
2	I	286	Total	C	N	O	S	0	0
			2255	1424	373	441	17		
2	J	286	Total	C	N	O	S	0	0
			2255	1424	373	441	17		
2	G	286	Total	C	N	O	S	0	0
			2249	1418	373	441	17		

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
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
G	-1	GLY	-	EXPRESSION TAG	UNP P54921
G	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
			480	293	89	97	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
			531	328	89	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.

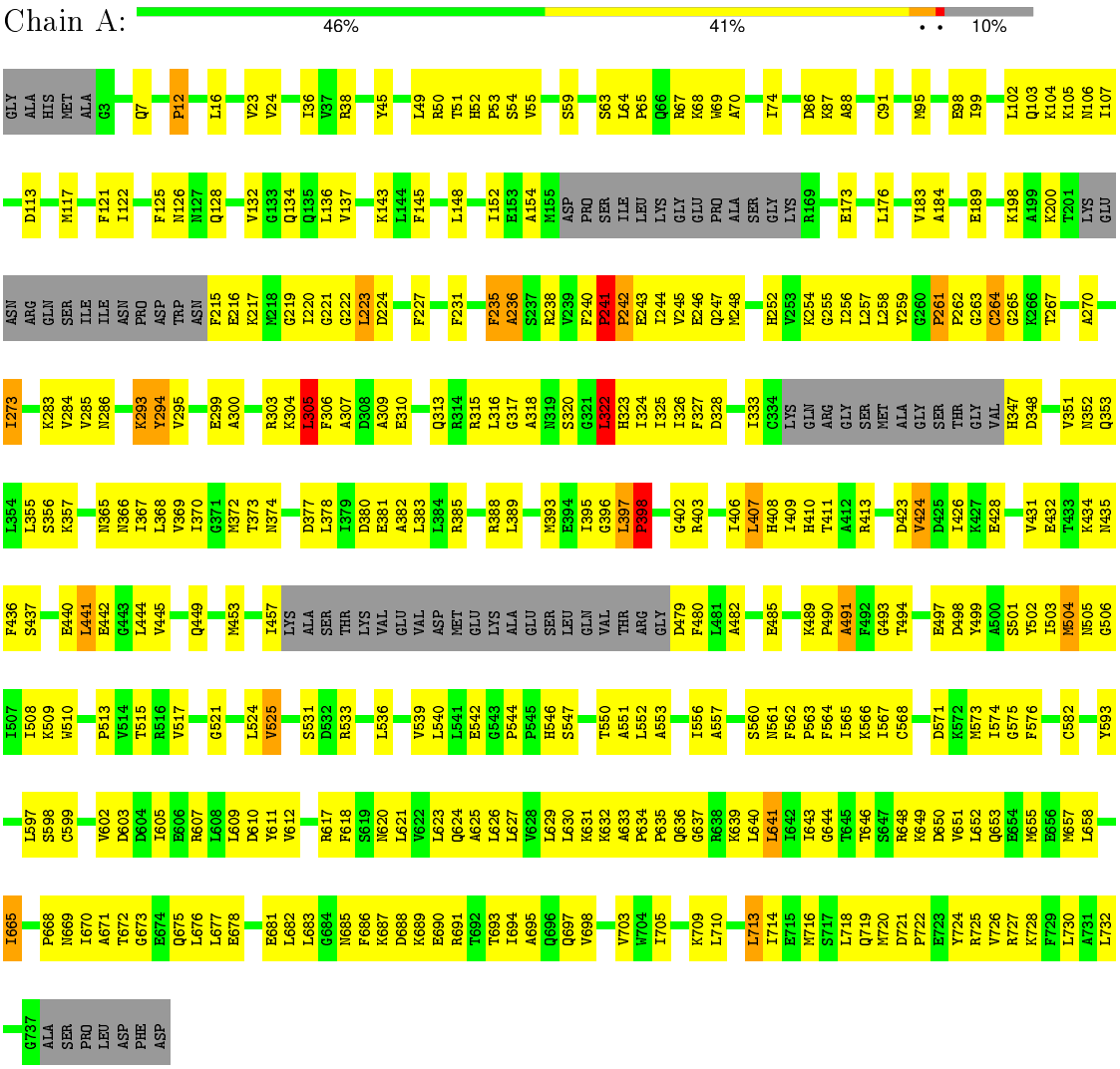
Mol	Chain	Residues	Atoms					AltConf	Trace
5	M	131	Total	C	N	O	S	0	0
			1039	616	195	219	9		

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 errors 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.

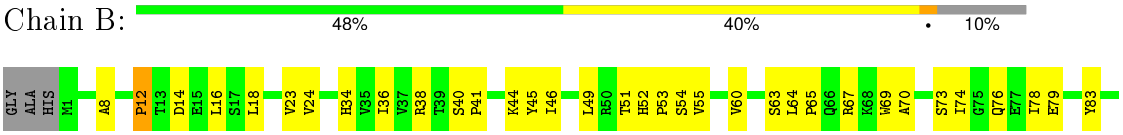
- Molecule 1: Vesicle-fusing ATPase

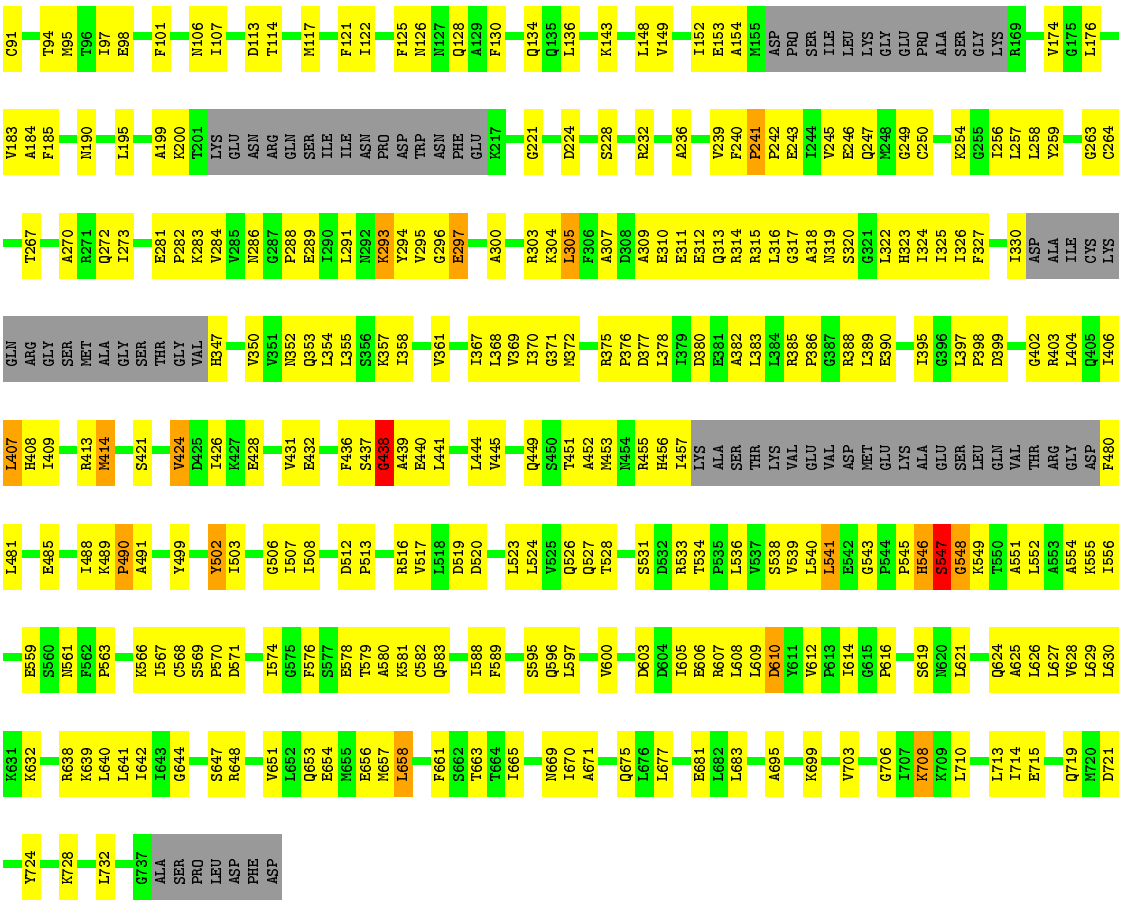
Chain A:



- Molecule 1: Vesicle-fusing ATPase

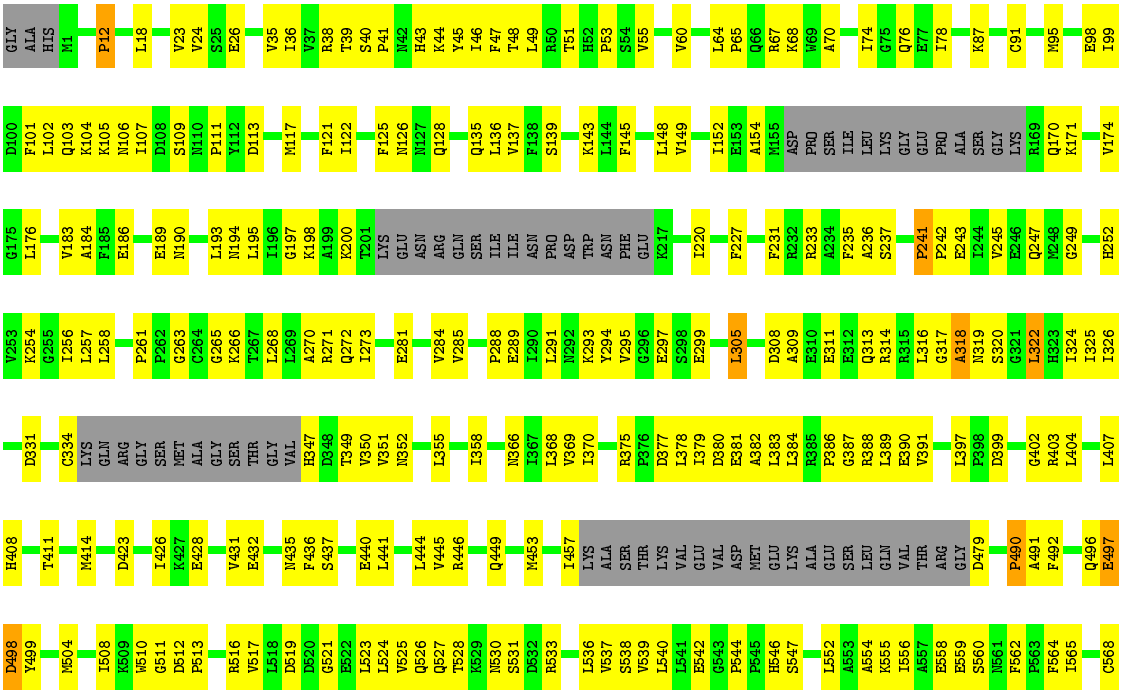
Chain B:

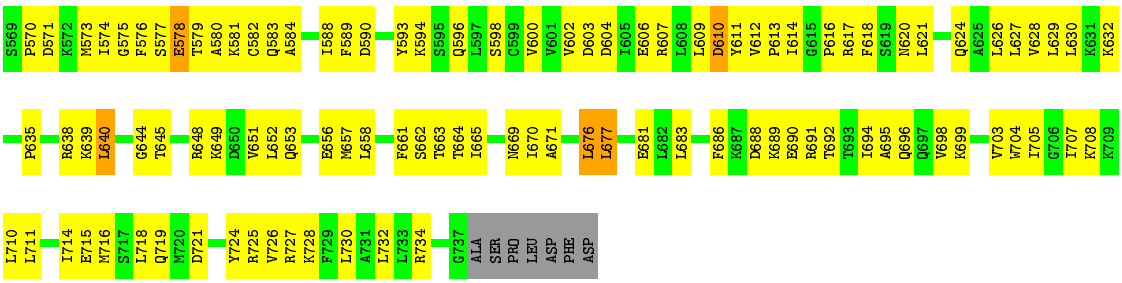




• Molecule 1: Vesicle-fusing ATPase

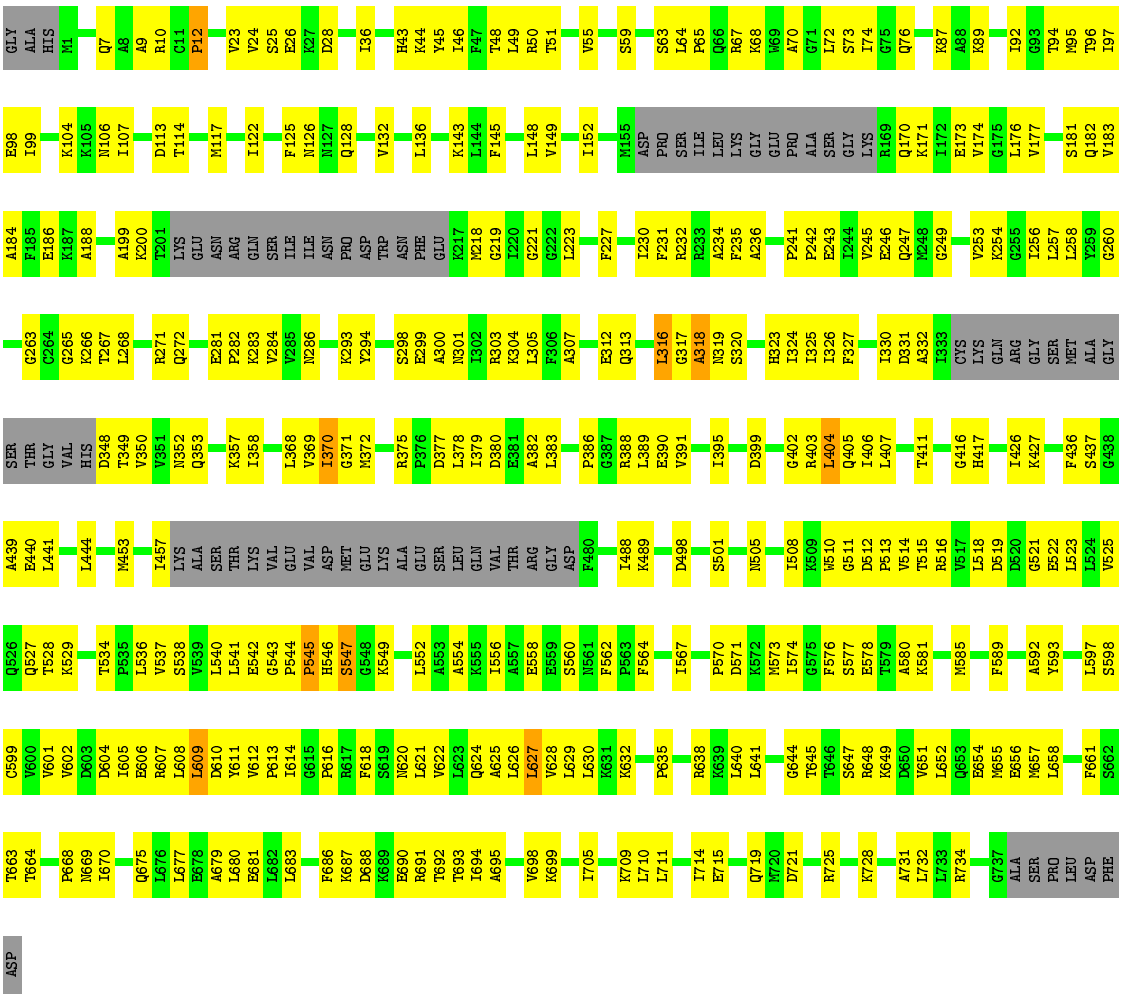
Chain C:





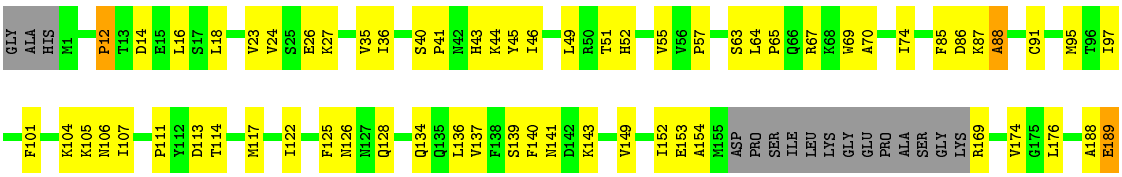
• Molecule 1: Vesicle-fusing ATPase

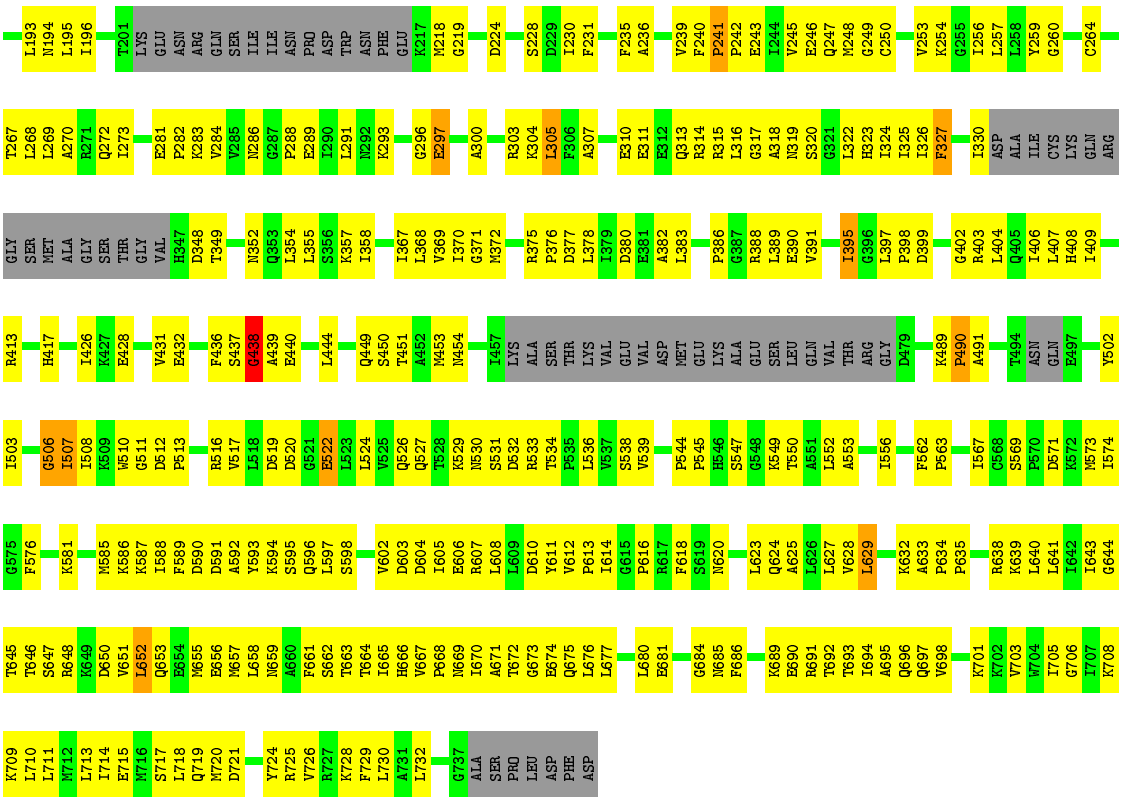
Chain D:



• Molecule 1: Vesicle-fusing ATPase

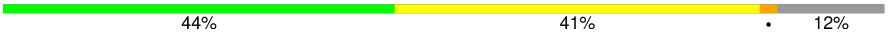
Chain E:

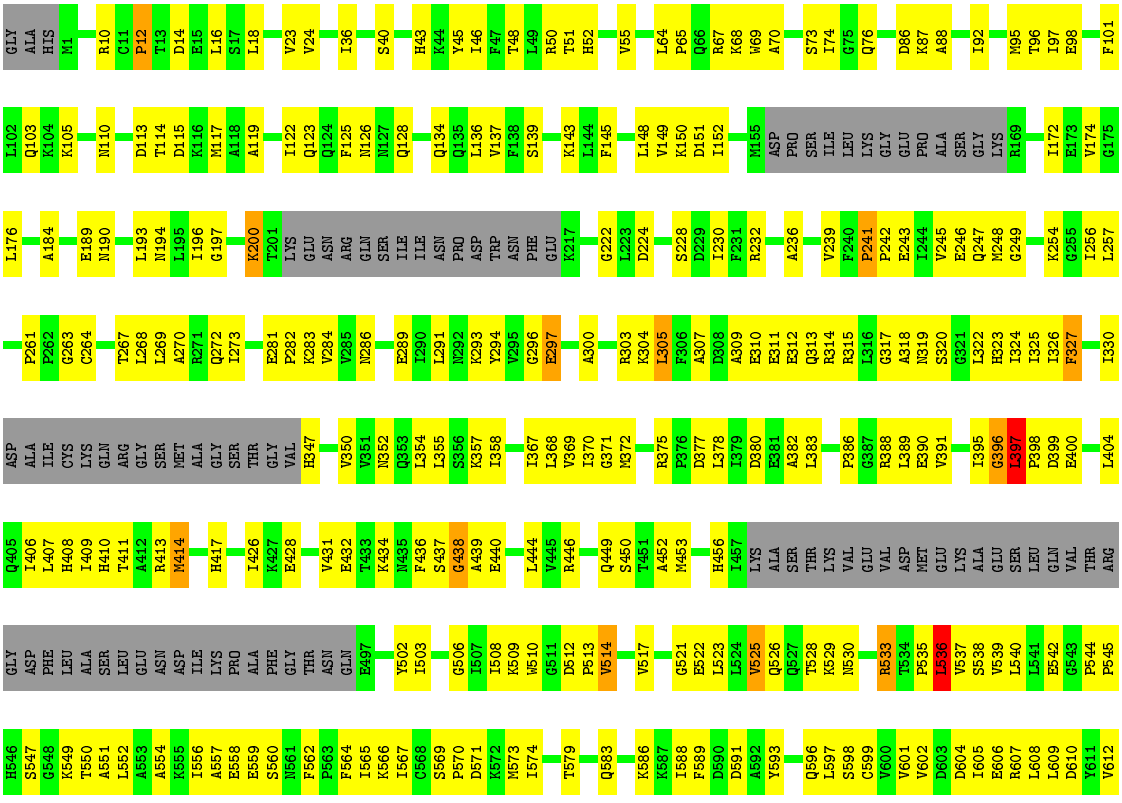


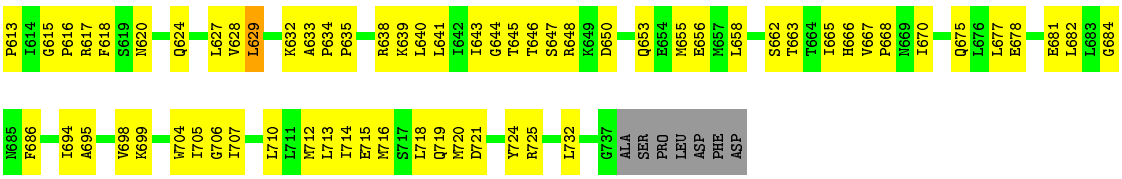


• Molecule 1: Vesicle-fusing ATPase

Chain F:

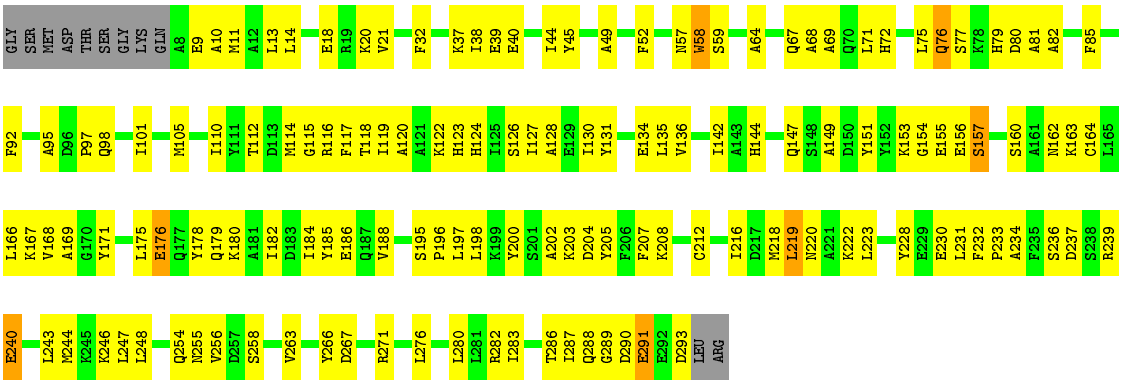






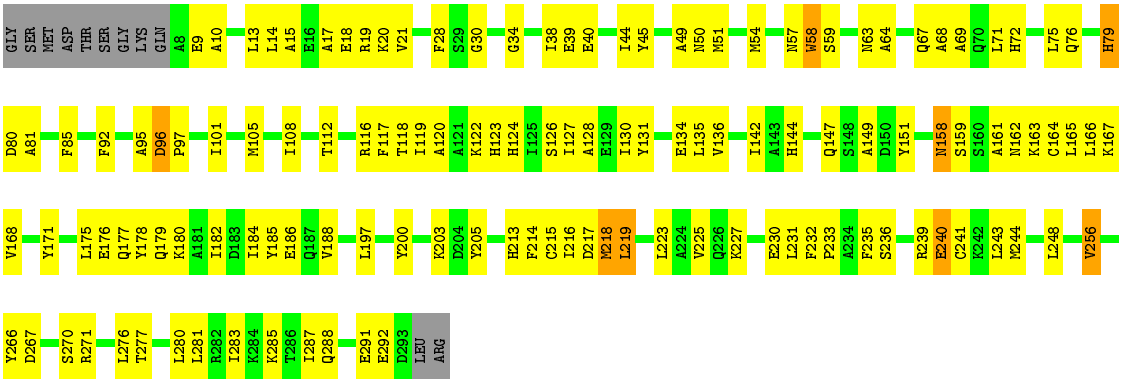
• Molecule 2: Alpha-soluble NSF attachment protein

Chain H:



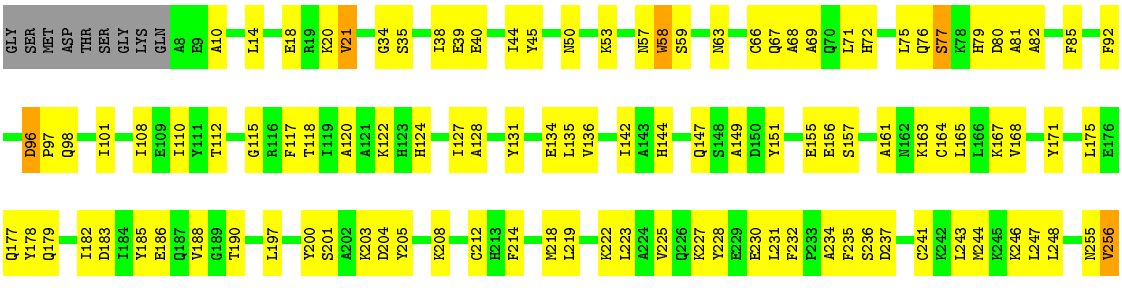
• Molecule 2: Alpha-soluble NSF attachment protein

Chain I:



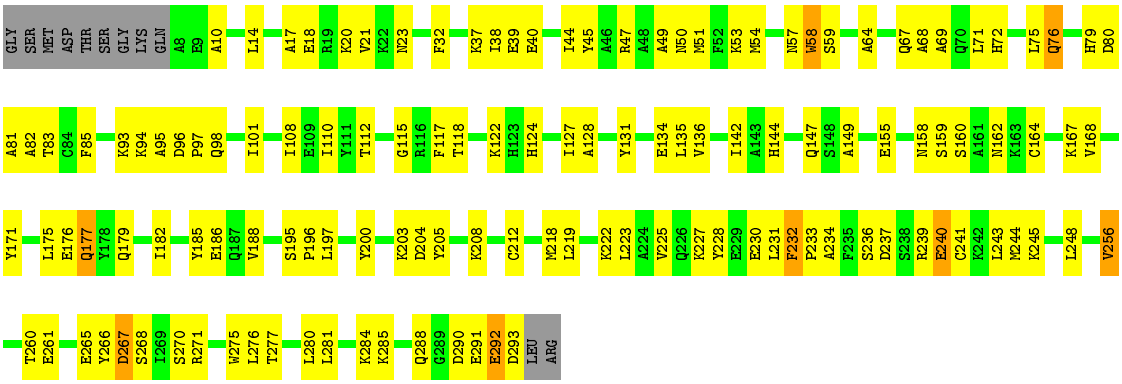
• Molecule 2: Alpha-soluble NSF attachment protein

Chain J:



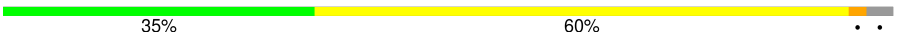
- Molecule 2: Alpha-soluble NSF attachment protein

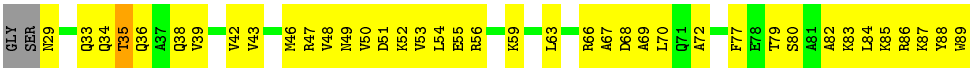
Chain G:



- Molecule 3: Vesicle-associated membrane protein 2

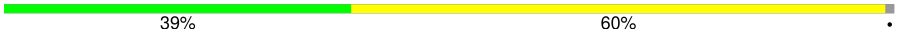
Chain K:

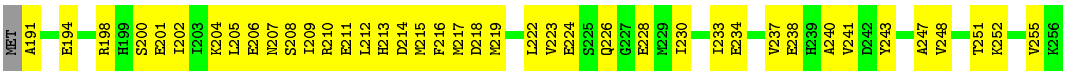




- Molecule 4: Syntaxin-1A

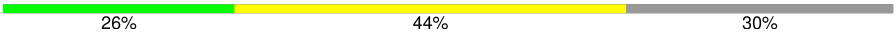
Chain L:

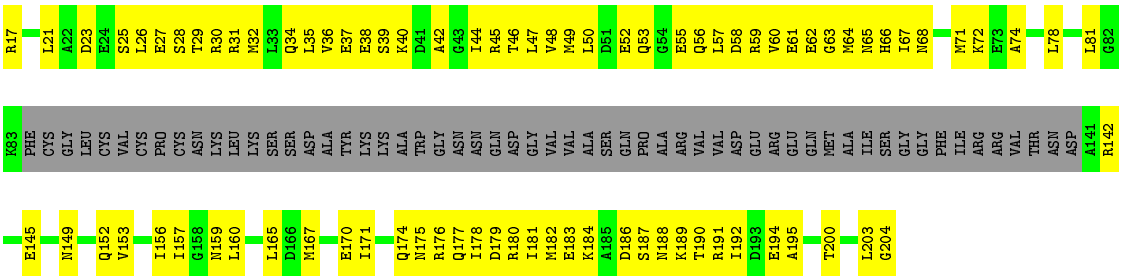




- Molecule 5: Synaptosomal-associated protein 25

Chain M:





4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	15249	Depositor
Resolution determination method	FSC 0.143	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	0.48	2/5099 (0.0%)	0.84	16/6906 (0.2%)
1	B	0.39	1/5062 (0.0%)	0.78	12/6854 (0.2%)
1	C	0.38	0/5094	0.73	5/6896 (0.1%)
1	D	0.39	0/5045	0.72	5/6834 (0.1%)
1	E	0.42	0/5057	0.79	8/6845 (0.1%)
1	F	0.43	1/4996 (0.0%)	0.79	10/6753 (0.1%)
2	G	0.37	0/2288	0.64	1/3077 (0.0%)
2	H	0.39	0/2280	0.65	2/3067 (0.1%)
2	I	0.32	0/2295	0.58	1/3086 (0.0%)
2	J	0.32	0/2295	0.53	0/3086
3	K	0.24	0/483	0.42	0/648
4	L	0.23	0/535	0.43	0/715
5	M	0.22	0/1039	0.43	0/1381
All	All	0.40	4/41568 (0.0%)	0.73	60/56148 (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	2
1	B	0	1
1	E	0	1
1	F	0	1
All	All	0	5

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	514	VAL	CB-CG1	6.36	1.66	1.52
1	A	551	ALA	CA-CB	-5.87	1.40	1.52
1	B	547	SER	C-O	5.22	1.33	1.23

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	235	PHE	CD1-CE1	-5.04	1.29	1.39

The worst 5 of 60 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	533	ARG	NE-CZ-NH1	-9.39	115.60	120.30
1	D	547	SER	C-N-CA	-8.93	103.55	122.30
1	E	629	LEU	CB-CG-CD1	-8.25	96.97	111.00
1	C	677	LEU	CB-CG-CD2	-8.21	97.04	111.00
1	C	322	LEU	CA-CB-CG	7.89	133.46	115.30

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	241	PRO	Peptide
1	A	261	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	5022	0	4913	339	0
1	B	4988	0	4912	300	0
1	C	5019	0	4930	335	0
1	D	4971	0	4877	315	0
1	E	4983	0	4898	329	0
1	F	4923	0	4898	327	0
2	G	2249	0	2192	129	0
2	H	2241	0	2182	141	0
2	I	2255	0	2199	121	0
2	J	2255	0	2199	118	0
3	K	480	0	465	65	0
4	L	531	0	522	63	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	M	1039	0	1020	128	0
All	All	40956	0	40207	2408	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 30.

The worst 5 of 2408 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:A:718:LEU:HD13	1:F:533:ARG:NH2	1.60	1.15
3:K:36:GLN:HA	4:L:205:LEU:HD13	1.31	1.11
1:A:490:PRO:HA	1:A:491:ALA:HB3	1.39	1.01
1:B:490:PRO:HA	1:B:491:ALA:HB3	1.47	0.97
2:H:116:ARG:NH2	3:K:68:ASP:OD2	2.01	0.92

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	666/747 (89%)	613 (92%)	36 (5%)	17 (3%)	7 45
1	B	661/747 (88%)	588 (89%)	59 (9%)	14 (2%)	9 50
1	C	666/747 (89%)	616 (92%)	38 (6%)	12 (2%)	11 53
1	D	663/747 (89%)	607 (92%)	46 (7%)	10 (2%)	13 57
1	E	658/747 (88%)	603 (92%)	41 (6%)	14 (2%)	9 50
1	F	644/747 (86%)	588 (91%)	43 (7%)	13 (2%)	9 51
2	G	284/297 (96%)	231 (81%)	41 (14%)	12 (4%)	3 34
2	H	284/297 (96%)	230 (81%)	46 (16%)	8 (3%)	6 44

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	I	284/297 (96%)	227 (80%)	48 (17%)	9 (3%)	5 41
2	J	284/297 (96%)	233 (82%)	40 (14%)	11 (4%)	4 36
3	K	59/63 (94%)	56 (95%)	2 (3%)	1 (2%)	11 55
4	L	64/67 (96%)	56 (88%)	8 (12%)	0	100 100
5	M	127/188 (68%)	124 (98%)	3 (2%)	0	100 100
All	All	5344/5988 (89%)	4772 (89%)	451 (8%)	121 (2%)	12 48

5 of 121 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	320	SER

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	510/638 (80%)	508 (100%)	2 (0%)	93 96
1	B	511/638 (80%)	509 (100%)	2 (0%)	93 96
1	C	511/638 (80%)	509 (100%)	2 (0%)	93 96
1	D	505/638 (79%)	502 (99%)	3 (1%)	90 95
1	E	509/638 (80%)	506 (99%)	3 (1%)	90 95
1	F	512/638 (80%)	507 (99%)	5 (1%)	82 92
2	G	234/244 (96%)	234 (100%)	0	100 100
2	H	232/244 (95%)	232 (100%)	0	100 100
2	I	235/244 (96%)	234 (100%)	1 (0%)	93 96
2	J	235/244 (96%)	234 (100%)	1 (0%)	93 96
3	K	49/54 (91%)	49 (100%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
4	L	59/61 (97%)	59 (100%)	0	100 100
5	M	113/161 (70%)	113 (100%)	0	100 100
All	All	4215/5080 (83%)	4196 (100%)	19 (0%)	92 96

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	370	ILE
1	E	322	LEU
1	F	514	VAL
1	D	316	LEU
1	F	536	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	352	ASN
1	F	194	ASN
5	M	65	ASN
1	E	527	GLN
1	E	659	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.