



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 03:18 pm BST

PDB ID : 2PI7
Title : Structure of the catalytic domain of the chick retinal neurite inhibitor-Receptor Protein Tyrosine Phosphatase CRYP-2/cPTPRO
Authors : Girish, T.S.; Gopal, B.
Deposited on : 2007-04-13
Resolution : 2.59 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
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) : 2.11

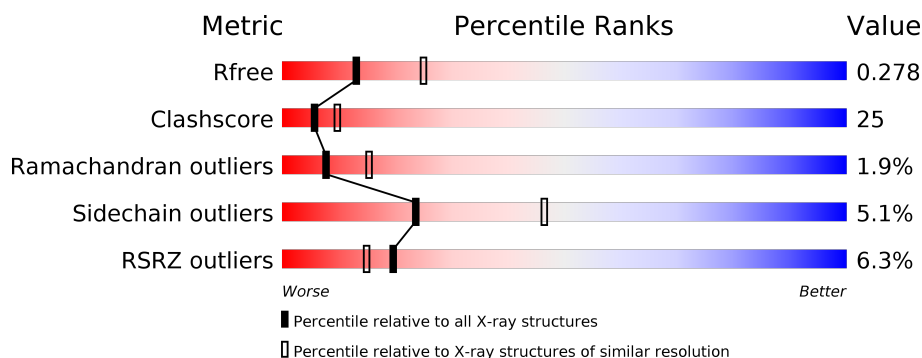
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.59 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}	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 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	312	<div> <div>5%</div> <div> <div></div> <div>58%</div> <div>33%</div> <div>• 6%</div> </div> </div>
1	B	312	<div> <div>6%</div> <div> <div></div> <div>52%</div> <div>36%</div> <div>• 7%</div> </div> </div>

2 Entry composition [i](#)

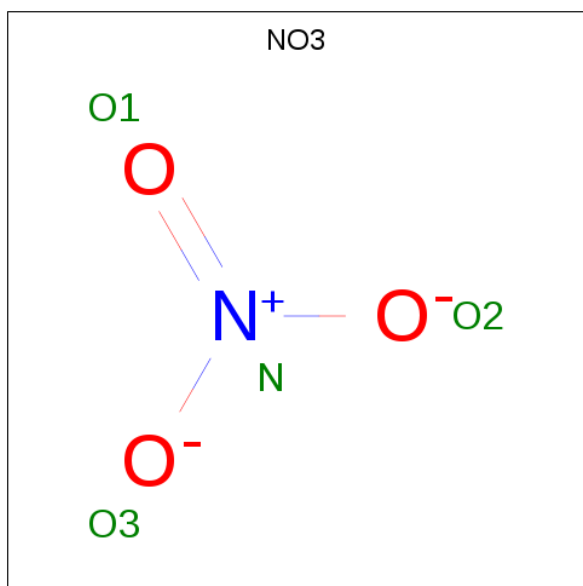
There are 3 unique types of molecules in this entry. The entry contains 4920 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 tyrosine phosphatase CRYP-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	294	Total	C	N	O	S	0	0	0
			2400	1530	408	444	18			
1	B	289	Total	C	N	O	S	0	0	0
			2330	1482	394	436	18			

- Molecule 2 is NITRATE ION (three-letter code: NO3) (formula: NO₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	N	O	0	0
			4	1	3		
2	B	1	Total	N	O	0	0
			4	1	3		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	110	Total 110	O 110	0	0
3	B	72	Total 72	O 72	0	0

These plots are drawn for all protein, RNA and DNA 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 ($\text{RSRZ} > 2$). 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.

Chain A:

Residue	State
Q1248	Red
K1249	Green
K1250	Green
K1251	Red
Q1252	Green
Q1253	Green
PHE	Green
CYS	Green
ILE	Green
SER	Green
ASP	Green
VAL	Green
ILE	Green
I1185	Red
H1186	Green
TYR	Green
GLU	Green
ASN	Green
VAL	Green
SER	Green
LYS	Green
SER	Green
T1158	Red
A1159	Red
E1163	Green
Q1167	Red
S1179	Green
P1182	Green
M1183	Green
I1184	Green
I1185	Red
H1186	Green
C1187	Green
G1190	Green
V1191	Green
G1192	Green
T1194	Green
I1198	Green
D1201	Green
W1202	Red
Q1205	Green
H1206	Green
F1212	Green
I1215	Green
L1216	Red
G1217	Green
L1218	Green
V1219	Green
M1222	Green
R1223	Green
S1224	Green
Y1225	Green
R1226	Green
M1229	Green
V1230	Red
Q1231	Green
T1232	Green
Q1235	Green
Y1236	Green
H1240	Green
Q1241	Green
C1242	Green
V1243	Green
Q1244	Green
L1245	Red
M1246	Green
H1247	Red
SER	Green
LYS	Green
ASN	Green
GLY	Red
I960	Red
I961	Red
K962	Red
R963	Green
R964	Green
L965	Green
T966	Green
I967	Green
Q970	Green
L971	Green
F974	Green
Y977	Green
K979	Green
D980	Green
N981	Green
A982	Green
K983	Green
F993	Green
L996	Green
K997	Green
H1005	Green
F1006	Green
A1007	Green
H1017	Green
R1018	Green
M1021	Red
I1022	Red
L1023	Green
F1027	Green
S1028	Green
R1031	Green
M1035	Red
H1036	Green
E1037	Red
N1045	Red
A1046	Green
P1050	Green
G1051	Green
Y1052	Green
N1053	Green
S1054	Green
P1055	Green

Chain B:

6% 52% 36% 7%

4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	68.32Å 68.32Å 245.02Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.56 – 2.59 19.91 – 2.61	Depositor EDS
% Data completeness (in resolution range)	97.3 (42.56-2.59) 99.4 (19.91-2.61)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.77 (at 2.59Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.227 , 0.287 0.217 , 0.278	Depositor DCC
R_{free} test set	1063 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å ²)	27.0	Xtriage
Anisotropy	0.138	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 50.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.067 for -h,-k,l	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	4920	wwPDB-VP
Average B, all atoms (Å ²)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.67% 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: NO3

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	0.38	0/2461	0.64	0/3334
1	B	0.37	0/2388	0.65	0/3238
All	All	0.38	0/4849	0.65	0/6572

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	2400	0	2307	116	0
1	B	2330	0	2209	119	0
2	A	4	0	0	0	0
2	B	4	0	0	0	0
3	A	110	0	0	7	0
3	B	72	0	0	9	0
All	All	4920	0	4516	235	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 25.

The worst 5 of 235 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:1051:GLY:HA2	1:A:1202:TRP:CZ2	1.77	1.17
1:A:1051:GLY:HA2	1:A:1202:TRP:CH2	1.91	1.05
1:B:1046:ALA:HB2	1:B:1061:THR:HG23	1.38	1.05
1:A:1046:ALA:HB2	1:A:1061:THR:HG23	1.41	1.02
1:A:1202:TRP:CD1	1:A:1218:LEU:HD11	2.00	0.97

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	292/312 (94%)	261 (89%)	28 (10%)	3 (1%)	15	32
1	B	287/312 (92%)	244 (85%)	35 (12%)	8 (3%)	5	7
All	All	579/624 (93%)	505 (87%)	63 (11%)	11 (2%)	8	15

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	1156	VAL
1	B	1157	PRO
1	B	1158	THR
1	A	1125	THR
1	B	1154	HIS

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	261/286 (91%)	246 (94%)	15 (6%)	20	41
1	B	251/286 (88%)	240 (96%)	11 (4%)	28	53
All	All	512/572 (90%)	486 (95%)	26 (5%)	24	46

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1202	TRP
1	A	1248	GLN
1	B	1216	LEU
1	A	1216	LEU
1	A	1247	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1240	HIS
1	B	967	ASN
1	B	1240	HIS
1	A	1241	GLN
1	A	1248	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

2 ligands are 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	NO3	B	2002	-	1,3,3	0.30	0	0,3,3	0.00	-
2	NO3	A	2001	-	1,3,3	0.21	0	0,3,3	0.00	-

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	294/312 (94%)	0.24	17 (5%) 23 17	9, 20, 39, 56	0
1	B	289/312 (92%)	0.46	20 (6%) 16 12	8, 23, 44, 62	0
All	All	583/624 (93%)	0.35	37 (6%) 20 15	8, 21, 42, 62	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	960	LEU	9.5
1	B	1156	VAL	8.3
1	B	1155	GLY	8.3
1	B	1035	MET	6.1
1	B	1157	PRO	5.9

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. 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
2	NO3	B	2002	4/4	0.98	0.12	21,22,23,25	0
2	NO3	A	2001	4/4	0.98	0.13	12,16,17,18	0

6.5 Other polymers [i](#)

There are no such residues in this entry.