wwPDB X-ray Structure Validation Summary Report (i)

Jan 26, 2022 – 04:01 PM EST

PDB ID	:	7MKK
Title	:	Crystal structure of Drosophila Panoramix in complex with Sov NTD
Authors	:	Wang, J.; Patel, D.J.
Deposited on	:	2021-04-24
Resolution	:	2.50 Å(reported)

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

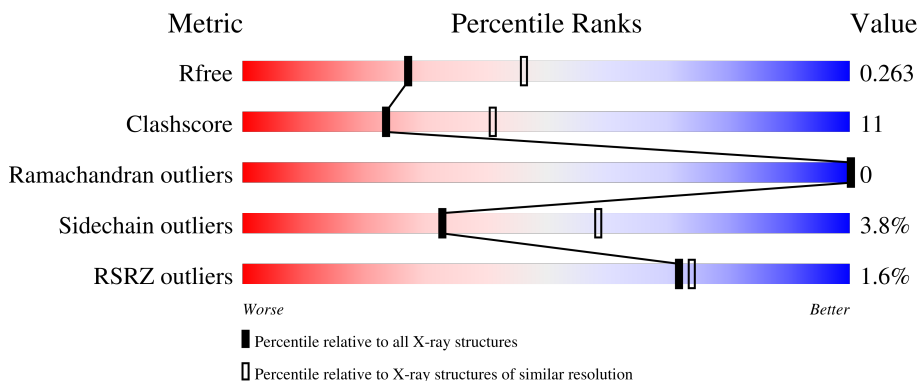|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.26
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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 2.50 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.
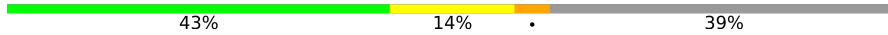
Metric	Percentile Ranks	Value
Rfree		0.263
Clashscore		11
Ramachandran outliers		0
Sidechain outliers		3.8%
RSRZ outliers		1.6%

Worse   *Better*

▊ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	78	3% · 63% · 27% · 9%
1	B	78	77% · 15% · 6%
1	E	78	62% · 28% · 9%
1	G	78	4% · 60% · 28% · 9%
2	C	28	4% · 71% · 7% · 18%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	28	57% 21% 21%
2	F	28	43% 14% • 39%
2	H	28	64% 14% 21%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3072 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 Small ovary, isoform A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	73	Total	C	N	O	S	0	0	0
			603	383	115	103	2			
1	A	71	Total	C	N	O	S	0	0	0
			581	370	110	99	2			
1	E	71	Total	C	N	O	S	0	0	0
			585	372	110	101	2			
1	G	71	Total	C	N	O	S	0	0	0
			585	372	110	101	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	13	MET	-	initiating methionine	UNP Q9W3W6
A	13	MET	-	initiating methionine	UNP Q9W3W6
E	13	MET	-	initiating methionine	UNP Q9W3W6
G	13	MET	-	initiating methionine	UNP Q9W3W6

- Molecule 2 is a protein called Protein panoramix.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	23	Total	C	N	O	S	0	0	0
			179	111	27	40	1			
2	D	22	Total	C	N	O	S	0	0	0
			171	107	25	38	1			
2	F	17	Total	C	N	O	S	0	0	0
			129	79	18	31	1			
2	H	22	Total	C	N	O	S	0	0	0
			171	107	25	38	1			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	82	SER	-	expression tag	UNP Q9W2H9
D	82	SER	-	expression tag	UNP Q9W2H9
F	82	SER	-	expression tag	UNP Q9W2H9
H	82	SER	-	expression tag	UNP Q9W2H9

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	30	Total 30	O 30	0	0
3	C	3	Total 3	O 3	0	0
3	A	10	Total 10	O 10	0	0
3	D	6	Total 6	O 6	0	0
3	E	13	Total 13	O 13	0	0
3	F	3	Total 3	O 3	0	0
3	H	3	Total 3	O 3	0	0

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.
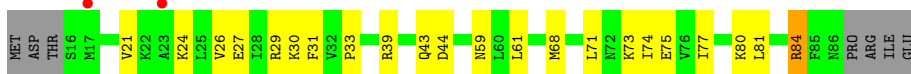
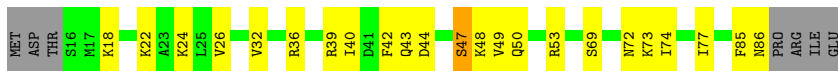- Molecule 1: Small ovary, isoform A

Chain B:
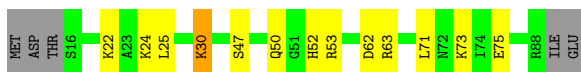
77% 15% 6%

MET ASP THR S16 K22 A23 K24 L25 K30 S47 Q50 G51 H52 R53 D62 R63 L71 N72 K73 I74 E75 R88 ILE GLU

- Molecule 1: Small ovary, isoform A

Chain A:

3% 63% 27% 9%

MET ASP THR S16 M17 V21 K22 A23 K24 L25 V26 E27 I28 R29 K30 F31 V32 P33 R39 Q43 D44 N59 L60 L61 M68 L71 N72 K73 I74 E75 V76 I77 K80 L81 R84 F85 N86 PRO ARG ILE GLU

- Molecule 1: Small ovary, isoform A

Chain E:

62% 28% 9%

MET ASP THR S16 M17 K18 K22 A23 K24 L25 V26 V32 R36 R39 I40 D41 F42 Q43 D44 S47 K48 V49 Q50 R53 S69 N72 K73 I74 I77 F85 N86 PRO ARG ILE GLU

- Molecule 1: Small ovary, isoform A

Chain G:

4% 60% 28% 9%

MET ASP THR S16 M17 K18 E19 K20 V21 K22 A23 K24 L25 V26 E27 I28 F34 V38 R39 I40 D41 V49 R53 L60 D65 V66 S67 L71 N72 K73 R84 F85 N86 PRO ARG ILE GLU

- Molecule 2: Protein panoramix

Chain C:

4% 71% 7% 18%

SER MET E84 D97 P103 F104 E105 N106 ASN SER PRO
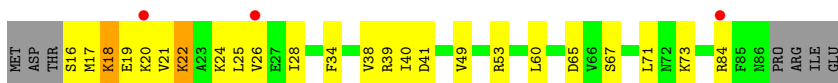
- Molecule 2: Protein panoramix

Chain D:

• Molecule 2: Protein panoramix

Chain F:

• Molecule 2: Protein panoramix

Chain H:
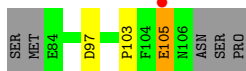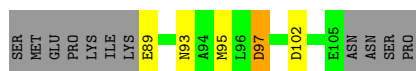
4   Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 64 2 2	Depositor
Cell constants a, b, c, α, β, γ	127.39Å 127.39Å 96.01Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.83 – 2.50 47.83 – 2.50	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.83-2.50) 100.0 (47.83-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	4.85 (at 2.51Å)	Xtriage
Refinement program	PHENIX 1.16_3549	Depositor
R, R_{free}	0.227 , 0.260 0.228 , 0.263	Depositor DCC
R_{free} test set	822 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	57.8	Xtriage
Anisotropy	0.088	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 51.8	EDS
L-test for twinning[2]	$<|L|> = 0.49, <L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3072	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.27% 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	A	0.34	0/586	0.46	0/782
1	B	0.28	0/609	0.50	0/813
1	E	0.28	0/590	0.45	0/787
1	G	0.52	0/590	0.61	0/787
2	C	0.43	0/181	0.49	0/244
2	D	0.30	0/173	0.50	0/233
2	F	0.29	0/130	0.49	0/176
2	H	0.27	0/173	0.44	0/233
All	All	0.36	0/3032	0.50	0/4055

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	581	0	627	20	0
1	B	603	0	651	12	0
1	E	585	0	631	15	0
1	G	585	0	631	20	0
2	C	179	0	167	2	0
2	D	171	0	161	7	0
2	F	129	0	111	4	0
2	H	171	0	161	4	0

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	10	0	0	1	0
3	B	30	0	0	4	0
3	C	3	0	0	0	0
3	D	6	0	0	0	0
3	E	13	0	0	0	0
3	F	3	0	0	0	0
3	H	3	0	0	0	0
All	All	3072	0	3140	69	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 69 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:E:85:PHE:O	1:E:86:ASN:ND2	2.04	0.90
1:B:53:ARG:HH22	2:D:88:LYS:HD3	1.42	0.83
1:A:73:LYS:HG2	2:D:100:LEU:HD11	1.72	0.72
1:G:18:LYS:O	1:G:21:VAL:CG1	2.38	0.71
1:B:30:LYS:O	3:B:101:HOH:O	2.10	0.69

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	69/78 (88%)	67 (97%)	2 (3%)	0	100	100
1	B	71/78 (91%)	70 (99%)	1 (1%)	0	100	100
1	E	69/78 (88%)	65 (94%)	4 (6%)	0	100	100
1	G	69/78 (88%)	62 (90%)	7 (10%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	21/28 (75%)	20 (95%)	1 (5%)	0	100	100
2	D	20/28 (71%)	19 (95%)	1 (5%)	0	100	100
2	F	15/28 (54%)	14 (93%)	1 (7%)	0	100	100
2	H	20/28 (71%)	19 (95%)	1 (5%)	0	100	100
All	All	354/424 (84%)	336 (95%)	18 (5%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	66/75 (88%)	63 (96%)	3 (4%)	27	51
1	B	69/75 (92%)	67 (97%)	2 (3%)	42	69
1	E	67/75 (89%)	65 (97%)	2 (3%)	41	68
1	G	67/75 (89%)	63 (94%)	4 (6%)	19	37
2	C	20/25 (80%)	19 (95%)	1 (5%)	24	46
2	D	19/25 (76%)	19 (100%)	0	100	100
2	F	14/25 (56%)	13 (93%)	1 (7%)	14	28
2	H	19/25 (76%)	19 (100%)	0	100	100
All	All	341/400 (85%)	328 (96%)	13 (4%)	33	58

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	47	SER
2	F	97	ASP
1	G	60	LEU
1	G	22	LYS
1	G	39	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	43	GLN
2	F	93	ASN

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 monosaccharides in this entry.

5.6   Ligand geometry ⓘ

There are no ligands in this entry.

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 ⓘ

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, 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	71/78 (91%)	-0.04	2 (2%)	53	56	44, 67, 98, 117	0
1	B	73/78 (93%)	-0.09	0	100	100	38, 49, 77, 96	0
1	E	71/78 (91%)	-0.26	0	100	100	42, 62, 87, 96	0
1	G	71/78 (91%)	0.21	3 (4%)	36	39	66, 81, 104, 113	0
2	C	23/28 (82%)	-0.03	1 (4%)	35	38	39, 58, 89, 95	0
2	D	22/28 (78%)	-0.28	0	100	100	48, 62, 83, 99	1 (4%)
2	F	17/28 (60%)	-0.17	0	100	100	49, 64, 92, 108	0
2	H	22/28 (78%)	-0.03	0	100	100	70, 80, 91, 114	0
All	All	370/424 (87%)	-0.06	6 (1%)	72	74	38, 66, 96, 117	1 (0%)

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	23	ALA	3.4
1	G	20	LYS	3.0
1	A	17	MET	2.9
2	C	105	GLU	2.9
1	G	26	VAL	2.3

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.