



wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 10:53 AM GMT

PDB ID : 4GE5
Title : Crystal structure of human protein tyrosine phosphatase PTPN9 (MEG2)
complex with compound 5
Authors : Zhang, Z.-Y.; Liu, S.; Zhang, S.
Deposited on : 2012-08-01
Resolution : 2.00 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

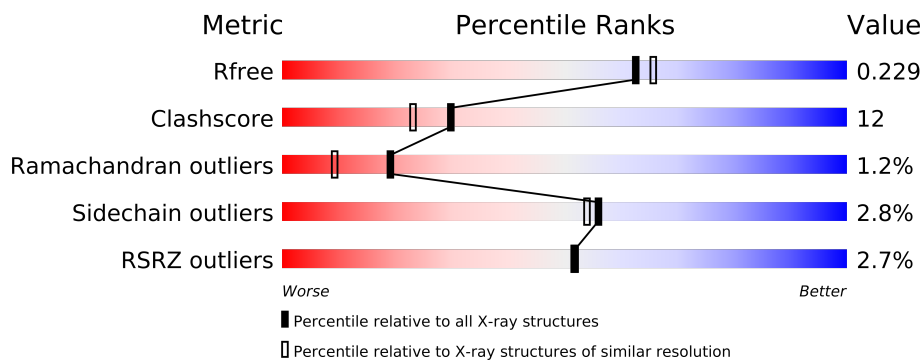
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.00 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}	66092	4888 (2.00-2.00)
Clashscore	79885	6188 (2.00-2.00)
Ramachandran outliers	78287	6102 (2.00-2.00)
Sidechain outliers	78261	6100 (2.00-2.00)
RSRZ outliers	66119	4890 (2.00-2.00)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	314	
1	B	314	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5230 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

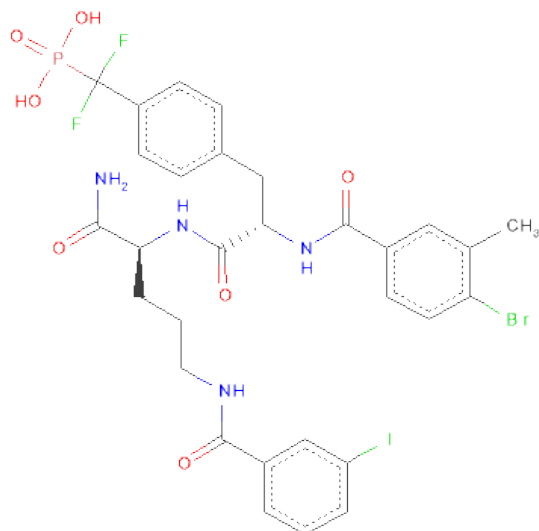
- Molecule 1 is a protein called Tyrosine-protein phosphatase non-receptor type 9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	298	Total	C	N	O	S	6	0	0
			2405	1518	424	447	16			
1	B	299	Total	C	N	O	S	0	0	0
			2423	1528	428	451	16			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	276	MET	-	INITIATING METHIONINE	UNP P43378
A	583	ALA	-	EXPRESSION TAG	UNP P43378
A	584	HIS	-	EXPRESSION TAG	UNP P43378
A	585	HIS	-	EXPRESSION TAG	UNP P43378
A	586	HIS	-	EXPRESSION TAG	UNP P43378
A	587	HIS	-	EXPRESSION TAG	UNP P43378
A	588	HIS	-	EXPRESSION TAG	UNP P43378
A	589	HIS	-	EXPRESSION TAG	UNP P43378
B	276	MET	-	INITIATING METHIONINE	UNP P43378
B	583	ALA	-	EXPRESSION TAG	UNP P43378
B	584	HIS	-	EXPRESSION TAG	UNP P43378
B	585	HIS	-	EXPRESSION TAG	UNP P43378
B	586	HIS	-	EXPRESSION TAG	UNP P43378
B	587	HIS	-	EXPRESSION TAG	UNP P43378
B	588	HIS	-	EXPRESSION TAG	UNP P43378
B	589	HIS	-	EXPRESSION TAG	UNP P43378

- Molecule 2 is N-(4-BROMO-3-METHYLBENZOYL)-4-[DIFLUORO(PHOSPHONO)METHYL]-L-PHENYLALANYL-N 5 -(3-IODOBENZOYL)-L-ORNITHINAMIDE (three-letter code: A89) (formula: C₃₀H₃₁BrF₂IN₄O₇P).



Mol	Chain	Residues	Atoms								ZeroOcc	AltConf
			Total	Br	C	F	I	N	O	P		
2	A	1	46	1	30	2	1	4	7	1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	187	Total	O	0	0
			187	187		
3	B	169	Total	O	0	0
			169	169		

4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	40.26Å 57.17Å 66.42Å 77.55° 78.23° 79.97°	Depositor
Resolution (Å)	50.00 – 2.00 30.03 – 1.76	Depositor EDS
% Data completeness (in resolution range)	91.3 (50.00-2.00) 89.3 (30.03-1.76)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.54 (at 1.76Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.187 , 0.223 0.191 , 0.229	Depositor DCC
R_{free} test set	1405 reflections (4.08%)	DCC
Wilson B-factor (Å ²)	17.7	Xtriage
Anisotropy	0.627	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 49.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 52574 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	5230	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.40% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: A89

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	0.33	0/2459	0.58	0/3325
1	B	0.31	0/2478	0.57	0/3350
All	All	0.32	0/4937	0.57	0/6675

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2405	0	2346	52	0
1	B	2423	0	2363	64	0
2	A	46	0	29	1	0
3	A	187	0	0	1	0
3	B	169	0	0	1	0
All	All	5230	0	4738	116	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 12.

The worst 5 of 116 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:575:ALA:HA	1:A:580:MET:HE3	1.57	0.84
1:B:581:VAL:HG12	1:B:582:SER:H	1.45	0.81
1:A:312:ARG:NE	1:A:312:ARG:HA	1.99	0.76
1:B:292:VAL:HG21	1:B:581:VAL:HG11	1.68	0.75
1:B:319:PHE:O	1:B:323:MET:HG2	1.86	0.75

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	294/314 (94%)	279 (95%)	11 (4%)	4 (1%)	16	7
1	B	295/314 (94%)	284 (96%)	8 (3%)	3 (1%)	22	12
All	All	589/628 (94%)	563 (96%)	19 (3%)	7 (1%)	19	9

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	350	SER
1	A	351	GLY
1	B	421	ASP
1	B	496	SER
1	A	558	ILE

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	263/278 (95%)	259 (98%)	4 (2%)	76	79

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	266/278 (96%)	255 (96%)	11 (4%)	41	35
All	All	529/556 (95%)	514 (97%)	15 (3%)	56	54

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	414	GLN
1	B	439	MET
1	B	528	LEU
1	B	409	ARG
1	B	488	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	489	ASN
1	A	559	GLN
1	B	414	GLN
1	A	456	GLN
1	B	289	GLN

5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and

the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	A89	A	601	-	48,48,48	1.70	5 (10%)	69,69,69	1.62	14 (20%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	A89	A	601	-	-	0/46/50/50	0/3/3/3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	A89	C24-C23	-6.68	1.36	1.50
2	A	601	A89	C17-C16	-5.87	1.37	1.50
2	A	601	A89	P1-O1	4.02	1.62	1.54
2	A	601	A89	P1-O3	3.21	1.61	1.54
2	A	601	A89	C21-I	2.18	2.16	2.10

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	601	A89	C17-C16-N4	4.80	128.52	117.17
2	A	601	A89	O6-C16-N4	-4.21	113.38	122.50
2	A	601	A89	C24-C23-N1	4.13	123.77	116.89
2	A	601	A89	C9-N1-C23	3.96	130.63	121.80
2	A	601	A89	O7-C23-N1	-3.78	115.61	122.44

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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	298/314 (94%)	-0.26	5 (1%) 67 67	12, 23, 41, 55	1 (0%)
1	B	299/314 (95%)	-0.01	11 (3%) 39 39	16, 28, 46, 56	0
All	All	597/628 (95%)	-0.14	16 (2%) 52 52	12, 25, 44, 56	1 (0%)

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	583	ALA	6.5
1	A	349	ARG	5.6
1	B	582	SER	5.4
1	B	283	PRO	5.0
1	B	584	HIS	4.7

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 carbohydrates in this entry.

6.4 Ligands ⓘ

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	RSR	LLDF	B-factors(\AA^2)	Q<0.9
2	A89	A	601	46/46	0.17	0.44	22,43,62,75	0

6.5 Other polymers ⓘ

There are no such residues in this entry.