



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 8, 2016 – 05:00 PM EDT

PDB ID : 5KYU
Title : crystal structure of Sec23 and TANGO1 peptide2 complex
Authors : Ma, W.; Goldberg, J.
Deposited on : 2016-07-22
Resolution : 3.51 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : unknown
Xtriage (Phenix) : 1.9-1692
EDS : rb-20027939
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027939

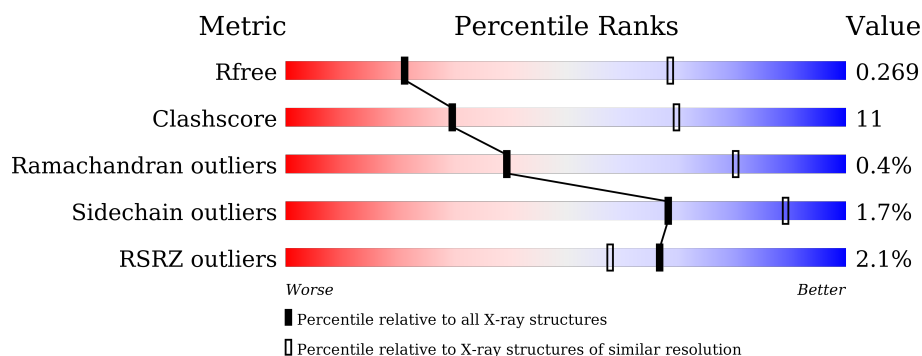
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.51 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}	91344	1089 (3.64-3.40)
Clashscore	102246	1197 (3.64-3.40)
Ramachandran outliers	100387	1159 (3.64-3.40)
Sidechain outliers	100360	1160 (3.64-3.40)
RSRZ outliers	91569	1096 (3.64-3.40)

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 on the lower 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 $\leq 5\%$. 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	A	765	<div> <div>%</div> <div> <div></div> <div>76%</div> <div>16%</div> <div>• 6%</div> </div> </div>
2	B	767	<div> <div>3%</div> <div> <div></div> <div>76%</div> <div>22%</div> <div>•</div> </div> </div>
3	C	8	<div> <div> <div>25%</div> <div>13%</div> <div>63%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	ZN	B	1101	-	-	X	-

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 11605 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 Protein transport protein Sec23A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	716	Total	C	N	O	S	0	0	0
			5678	3617	975	1046	40			

- Molecule 2 is a protein called Protein transport protein Sec24D.

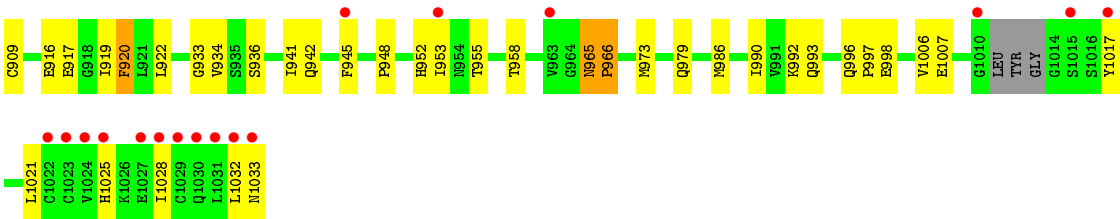
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	764	Total	C	N	O	S	0	0	0
			5904	3758	993	1101	52			

- Molecule 3 is a protein called TANGO1 peptide2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	3	Total	C	N	O	0	0	0
			21	15	3	3			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Zn	0	0
			1	1		
4	A	1	Total	Zn	0	0
			1	1		



● Molecule 3: TANGO1 peptide3



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	103.43Å 140.93Å 150.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.55 – 3.51 48.55 – 3.51	Depositor EDS
% Data completeness (in resolution range)	97.2 (48.55-3.51) 91.3 (48.55-3.51)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.36 (at 3.48Å)	Xtriage
Refinement program	PHENIX (1.10 _2155: ???)	Depositor
R, R_{free}	0.217 , 0.267 0.227 , 0.269	Depositor DCC
R_{free} test set	1879 reflections (7.32%)	DCC
Wilson B-factor (Å ²)	113.4	Xtriage
Anisotropy	0.626	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 105.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	11605	wwPDB-VP
Average B, all atoms (Å ²)	148.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.27% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.31	0/5811	0.46	0/7870
2	B	0.33	0/6029	0.51	0/8181
3	C	0.27	0/23	0.35	0/32
All	All	0.32	0/11863	0.49	0/16083

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	5678	0	5616	97	0
2	B	5904	0	5808	161	0
3	C	21	0	21	1	0
4	A	1	0	0	0	0
4	B	1	0	0	2	0
All	All	11605	0	11445	257	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 11.

The worst 5 of 257 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:385:GLN:HE21	2:B:390:ASN:HB3	1.32	0.93
2:B:389:CYS:HG	4:B:1101:ZN:ZN	0.75	0.92
1:A:15:ASP:OD1	1:A:115:SER:OG	1.90	0.90
2:B:403:LEU:HD12	2:B:404:ASP:H	1.35	0.89
2:B:916:GLU:O	2:B:933:GLY:HA3	1.75	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	A	708/765 (92%)	675 (95%)	30 (4%)	3 (0%)	39	80
2	B	760/767 (99%)	698 (92%)	59 (8%)	3 (0%)	39	80
3	C	1/8 (12%)	1 (100%)	0	0	100	100
All	All	1469/1540 (95%)	1374 (94%)	89 (6%)	6 (0%)	39	80

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	35	VAL
1	A	52	PRO
2	B	966	PRO
1	A	125	PRO
2	B	965	ASN

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	624/666 (94%)	615 (99%)	9 (1%)	74	91
2	B	651/681 (96%)	638 (98%)	13 (2%)	63	87
3	C	3/7 (43%)	3 (100%)	0	100	100
All	All	1278/1354 (94%)	1256 (98%)	22 (2%)	68	89

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	307	ASN
2	B	398	PHE
2	B	834	LEU
2	B	323	THR
2	B	325	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	390	ASN
2	B	824	ASN
2	B	412	HIS
2	B	365	ASN
2	B	393	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

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.

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	716/765 (93%)	-0.04	6 (0%) 87 80	60, 139, 207, 258	0
2	B	764/767 (99%)	0.07	25 (3%) 50 41	86, 139, 234, 393	0
3	C	3/8 (37%)	1.38	0 100 100	291, 291, 302, 304	0
All	All	1483/1540 (96%)	0.02	31 (2%) 67 58	60, 139, 217, 393	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	830	ALA	11.0
2	B	832	SER	8.9
2	B	831	ALA	7.5
2	B	1029	CYS	7.0
2	B	1031	LEU	5.4

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 carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
4	ZN	B	1101	1/1	0.89	0.20	-0.09	155,155,155,155	0
4	ZN	A	801	1/1	0.96	0.11	-1.37	111,111,111,111	0

6.5 Other polymers [i](#)

There are no such residues in this entry.