



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 16, 2023 – 12:10 PM EST

PDB ID : 4P6Z
Title : Crystal structure of the human BST2 cytoplasmic domain and the HIV-1 Vpu cytoplasmic domain bound to the clathrin adaptor protein complex 1 (AP1) core
Authors : Jia, X.; Xiong, Y.
Deposited on : 2014-03-25
Resolution : 3.00 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 ⓘ 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

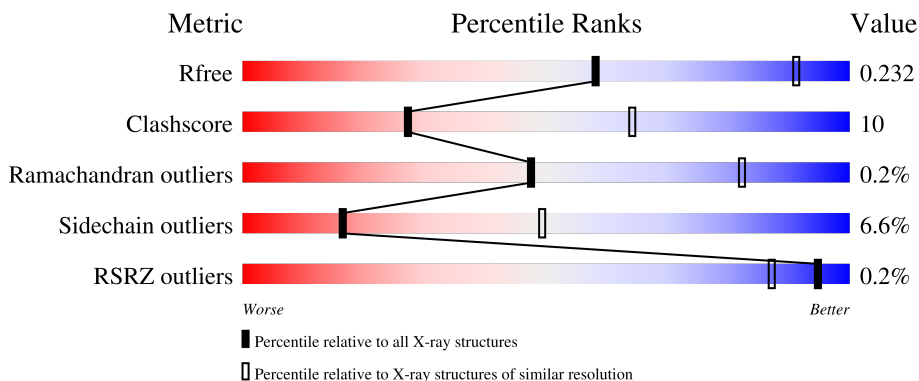
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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 $\leq 5\%$. 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	G	627	
2	S	158	
3	M	423	
4	B	600	
5	V	63	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	T	25	 32% 8% 56%

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 13940 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-1 complex subunit gamma-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	G	587	4645	2922	817	867	39	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-13	MET	-	initiating methionine	UNP P22892
G	-12	GLY	-	expression tag	UNP P22892
G	-11	SER	-	expression tag	UNP P22892
G	-10	SER	-	expression tag	UNP P22892
G	-9	HIS	-	expression tag	UNP P22892
G	-8	HIS	-	expression tag	UNP P22892
G	-7	HIS	-	expression tag	UNP P22892
G	-6	HIS	-	expression tag	UNP P22892
G	-5	HIS	-	expression tag	UNP P22892
G	-4	HIS	-	expression tag	UNP P22892
G	-3	SER	-	expression tag	UNP P22892
G	-2	GLN	-	expression tag	UNP P22892
G	-1	ASP	-	expression tag	UNP P22892
G	0	PRO	-	expression tag	UNP P22892

- Molecule 2 is a protein called AP-1 complex subunit sigma-1A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	S	147	1231	796	203	221	11	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
S	11	ARG	GLN	engineered mutation	UNP P61966

- Molecule 3 is a protein called AP-1 complex subunit mu-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	M	415	3370	2166	569	621	14	0	1	0

- Molecule 4 is a protein called AP-1 complex subunit beta-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	B	570	4509	2882	739	861	27	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-15	MET	-	initiating methionine	UNP Q10567
B	-14	GLY	-	expression tag	UNP Q10567
B	-13	SER	-	expression tag	UNP Q10567
B	-12	SER	-	expression tag	UNP Q10567
B	-11	HIS	-	expression tag	UNP Q10567
B	-10	HIS	-	expression tag	UNP Q10567
B	-9	HIS	-	expression tag	UNP Q10567
B	-8	HIS	-	expression tag	UNP Q10567
B	-7	HIS	-	expression tag	UNP Q10567
B	-6	HIS	-	expression tag	UNP Q10567
B	-5	SER	-	expression tag	UNP Q10567
B	-4	GLN	-	expression tag	UNP Q10567
B	-3	ASP	-	expression tag	UNP Q10567
B	-2	PRO	-	expression tag	UNP Q10567
B	-1	ASN	-	expression tag	UNP Q10567
B	0	SER	-	expression tag	UNP Q10567
B	431	ALA	THR	engineered mutation	UNP Q10567
B	476	LYS	GLU	engineered mutation	UNP Q10567

- Molecule 5 is a protein called Protein Vpu.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
5	V	10	78	47	11	20	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
V	18	GLY	-	expression tag	UNP P19554

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
V	19	SER	-	expression tag	UNP P19554
V	20	ASP	-	expression tag	UNP P19554
V	21	GLU	-	expression tag	UNP P19554
V	22	ALA	-	expression tag	UNP P19554
V	23	SER	-	expression tag	UNP P19554
V	24	GLU	-	expression tag	UNP P19554
V	25	GLY	-	expression tag	UNP P19554
V	26	SER	-	expression tag	UNP P19554
V	27	GLY	-	expression tag	UNP P19554

- Molecule 6 is a protein called Bone marrow stromal antigen 2.


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	T	11	93	58	14	19	2	0	0	0

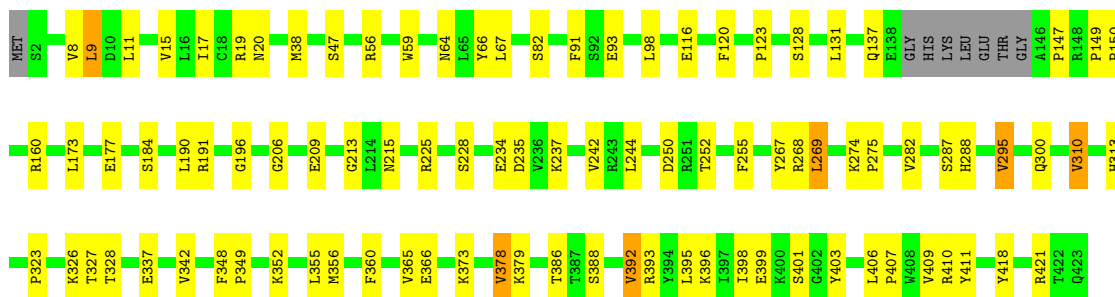
There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
T	-3	ALA	-	expression tag	UNP Q10589
T	-2	GLY	-	expression tag	UNP Q10589
T	-1	PHE	-	expression tag	UNP Q10589
T	0	SER	-	expression tag	UNP Q10589

- Molecule 7 is water.

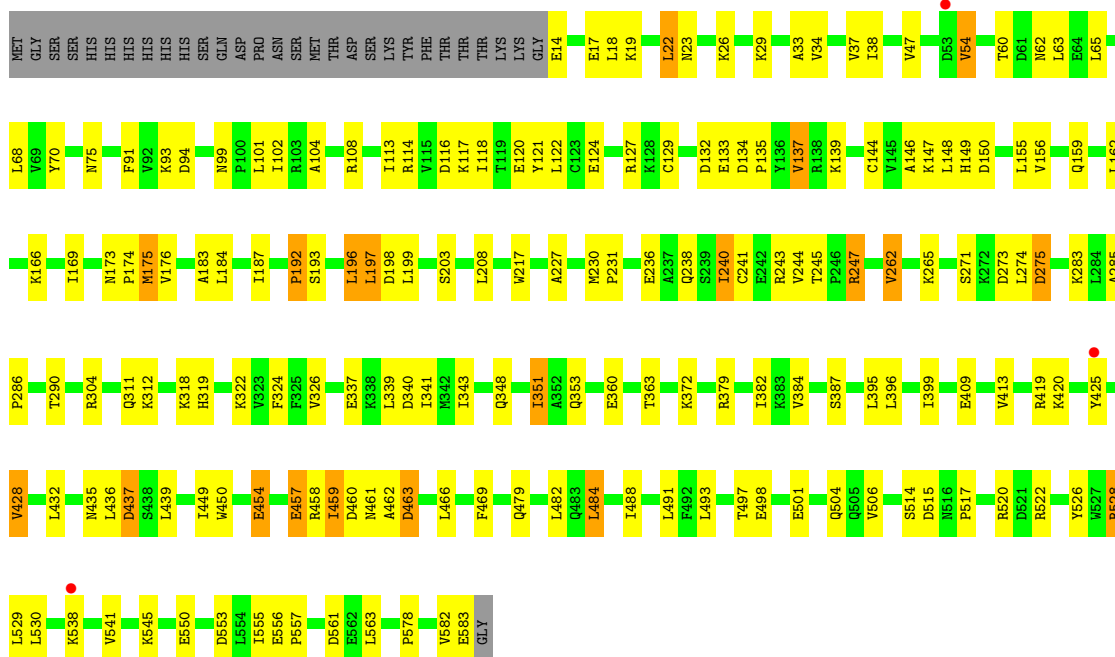
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	G	4	Total 4	O 4	0	0
7	S	3	Total 3	O 3	0	0
7	M	5	Total 5	O 5	0	0
7	B	2	Total 2	O 2	0	0

Chain M:  76% 21% ..



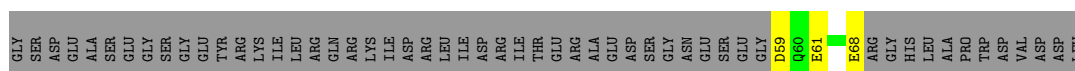
• Molecule 4: AP-1 complex subunit beta-1

Chain B:  66% 26% • 5%




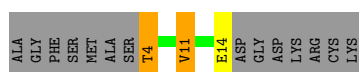
• Molecule 5: Protein Vpu

Chain V:  11% 5% 84%



• Molecule 6: Bone marrow stromal antigen 2

Chain T:  32% 8% 56%



4 Data and refinement statistics i

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α , β , γ	160.54Å 160.54Å 118.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.00 48.76 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.2 (50.00-3.00) 99.4 (48.76-3.00)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.37 (at 3.01Å)	Xtrriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.186 , 0.229 0.189 , 0.232	Depositor DCC
R_{free} test set	3036 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å ²)	89.2	Xtrriage
Anisotropy	0.128	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 57.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.031 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	13940	wwPDB-VP
Average B, all atoms (Å ²)	107.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	G	0.62	1/4710 (0.0%)	0.79	5/6357 (0.1%)
2	S	0.72	0/1252	0.77	1/1674 (0.1%)
3	M	0.78	0/3450	0.74	2/4662 (0.0%)
4	B	0.68	1/4579 (0.0%)	0.73	4/6210 (0.1%)
5	V	0.84	0/77	0.75	0/103
6	T	0.76	0/95	0.90	0/128
All	All	0.69	2/14163 (0.0%)	0.76	12/19134 (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
4	B	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	528	ARG	NE-CZ	20.28	1.59	1.33
1	G	83	ARG	CZ-NH2	-11.33	1.18	1.33

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	83	ARG	NE-CZ-NH1	26.94	133.77	120.30
4	B	528	ARG	NE-CZ-NH2	-19.83	110.39	120.30
1	G	83	ARG	NE-CZ-NH2	-18.02	111.29	120.30
4	B	528	ARG	CD-NE-CZ	-8.26	112.03	123.60
1	G	221	ARG	NE-CZ-NH1	5.91	123.25	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	B	528	ARG	Sidechain

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	G	4645	0	4768	121	0
2	S	1231	0	1259	27	0
3	M	3370	0	3386	55	0
4	B	4509	0	4645	108	0
5	V	78	0	70	2	0
6	T	93	0	82	3	0
7	B	2	0	0	0	0
7	G	4	0	0	0	0
7	M	5	0	0	0	0
7	S	3	0	0	0	0
All	All	13940	0	14210	284	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 10.

The worst 5 of 284 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)
3:M:149:PRO:HD3	4:B:101:LEU:HB2	1.50	0.94
2:S:11:ARG:HH22	5:V:59:ASP:HB2	1.35	0.90
2:S:11:ARG:NH2	5:V:59:ASP:HB2	1.88	0.88
3:M:410:ARG:HH22	6:T:4:THR:HA	1.40	0.85
1:G:80:THR:HG23	2:S:142:GLN:HE21	1.42	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	G	585/627 (93%)	547 (94%)	36 (6%)	2 (0%)	41	76
2	S	145/158 (92%)	142 (98%)	2 (1%)	1 (1%)	22	60
3	M	412/423 (97%)	400 (97%)	12 (3%)	0	100	100
4	B	568/600 (95%)	542 (95%)	25 (4%)	1 (0%)	47	82
5	V	8/63 (13%)	7 (88%)	1 (12%)	0	100	100
6	T	9/25 (36%)	9 (100%)	0	0	100	100
All	All	1727/1896 (91%)	1647 (95%)	76 (4%)	4 (0%)	47	82

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	452	HIS
2	S	120	GLY
1	G	490	GLY
4	B	192	PRO

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	G	521/557 (94%)	486 (93%)	35 (7%)	16	49
2	S	135/144 (94%)	127 (94%)	8 (6%)	19	54
3	M	378/383 (99%)	362 (96%)	16 (4%)	30	66

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	B	509/536 (95%)	470 (92%)	39 (8%)	13	42
5	V	9/53 (17%)	7 (78%)	2 (22%)	1	4
6	T	11/21 (52%)	8 (73%)	3 (27%)	0	2
All	All	1563/1694 (92%)	1460 (93%)	103 (7%)	16	49

5 of 103 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	M	392	VAL
4	B	247	ARG
5	V	68	GLU
4	B	22	LEU
4	B	155	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	G	340	HIS
1	G	374	ASN
4	B	195	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 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	G	587/627 (93%)	-0.29	1 (0%) 95 87	55, 126, 180, 212	0
2	S	147/158 (93%)	-0.45	0 100 100	55, 90, 119, 170	0
3	M	415/423 (98%)	-0.45	0 100 100	47, 75, 113, 163	0
4	B	570/600 (95%)	-0.33	3 (0%) 91 75	58, 107, 168, 224	0
5	V	10/63 (15%)	-0.29	0 100 100	83, 87, 135, 144	0
6	T	11/25 (44%)	0.46	0 100 100	86, 111, 134, 142	0
All	All	1740/1896 (91%)	-0.35	4 (0%) 95 87	47, 103, 169, 224	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	278	THR	3.2
4	B	53	ASP	2.2
4	B	538	LYS	2.2
4	B	425	TYR	2.0

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.