wwPDB X-ray Structure Validation Summary Report (i)

Jan 20, 2024 – 01:14 pm GMT

PDB ID	:	7OG1
Title	:	AP2 clathrin adaptor core in complex with cargo peptide and FCHO2
Authors	:	Zaccai, N.R.; Kelly, B.T.; Evans, P.R.; Owen, D.J.
Deposited on	:	2021-05-05
Resolution	:	3.25 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:

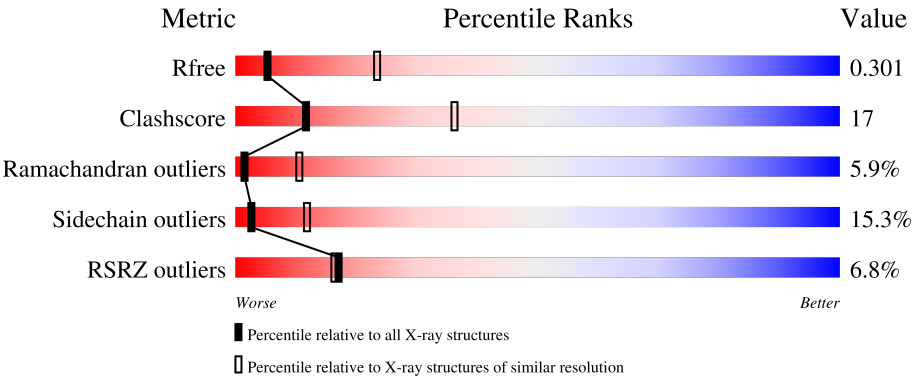
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.25 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.



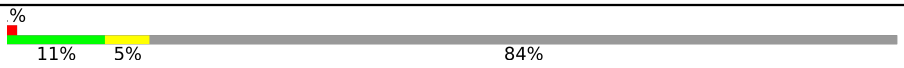
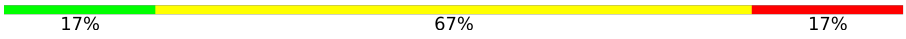
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1191 (3.30-3.22)
Clashscore	141614	1251 (3.30-3.22)
Ramachandran outliers	138981	1229 (3.30-3.22)
Sidechain outliers	138945	1228 (3.30-3.22)
RSRZ outliers	127900	1154 (3.30-3.22)

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	AAA	621	14% 60% 31% 8%
2	BBB	591	2% 51% 35% 9%
3	CCC	446	98%
3	MMM	446	53% 32% 6% 9%
4	SSS	142	14% 43% 47% 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	DDD	152	
5	GGG	152	
6	PPP	6	

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 14391 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 AP-2 complex subunit alpha-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	613	Total	C	N	O	S	0	0	0
			4836	3081	833	901	21			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	272	GLU	-	insertion	UNP P18484

- Molecule 2 is a protein called AP-2 complex subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	BBB	572	Total	C	N	O	S	0	0	0
			4527	2882	752	868	25			

- Molecule 3 is a protein called AP-2 complex subunit mu.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	MMM	406	Total	C	N	O	S	0	0	0
			3254	2087	566	582	19			
3	CCC	9	Total	C	N	O	S	0	0	0
			72	45	12	14	1			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
MMM	237	MET	-	insertion	UNP P84092
MMM	238	GLU	-	insertion	UNP P84092
MMM	239	GLN	-	insertion	UNP P84092
MMM	240	LYS	-	insertion	UNP P84092
MMM	241	LEU	-	insertion	UNP P84092
MMM	242	ILE	-	insertion	UNP P84092
MMM	243	SER	-	insertion	UNP P84092

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
MMM	244	GLU	-	insertion	UNP P84092
MMM	245	GLU	-	insertion	UNP P84092
MMM	246	ASP	-	insertion	UNP P84092
MMM	247	LEU	-	insertion	UNP P84092
CCC	237	MET	-	insertion	UNP P84092
CCC	238	GLU	-	insertion	UNP P84092
CCC	239	GLN	-	insertion	UNP P84092
CCC	240	LYS	-	insertion	UNP P84092
CCC	241	LEU	-	insertion	UNP P84092
CCC	242	ILE	-	insertion	UNP P84092
CCC	243	SER	-	insertion	UNP P84092
CCC	244	GLU	-	insertion	UNP P84092
CCC	245	GLU	-	insertion	UNP P84092
CCC	246	ASP	-	insertion	UNP P84092
CCC	247	LEU	-	insertion	UNP P84092

- Molecule 4 is a protein called AP-2 complex subunit sigma.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	SSS	142	Total	C	N	O	S	0	0	0
			1200	778	200	215	7			

- Molecule 5 is a protein called F-BAR domain only protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	GGG	32	Total	C	N	O	S	0	0	0
			245	153	43	47	2			
5	DDD	24	Total	C	N	O		0	0	0
			200	122	32	46				

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
GGG	309	GLY	-	expression tag	UNP Q0JRZ9
GGG	310	SER	-	expression tag	UNP Q0JRZ9
GGG	311	PRO	-	expression tag	UNP Q0JRZ9
GGG	312	GLU	-	expression tag	UNP Q0JRZ9
GGG	313	PHE	-	expression tag	UNP Q0JRZ9
GGG	445	GLU	-	expression tag	UNP Q0JRZ9
GGG	446	PHE	-	expression tag	UNP Q0JRZ9
GGG	447	PRO	-	expression tag	UNP Q0JRZ9
GGG	448	GLY	-	expression tag	UNP Q0JRZ9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
GGG	449	ARG	-	expression tag	UNP Q0JRZ9
GGG	450	PRO	-	expression tag	UNP Q0JRZ9
GGG	451	HIS	-	expression tag	UNP Q0JRZ9
GGG	452	HIS	-	expression tag	UNP Q0JRZ9
GGG	453	HIS	-	expression tag	UNP Q0JRZ9
GGG	454	HIS	-	expression tag	UNP Q0JRZ9
GGG	455	HIS	-	expression tag	UNP Q0JRZ9
GGG	456	HIS	-	expression tag	UNP Q0JRZ9
GGG	457	HIS	-	expression tag	UNP Q0JRZ9
GGG	458	HIS	-	expression tag	UNP Q0JRZ9
GGG	459	HIS	-	expression tag	UNP Q0JRZ9
GGG	460	HIS	-	expression tag	UNP Q0JRZ9
DDD	309	GLY	-	expression tag	UNP Q0JRZ9
DDD	310	SER	-	expression tag	UNP Q0JRZ9
DDD	311	PRO	-	expression tag	UNP Q0JRZ9
DDD	312	GLU	-	expression tag	UNP Q0JRZ9
DDD	313	PHE	-	expression tag	UNP Q0JRZ9
DDD	445	GLU	-	expression tag	UNP Q0JRZ9
DDD	446	PHE	-	expression tag	UNP Q0JRZ9
DDD	447	PRO	-	expression tag	UNP Q0JRZ9
DDD	448	GLY	-	expression tag	UNP Q0JRZ9
DDD	449	ARG	-	expression tag	UNP Q0JRZ9
DDD	450	PRO	-	expression tag	UNP Q0JRZ9
DDD	451	HIS	-	expression tag	UNP Q0JRZ9
DDD	452	HIS	-	expression tag	UNP Q0JRZ9
DDD	453	HIS	-	expression tag	UNP Q0JRZ9
DDD	454	HIS	-	expression tag	UNP Q0JRZ9
DDD	455	HIS	-	expression tag	UNP Q0JRZ9
DDD	456	HIS	-	expression tag	UNP Q0JRZ9
DDD	457	HIS	-	expression tag	UNP Q0JRZ9
DDD	458	HIS	-	expression tag	UNP Q0JRZ9
DDD	459	HIS	-	expression tag	UNP Q0JRZ9
DDD	460	HIS	-	expression tag	UNP Q0JRZ9

- Molecule 6 is a protein called TGN38 CARGO PEPTIDE.

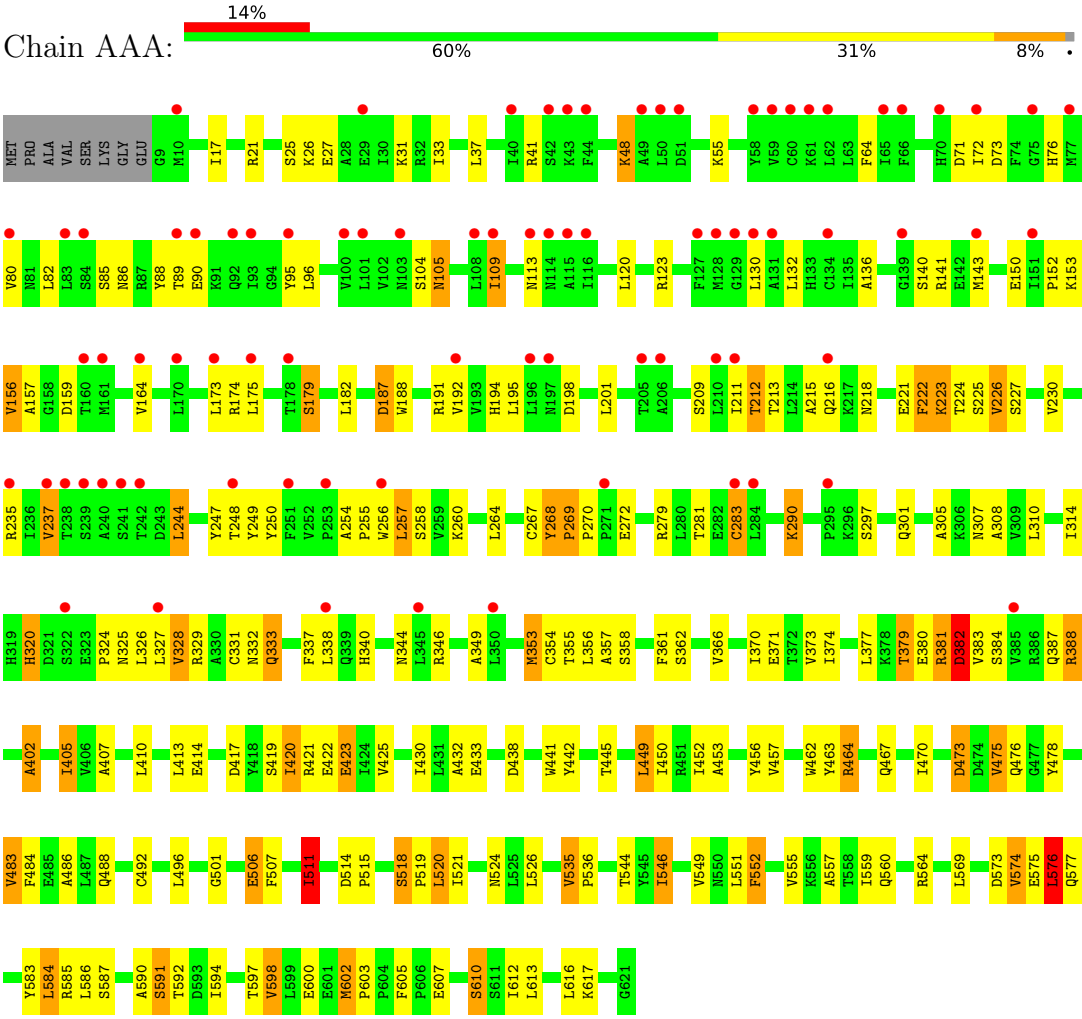
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
6	PPP	6	Total	C	N	O	0	0	0
			57	34	11	12			

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.

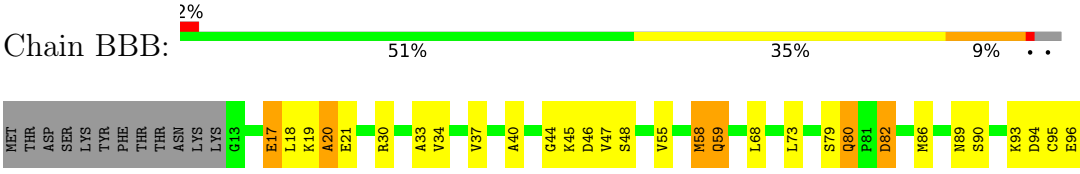
- Molecule 1: AP-2 complex subunit alpha-2

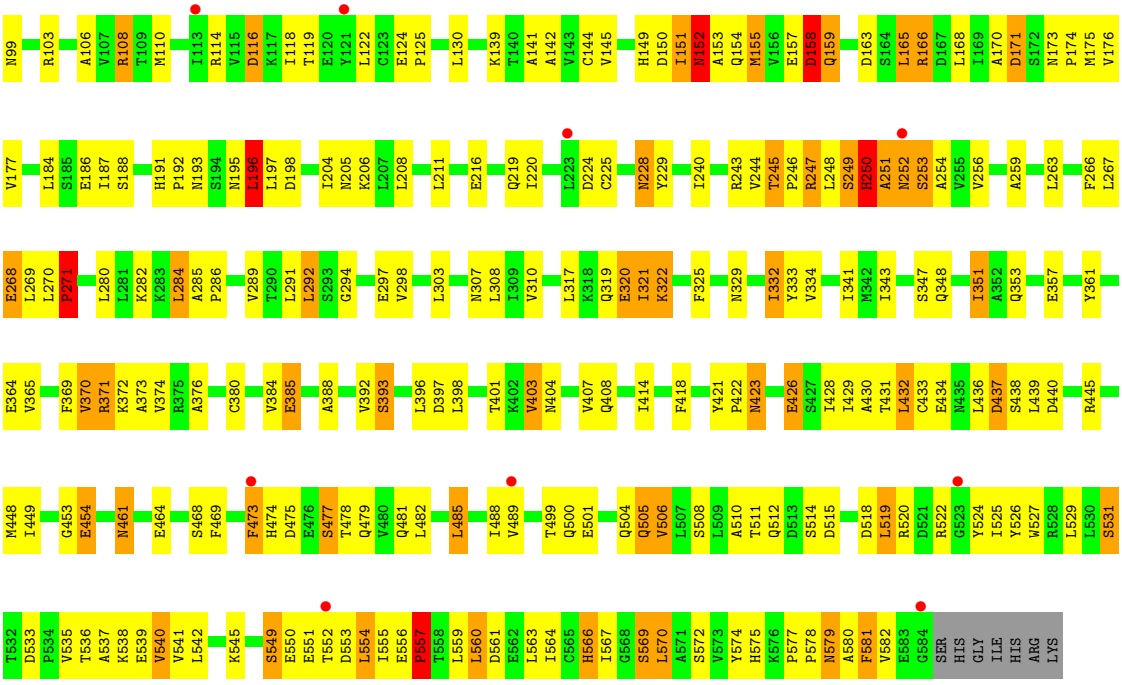
Chain AAA:



- Molecule 2: AP-2 complex subunit beta

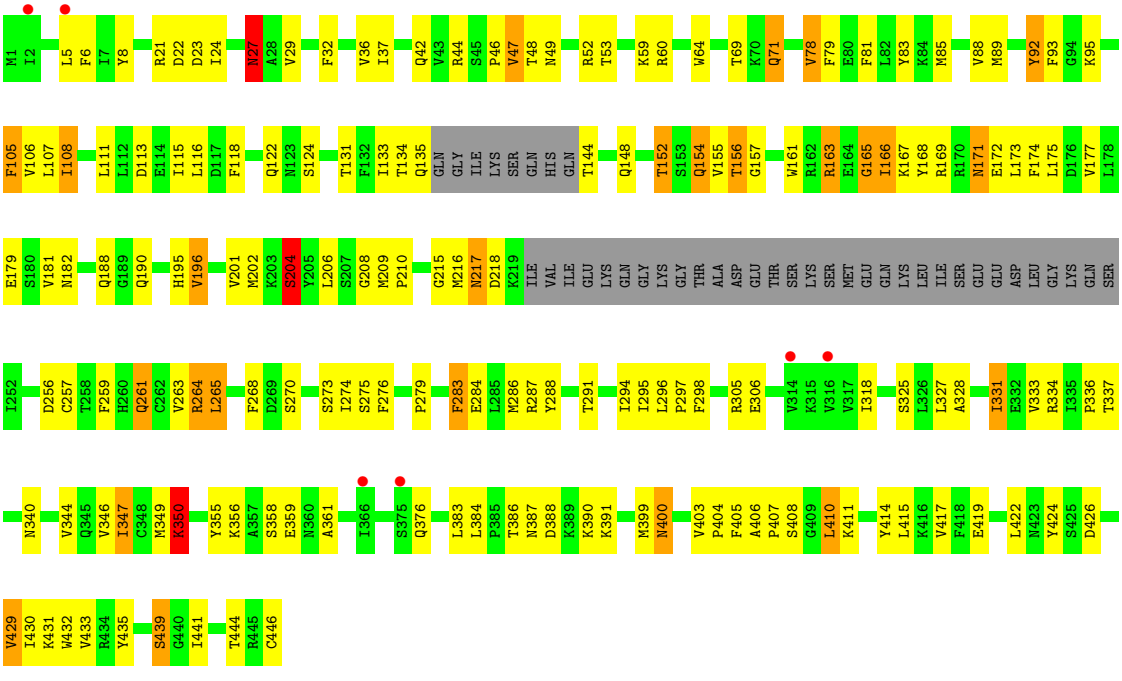
Chain BBB:





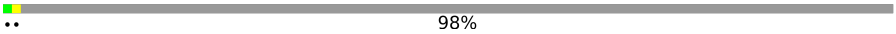
• Molecule 3: AP-2 complex subunit mu

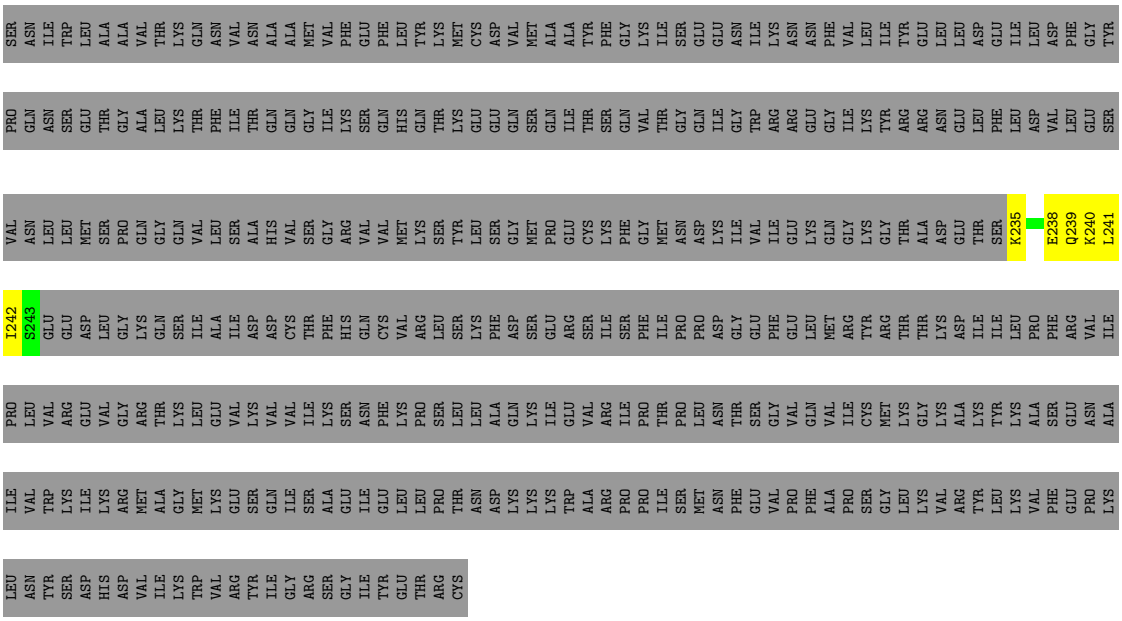
Chain MMM:



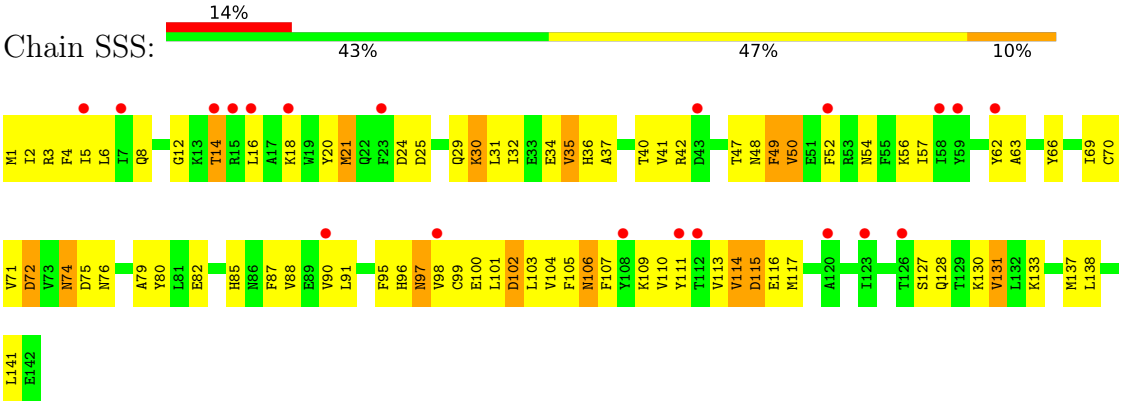
• Molecule 3: AP-2 complex subunit mu

Chain CCC:

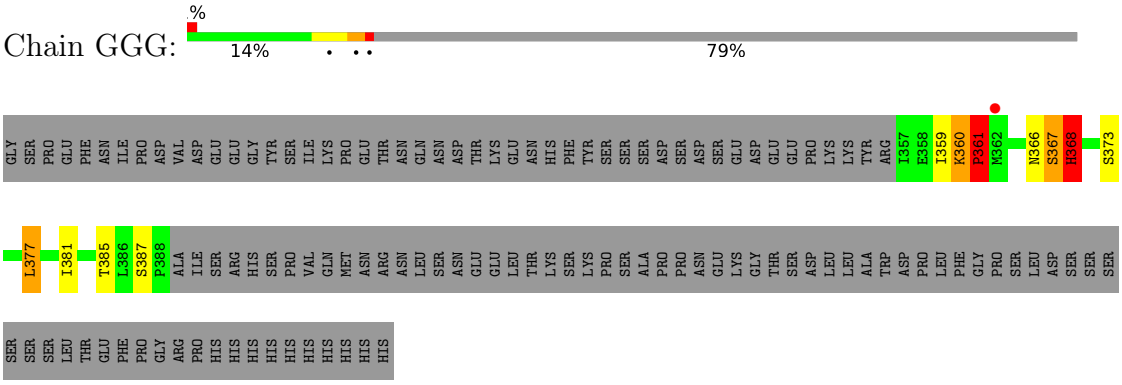




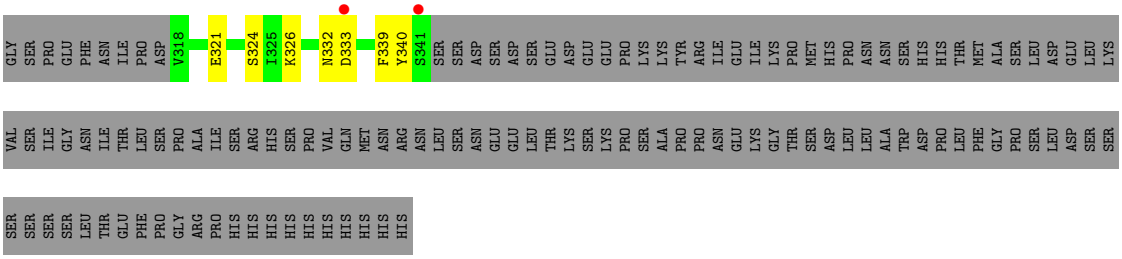
• Molecule 4: AP-2 complex subunit sigma



• Molecule 5: F-BAR domain only protein 2

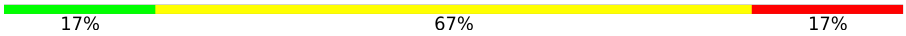


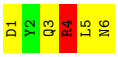
• Molecule 5: F-BAR domain only protein 2



• Molecule 6: TGN38 CARGO PEPTIDE

Chain PPP:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	92.58Å 150.07Å 96.43Å 90.00° 112.65° 90.00°	Depositor
Resolution (Å)	76.66 – 3.25 76.54 – 3.25	Depositor EDS
% Data completeness (in resolution range)	99.9 (76.66-3.25) 100.0 (76.54-3.25)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.28 (at 3.26Å)	Xtriage
Refinement program	REFMAC 5.8.0267, PHENIX v1.19	Depositor
R, R_{free}	0.207 , 0.306 0.209 , 0.301	Depositor DCC
R_{free} test set	1881 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	114.5	Xtriage
Anisotropy	0.353	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 103.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.000 for l,-k,h	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	14391	wwPDB-VP
Average B, all atoms (Å^2)	145.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.83% 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	AAA	0.72	0/4921	0.85	0/6669
2	BBB	0.72	0/4597	0.90	0/6237
3	CCC	0.81	0/71	0.79	0/91
3	MMM	0.72	0/3319	0.89	0/4473
4	SSS	0.72	0/1224	0.83	0/1650
5	DDD	0.74	0/204	0.84	0/275
5	GGG	0.76	0/250	0.97	1/339 (0.3%)
6	PPP	0.63	0/57	0.77	0/74
All	All	0.72	0/14643	0.88	1/19808 (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
2	BBB	0	4
5	GGG	0	1
All	All	0	5

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	GGG	361	PRO	N-CA-CB	-5.33	96.73	102.60

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	BBB	158	ASP	Peptide
2	BBB	250	HIS	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
2	BBB	329	ASN	Peptide
2	BBB	549	SER	Peptide
5	GGG	368	HIS	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	AAA	4836	0	4952	158	0
2	BBB	4527	0	4638	202	0
3	CCC	72	0	80	3	0
3	MMM	3254	0	3318	114	0
4	SSS	1200	0	1195	68	0
5	DDD	200	0	171	3	0
5	GGG	245	0	247	14	0
6	PPP	57	0	53	4	0
All	All	14391	0	14654	507	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 17.

The worst 5 of 507 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:BBB:122:LEU:O	2:BBB:125:PRO:HD2	1.66	0.96
2:BBB:249:SER:HB3	3:MMM:259:PHE:H	1.30	0.94
2:BBB:165:LEU:HD23	2:BBB:184:LEU:HD13	1.50	0.93
3:MMM:433:VAL:HG23	6:PPP:5:LEU:HD22	1.48	0.92
3:MMM:175:LEU:HD11	3:MMM:415:LEU:HD22	1.56	0.87

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 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	AAA	611/621 (98%)	465 (76%)	116 (19%)	30 (5%)	2 14
2	BBB	570/591 (96%)	415 (73%)	113 (20%)	42 (7%)	1 7
3	CCC	7/446 (2%)	4 (57%)	2 (29%)	1 (14%)	0 1
3	MMM	400/446 (90%)	290 (72%)	93 (23%)	17 (4%)	2 16
4	SSS	140/142 (99%)	103 (74%)	27 (19%)	10 (7%)	1 7
5	DDD	22/152 (14%)	13 (59%)	8 (36%)	1 (4%)	2 15
5	GGG	30/152 (20%)	23 (77%)	4 (13%)	3 (10%)	0 3
6	PPP	4/6 (67%)	3 (75%)	0	1 (25%)	0 0
All	All	1784/2556 (70%)	1316 (74%)	363 (20%)	105 (6%)	1 10

5 of 105 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	105	ASN
1	AAA	224	THR
1	AAA	269	PRO
1	AAA	362	SER
1	AAA	380	GLU

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	AAA	537/543 (99%)	458 (85%)	79 (15%)	3 13

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	BBB	513/532 (96%)	440 (86%)	73 (14%)	3 15
3	CCC	9/398 (2%)	7 (78%)	2 (22%)	1 3
3	MMM	358/398 (90%)	297 (83%)	61 (17%)	2 9
4	SSS	131/131 (100%)	113 (86%)	18 (14%)	3 16
5	DDD	23/142 (16%)	19 (83%)	4 (17%)	2 8
5	GGG	30/142 (21%)	24 (80%)	6 (20%)	1 5
6	PPP	6/6 (100%)	3 (50%)	3 (50%)	0 0
All	All	1607/2292 (70%)	1361 (85%)	246 (15%)	2 12

5 of 246 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	BBB	365	VAL
4	SSS	54	ASN
2	BBB	549	SER
4	SSS	30	LYS
5	GGG	387	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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	AAA	613/621 (98%)	0.61	84 (13%) 3 3	84, 165, 244, 320	0
2	BBB	572/591 (96%)	0.18	9 (1%) 72 69	82, 116, 165, 244	0
3	CCC	9/446 (2%)	-0.65	0 100 100	164, 179, 205, 206	0
3	MMM	406/446 (91%)	0.11	6 (1%) 73 71	80, 123, 185, 245	0
4	SSS	142/142 (100%)	0.70	20 (14%) 2 2	134, 174, 223, 242	0
5	DDD	24/152 (15%)	0.19	2 (8%) 11 11	162, 191, 212, 215	0
5	GGG	32/152 (21%)	-0.29	1 (3%) 49 47	91, 141, 168, 184	0
6	PPP	6/6 (100%)	-0.24	0 100 100	121, 141, 155, 174	0
All	All	1804/2556 (70%)	0.34	122 (6%) 17 16	80, 134, 224, 320	0

The worst 5 of 122 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	103	ASN	7.6
1	AAA	113	ASN	7.6
1	AAA	93	ILE	6.7
1	AAA	51	ASP	6.3
1	AAA	130	LEU	5.8

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

There are no such residues in this entry.