



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 5, 2024 – 12:08 PM EDT

PDB ID : 5KYY
Title : Crystal structure of Sec23 and TANGO1 peptide4 complex
Authors : Ma, W.; Goldberg, J.
Deposited on : 2016-07-22
Resolution : 3.40 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)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

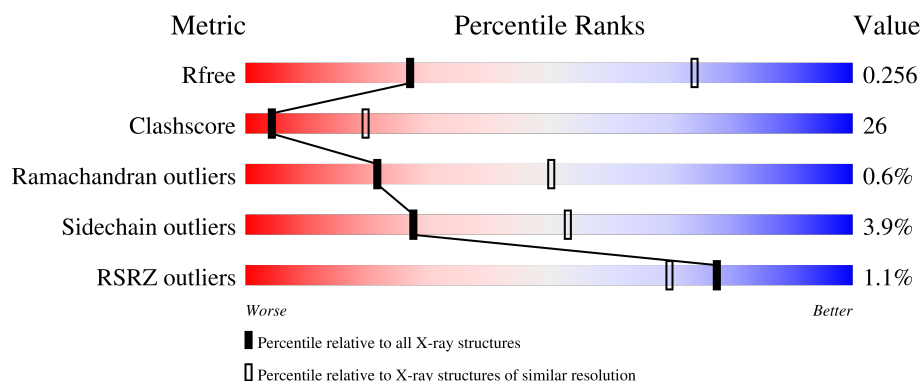
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.40 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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1140 (3.46-3.34)
Clashscore	180529	1172 (3.46-3.34)
Ramachandran outliers	177936	1172 (3.46-3.34)
Sidechain outliers	177891	1172 (3.46-3.34)
RSRZ outliers	164620	1140 (3.46-3.34)

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	A	765	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> % 58% 34% 6% </div> </div>
2	B	770	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> % 58% 39% . </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
3	ZN	A	801	-	-	X	-

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11682 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
			5676	3619	975	1042	40			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	767	Total	C	N	O	S	0	0	0
			6004	3822	1014	1114	54			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	ALA	-	expression tag	UNP O94855
B	2	MET	-	expression tag	UNP O94855
B	3	GLY	-	expression tag	UNP O94855

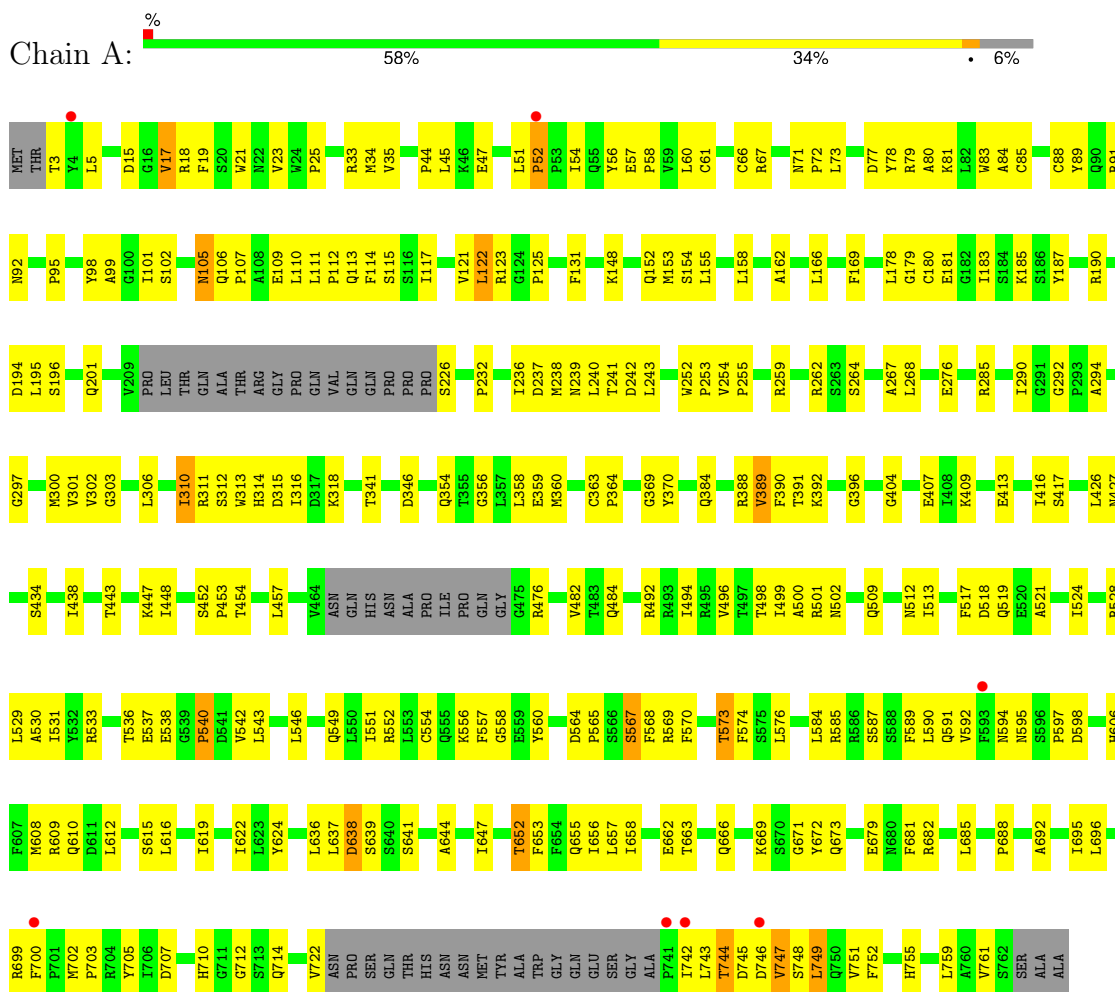
- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

3 Residue-property plots

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: Protein transport protein Sec23A



Q1030	T958	P860	R759	L650	G576	K472	Q376
L1031	L959	E861	R760	V651	K577	I473	F377
L1032	P961	Q870	N765	P652	L578	K474	I378
N1033	E962	R871	L766	P653	F579	K475	E379
	Y963	Q872	L768	T655	S583	Q478	R383
	G964	M875	N769	K661	S584	I484	Y384
	N965	T876	C770	T662	L585		Q385
	P966	Q889	S771	N663	T587		C386
	Y967			N664	A588		G387
	S968				E589		F388
	Q969				A590		C389
	Q970				F591		N390
					G592		C391
	M973				K593		V392
	I974				L594		N393
	N975				D599		Q401
	G976				K600		F497
	I977				G601		H402
	I978				L602		L403
	Q979				T605		I406
					D606		G407
	R982				K607		R408
	S985				E608		R409
					K609		L410
	L988				L610		D411
	K992				L611		H412
	Q993				F612		K415
					Q613		P416
	Q996				F614		E417
	P997				M617		L418
	E998				S621		S419
	N999				K624		A427
	V1000				D625		Y431
	F1001				G626		P441
	R1002				V627		A442
					A628		F443
					H629		M446
					S632		N453
	K1009				V633		L458
	G1010				L637		V459
LEU	S1015				F638		I462
TYR	S1016				P639		C463
GLY	Y1017				S640		E464
	D1018				Q641		E465
	F1020				Y642		L466
	L1021				V643		K467
	C1022				Q747		T468
	V1024				L751		M469
	H1025				Q758		L470
	K1026						E471
	E1027						
	I1028						
	C1029						

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	101.39Å 139.06Å 149.44Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.01 – 3.40 48.01 – 3.40	Depositor EDS
% Data completeness (in resolution range)	83.9 (48.01-3.40) 81.4 (48.01-3.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.67 (at 3.40Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.177 , 0.253 0.199 , 0.256	Depositor DCC
R_{free} test set	1985 reflections (7.96%)	wwPDB-VP
Wilson B-factor (Å ²)	82.7	Xtriage
Anisotropy	0.330	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 52.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11682	wwPDB-VP
Average B, all atoms (Å ²)	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.63% 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

5.1 Standard geometry

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	0.61	0/5809	0.73	0/7868
2	B	0.67	0/6132	0.76	0/8309
All	All	0.64	0/11941	0.75	0/16177

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

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	5676	0	5627	290	0
2	B	6004	0	5968	331	0
3	A	1	0	0	2	0
3	B	1	0	0	0	0
All	All	11682	0	11595	614	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 26.

The worst 5 of 614 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:51:LEU:HD21	1:A:117:ILE:HA	1.43	0.96
2:B:967:TYR:O	2:B:969:GLN:N	1.99	0.96
1:A:652:THR:HG23	1:A:655:GLN:H	1.29	0.94
2:B:366:ARG:NH2	2:B:391:CYS:HB2	1.83	0.93
2:B:931:TRP:CZ2	2:B:993:GLN:HG3	2.04	0.93

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%)	666 (94%)	39 (6%)	3 (0%)	30	60
2	B	763/770 (99%)	718 (94%)	39 (5%)	6 (1%)	16	44
All	All	1471/1535 (96%)	1384 (94%)	78 (5%)	9 (1%)	22	50

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	304	ASP
2	B	411	ASP
2	B	968	SER
2	B	998	GLU
1	A	389	VAL

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%)	601 (96%)	23 (4%)	29	54
2	B	672/682 (98%)	645 (96%)	27 (4%)	27	52
All	All	1296/1348 (96%)	1246 (96%)	50 (4%)	27	53

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	359	SER
2	B	741	ASP
2	B	1031	LEU
2	B	468	THR
2	B	617	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	617	ASN
2	B	668	HIS
2	B	664	ASN
2	B	816	HIS
1	A	655	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no oligosaccharides 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 [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	716/765 (93%)	-0.39	7 (0%) 79 71	26, 71, 129, 225	0
2	B	767/770 (99%)	-0.35	10 (1%) 74 66	31, 74, 160, 275	0
All	All	1483/1535 (96%)	-0.37	17 (1%) 77 70	26, 72, 149, 275	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	742	ILE	5.0
1	A	741	PRO	3.8
2	B	1015	SER	3.0
2	B	857	LEU	2.9
2	B	523	LEU	2.7

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](#)

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. 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	B-factors(\AA^2)	Q<0.9
3	ZN	B	1101	1/1	0.89	0.10	81,81,81,81	0
3	ZN	A	801	1/1	0.96	0.10	70,70,70,70	0

6.5 Other polymers [i](#)

There are no such residues in this entry.