wwPDB X-ray Structure Validation Summary Report (i)

Apr 28, 2024 – 03:36 am BST

PDB ID	:	2CAZ
Title	:	ESCRT-I core
Authors	:	Gill, D.J.; Teo, H.; Sun, J.; Perisic, O.; Veprintsev, D.B.; Vallis, Y.; Emr, S.D.; Williams, R.L.
Deposited on	:	2005-12-23
Resolution	:	3.60 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 (1)) were used in the production of this report:

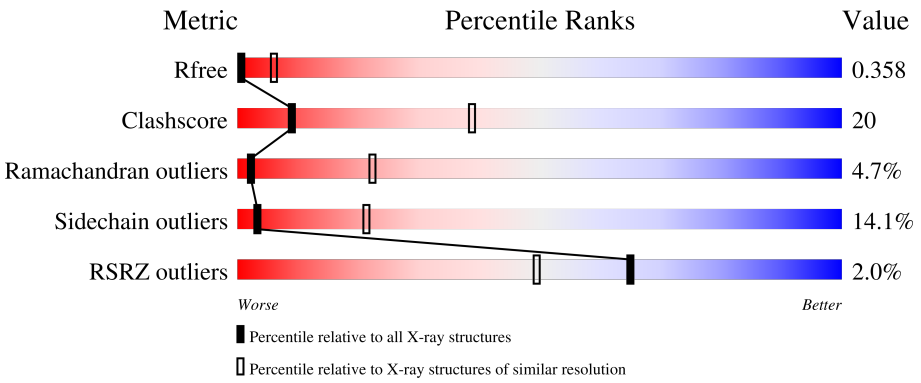
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 3.60 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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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	82	49% 17% 5% • 28%
1	D	82	48% 22% • • 26%
2	B	155	% 30% 29% 5% • 35%
2	E	155	% 26% 31% 8% • 34%
3	C	85	5% 47% 22% 6% • 24%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	85	4% 45% 14% 5% 36%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3602 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 SUPPRESSOR PROTEIN STP22 OF TEMPERATURE-SENSITIVE ALPHA-FACTOR RECEPTOR AND ARGININE PERMEASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	59	484	306	91	85	2	0	0	1
1	D	61	499	317	93	87	2	0	0	1

- Molecule 2 is a protein called VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN VPS28.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	101	813	515	131	166	1	0	0	1
2	E	102	820	520	132	167	1	0	0	1

- Molecule 3 is a protein called PROTEIN SRN2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	65	536	331	93	112	0	0	1
3	F	54	450	277	81	92	0	0	1

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	140	ASP	GLU	conflict	UNP Q99176
C	143	GLU	LYS	conflict	UNP Q99176

3 Residue-property plots (i)

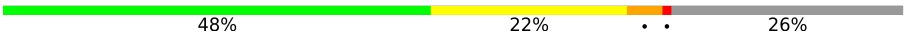
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.

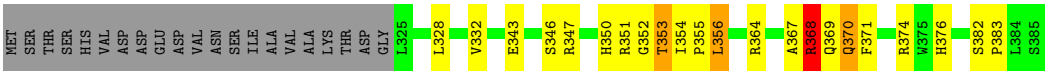
• Molecule 1: SUPPRESSOR PROTEIN STP22 OF TEMPERATURE-SENSITIVE ALPHA-FACTOR RECEPTOR AND ARGININE PERMEASE

Chain A:

• Molecule 1: SUPPRESSOR PROTEIN STP22 OF TEMPERATURE-SENSITIVE ALPHA-FACTOR RECEPTOR AND ARGININE PERMEASE

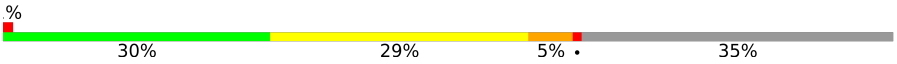
Chain D:

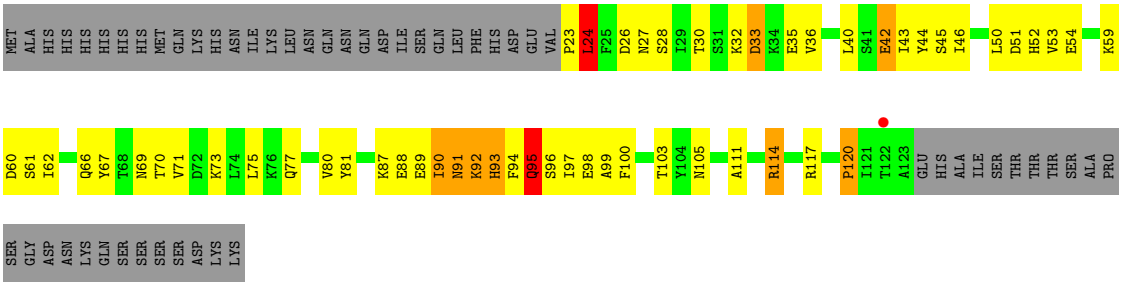




• Molecule 2: VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN VPS28

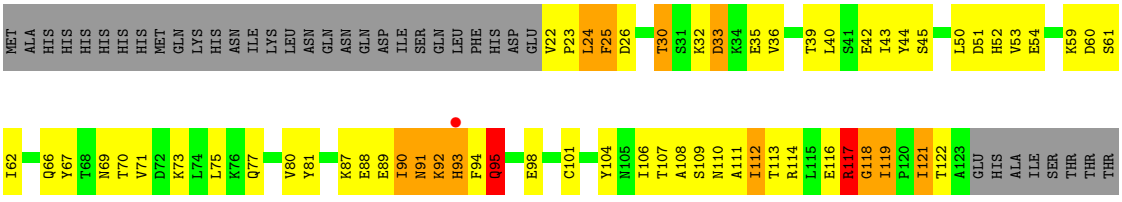
Chain B:

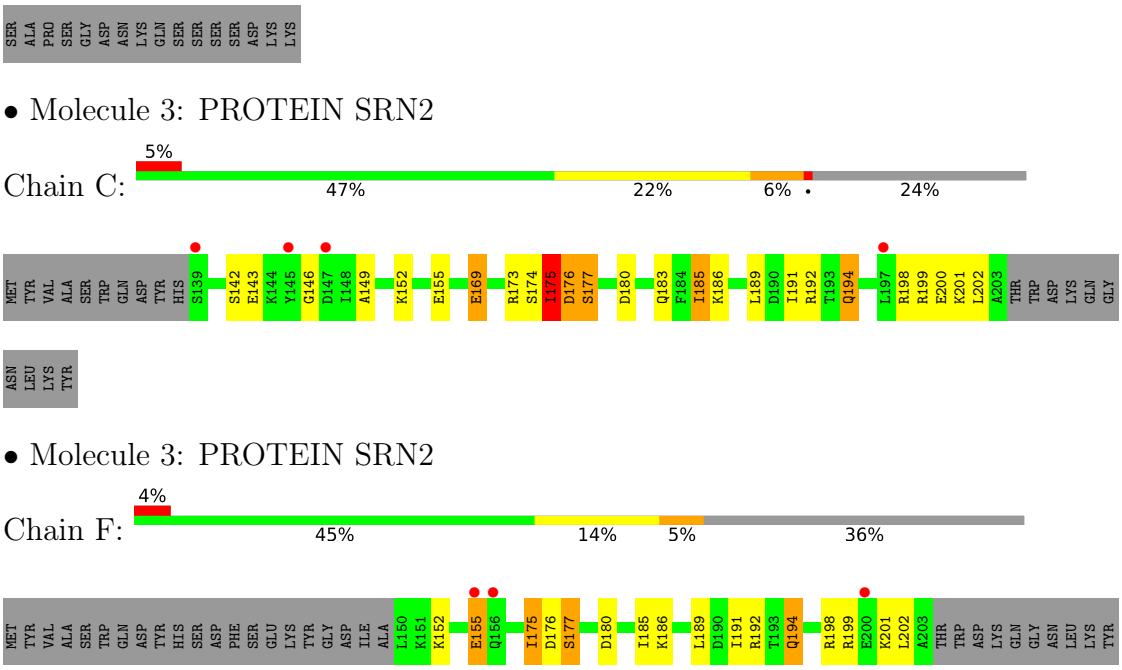




• Molecule 2: VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN VPS28

Chain E:




SER ALA PRO SER GLY ASP ASN LYS GLN SER SER SER SER ASP LYS LYS
• Molecule 3: PROTEIN SRN2
Chain C:
5%
47%
22%
6%
24%
MET TYR VAL ALA SER TRP GLN ASP TYR HIS S139 S142 E143 K144 Y145 G146 D147 I148 A149 K152 E155 E169 R173 S174 I176 D176 S177 D180 Q183 F184 I185 K186 L189 D190 I191 R192 T193 Q194 L197 R198 R199 E200 K201 L202 A203 THR TRP ASP LYS GLN GLY
ASN LEU LYS TYR
• Molecule 3: PROTEIN SRN2
Chain F:
4%
45%
14%
5%
36%
MET TYR VAL ALA SER TRP GLN ASP TYR HIS SER ASP PHE SER GLU LYS TYR GLY ASP ILE ALA L150 K151 K152 E155 Q156 I175 D176 S177 D180 I185 K186 L189 D190 I191 R192 T193 Q194 R198 R199 E200 K201 L202 A203 THR TRP ASP LYS GLN GLY ASN LEU LYS TYR

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α, β, γ	167.91Å 167.91Å 50.24Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	145.86 – 3.60 43.11 – 3.60	Depositor EDS
% Data completeness (in resolution range)	99.9 (145.86-3.60) 99.9 (43.11-3.60)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.45 (at 3.57Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.328 , 0.357 0.328 , 0.358	Depositor DCC
R_{free} test set	465 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å^2)	134.8	Xtriage
Anisotropy	0.420	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 27.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.049 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	3602	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.57% 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 (i)

5.1 Standard geometry (i)

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	1.29	5/492 (1.0%)	0.81	3/666 (0.5%)
1	D	0.84	2/508 (0.4%)	0.81	2/689 (0.3%)
2	B	0.55	0/825	0.63	0/1117
2	E	0.56	0/832	0.66	0/1128
3	C	0.83	1/541 (0.2%)	0.57	0/723
3	F	0.45	0/453	0.60	0/605
All	All	0.77	8/3651 (0.2%)	0.68	5/4928 (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
2	E	0	1

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	368	ARG	CZ-NH2	20.61	1.59	1.33
3	C	142	SER	CB-OG	16.11	1.63	1.42
1	D	368	ARG	CZ-NH2	12.78	1.49	1.33
1	A	326	ASN	CG-ND2	11.83	1.62	1.32
1	A	368	ARG	NE-CZ	7.28	1.42	1.33

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	368	ARG	NE-CZ-NH2	-12.34	114.13	120.30
1	A	326	ASN	OD1-CG-ND2	10.52	146.10	121.90
1	A	326	ASN	CB-CG-ND2	-6.26	101.68	116.70
1	D	368	ARG	NH1-CZ-NH2	5.73	125.70	119.40
1	A	368	ARG	CD-NE-CZ	5.04	130.66	123.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	E	24	LEU	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	484	0	491	27	0
1	D	499	0	509	27	0
2	B	813	0	809	43	0
2	E	820	0	817	55	0
3	C	536	0	529	13	0
3	F	450	0	455	10	0
All	All	3602	0	3610	146	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 20.

The worst 5 of 146 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:40:LEU:HA	2:B:43:ILE:HD12	1.42	1.02
2:E:40:LEU:HA	2:E:43:ILE:HD12	1.49	0.95
1:D:368:ARG:HG3	1:D:368:ARG:HH11	1.32	0.92
2:B:54:GLU:OE1	2:B:114:ARG:HD2	1.75	0.85
2:E:42:GLU:HG2	2:E:81:TYR:CZ	2.12	0.84

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	A	57/82 (70%)	50 (88%)	7 (12%)	0	100	100
1	D	59/82 (72%)	50 (85%)	9 (15%)	0	100	100
2	B	99/155 (64%)	69 (70%)	23 (23%)	7 (7%)	1	14
2	E	100/155 (64%)	65 (65%)	28 (28%)	7 (7%)	1	14
3	C	63/85 (74%)	44 (70%)	15 (24%)	4 (6%)	1	17
3	F	52/85 (61%)	37 (71%)	13 (25%)	2 (4%)	3	27
All	All	430/644 (67%)	315 (73%)	95 (22%)	20 (5%)	2	22

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	90	ILE
2	B	92	LYS
3	C	175	ILE
2	E	92	LYS
2	E	95	GLN

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	53/74 (72%)	46 (87%)	7 (13%)	4	23
1	D	55/74 (74%)	49 (89%)	6 (11%)	6	32
2	B	95/145 (66%)	81 (85%)	14 (15%)	3	20

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	E	96/145 (66%)	79 (82%)	17 (18%)	2	11
3	C	61/79 (77%)	53 (87%)	8 (13%)	4	23
3	F	52/79 (66%)	46 (88%)	6 (12%)	5	29
All	All	412/596 (69%)	354 (86%)	58 (14%)	3	21

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	202	LEU
3	F	189	LEU
2	E	26	ASP
3	F	176	ASP
2	E	119	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	27	ASN
2	E	77	GLN
2	E	95	GLN
2	E	91	ASN
2	E	93	HIS

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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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	59/82 (71%)	-0.07	0	100	100	72, 72, 72, 72	0
1	D	61/82 (74%)	0.04	0	100	100	72, 72, 72, 72	0
2	B	101/155 (65%)	-0.14	1 (0%)	82	70	72, 72, 72, 72	0
2	E	102/155 (65%)	-0.16	1 (0%)	82	70	72, 72, 72, 72	0
3	C	65/85 (76%)	0.30	4 (6%)	20	12	72, 72, 72, 72	0
3	F	54/85 (63%)	0.26	3 (5%)	24	14	72, 72, 72, 72	0
All	All	442/644 (68%)	0.00	9 (2%)	65	49	72, 72, 72, 72	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	139	SER	6.4
3	C	147	ASP	3.3
3	F	156	GLN	2.9
3	F	200	GLU	2.5
3	C	197	LEU	2.4

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.