



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:58 PM GMT

PDB ID : 4K2P
Title : The Structure of a Quintuple Mutant of the Tiam1 PH-CC-Ex Domain
Authors : Joshi, M.; Gakhar, L.; Fuentes, E.J.
Deposited on : 2013-04-09
Resolution : 1.98 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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

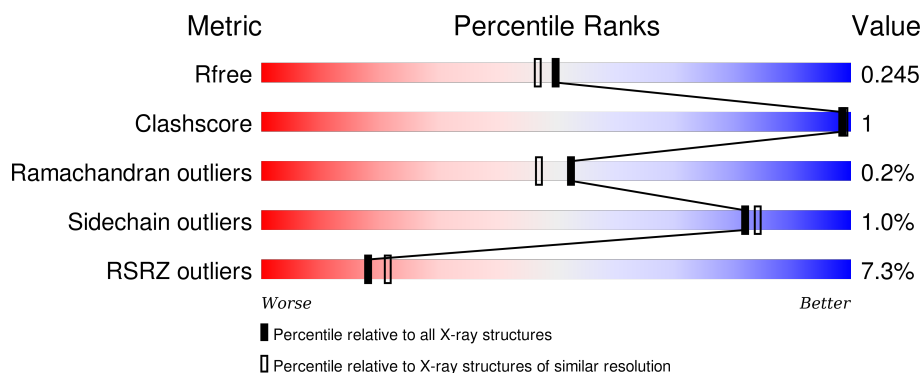
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.98 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	8664 (2.00-1.96)
Clashscore	102246	9905 (2.00-1.96)
Ramachandran outliers	100387	9792 (2.00-1.96)
Sidechain outliers	100360	9791 (2.00-1.96)
RSRZ outliers	91569	8679 (2.00-1.96)

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	276	<div> <div>5%</div> <div>85%</div> <div>14%</div> </div>
1	B	276	<div> <div>7%</div> <div>80%</div> <div>6%</div> <div>14%</div> </div>
1	C	276	<div> <div>7%</div> <div>81%</div> <div>16%</div> </div>
1	D	276	<div> <div>6%</div> <div>83%</div> <div>16%</div> </div>

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 15771 atoms, of which 7540 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 T-lymphoma invasion and metastasis-inducing protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	238	Total	C	H	N	O	S	0	2	0
			3815	1213	1919	336	339	8			
1	B	238	Total	C	H	N	O	S	0	0	0
			3802	1210	1911	335	338	8			
1	C	233	Total	C	H	N	O	S	0	0	0
			3676	1179	1837	323	329	8			
1	D	232	Total	C	H	N	O	S	0	2	0
			3724	1188	1873	325	330	8			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	427	GLY	-	EXPRESSION TAG	UNP Q13009
A	428	SER	-	EXPRESSION TAG	UNP Q13009
A	580	LEU	MET	ENGINEERED MUTATION	UNP Q13009
A	586	LEU	MET	ENGINEERED MUTATION	UNP Q13009
A	596	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
A	597	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
A	598	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
B	427	GLY	-	EXPRESSION TAG	UNP Q13009
B	428	SER	-	EXPRESSION TAG	UNP Q13009
B	580	LEU	MET	ENGINEERED MUTATION	UNP Q13009
B	586	LEU	MET	ENGINEERED MUTATION	UNP Q13009
B	596	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
B	597	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
B	598	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
C	427	GLY	-	EXPRESSION TAG	UNP Q13009
C	428	SER	-	EXPRESSION TAG	UNP Q13009
C	580	LEU	MET	ENGINEERED MUTATION	UNP Q13009
C	586	LEU	MET	ENGINEERED MUTATION	UNP Q13009
C	596	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
C	597	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
C	598	ALA	LYS	ENGINEERED MUTATION	UNP Q13009

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	427	GLY	-	EXPRESSION TAG	UNP Q13009
D	428	SER	-	EXPRESSION TAG	UNP Q13009
D	580	LEU	MET	ENGINEERED MUTATION	UNP Q13009
D	586	LEU	MET	ENGINEERED MUTATION	UNP Q13009
D	596	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
D	597	ALA	LYS	ENGINEERED MUTATION	UNP Q13009
D	598	ALA	LYS	ENGINEERED MUTATION	UNP Q13009

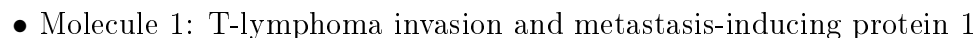
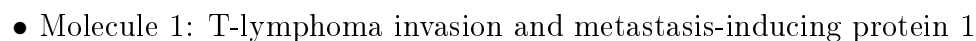
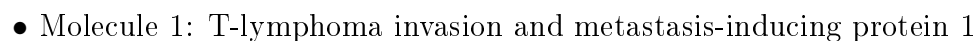
- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Ca 1 1	0	0
2	A	2	Total Ca 2 2	0	0
2	D	1	Total Ca 1 1	0	0
2	C	1	Total Ca 1 1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	192	Total O 192 192	0	0
3	B	195	Total O 195 195	0	0
3	C	201	Total O 201 201	0	0
3	D	161	Total O 161 161	0	0

- Molecule 1: T-lymphoma invasion and metastasis-inducing protein 1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	80.51Å 70.11Å 87.75Å 90.00° 90.19° 90.00°	Depositor
Resolution (Å)	19.70 – 1.98 19.70 – 1.98	Depositor EDS
% Data completeness (in resolution range)	97.5 (19.70-1.98) 97.7 (19.70-1.98)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.34 (at 1.97Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.192 , 0.236 0.206 , 0.245	Depositor DCC
R_{free} test set	3358 reflections (5.31%)	DCC
Wilson B-factor (Å ²)	23.6	Xtriage
Anisotropy	0.761	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.40 , 49.3	EDS
Estimated twinning fraction	0.062 for h,-k,-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	2 of 66586 reflections (0.003%)	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	15771	wwPDB-VP
Average B, all atoms (Å ²)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 44.83 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.4433e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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: CA

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	0.32	0/1947	0.47	0/2625
1	B	0.32	0/1933	0.50	0/2607
1	C	0.33	0/1877	0.48	0/2531
1	D	0.31	0/1899	0.47	0/2555
All	All	0.32	0/7656	0.48	0/10318

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	1896	1919	1895	1	0
1	B	1891	1911	1906	8	0
1	C	1839	1837	1838	7	0
1	D	1851	1873	1849	1	0
2	A	2	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	192	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	195	0	0	1	0
3	C	201	0	0	0	0
3	D	161	0	0	0	0
All	All	8231	7540	7488	14	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 1.

The worst 5 of 14 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:B:551:ALA:HA	1:B:562:LEU:CD2	2.39	0.52
1:B:449:LYS:HD2	1:C:653:ARG:HD3	1.91	0.52
1:C:482:ARG:O	1:C:483:SER:HB3	2.09	0.51
1:B:449:LYS:CD	1:C:653:ARG:HD3	2.45	0.46
1:C:645:ARG:HB3	1:C:646:PRO:HD3	1.98	0.45

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	238/276 (86%)	232 (98%)	6 (2%)	0	100	100
1	B	236/276 (86%)	228 (97%)	7 (3%)	1 (0%)	39	31
1	C	225/276 (82%)	220 (98%)	4 (2%)	1 (0%)	39	31
1	D	227/276 (82%)	223 (98%)	4 (2%)	0	100	100
All	All	926/1104 (84%)	903 (98%)	21 (2%)	2 (0%)	52	47

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	483	SER
1	B	459	ARG

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	204/235 (87%)	203 (100%)	1 (0%)	92	93
1	B	203/235 (86%)	199 (98%)	4 (2%)	63	63
1	C	196/235 (83%)	195 (100%)	1 (0%)	92	93
1	D	199/235 (85%)	197 (99%)	2 (1%)	82	84
All	All	802/940 (85%)	794 (99%)	8 (1%)	82	84

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	487	HIS
1	D	472	CYS
1	C	472	CYS
1	B	472	CYS
1	B	640	LEU

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

Mol	Chain	Res	Type
1	C	444	ASN
1	C	493	HIS

5.3.3 RNA ⓘ

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 [i](#)

Of 5 ligands modelled in this entry, 5 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	238/276 (86%)	0.41	14 (5%) 26 30	20, 33, 58, 68	0
1	B	238/276 (86%)	0.52	20 (8%) 14 16	22, 34, 70, 92	0
1	C	233/276 (84%)	0.40	18 (7%) 16 19	19, 31, 64, 85	0
1	D	232/276 (84%)	0.56	17 (7%) 18 21	22, 39, 65, 83	0
All	All	941/1104 (85%)	0.47	69 (7%) 18 21	19, 34, 64, 92	0

The worst 5 of 69 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	451	ASN	5.9
1	D	487	HIS	4.9
1	B	487	HIS	4.6
1	A	592	THR	4.4
1	B	458	THR	4.3

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
2	CA	A	1002	1/1	0.98	0.07	-2.64	46,46,46,46	0
2	CA	C	1001	1/1	0.98	0.18	-	45,45,45,45	0
2	CA	B	1001	1/1	0.99	0.18	-	45,45,45,45	0
2	CA	D	1001	1/1	0.98	0.17	-	45,45,45,45	0
2	CA	A	1001	1/1	0.94	0.17	-	53,53,53,53	0

6.5 Other polymers [i](#)

There are no such residues in this entry.