wwPDB X-ray Structure Validation Summary Report (i)

Oct 31, 2023 – 11:32 AM JST

PDB ID	:	5CRU
Title	:	Crystal structure of the Bro domain of HD-PTP
Authors	:	Lee, J.; Ku, B.; Kim, S.J.
Deposited on	:	2015-07-23
Resolution	:	2.40 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 (i) 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 (1)) were used in the production of this report:

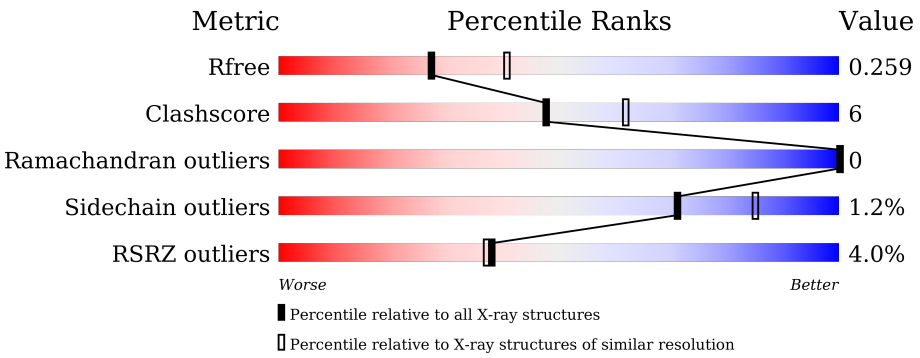
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.40 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}	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	361	3% 86% 14%
1	B	361	4% 87% 11%
1	C	361	3% 85% 14%
1	D	361	6% 83% 16%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 11411 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 Tyrosine-protein phosphatase non-receptor type 23.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	360	2853	1823	489	521	20	0	1	0
1	B	357	2833	1811	485	518	19	0	1	0
1	C	360	2853	1823	489	521	20	0	1	0
1	D	357	2828	1806	485	518	19	0	0	0

- Molecule 2 is water.

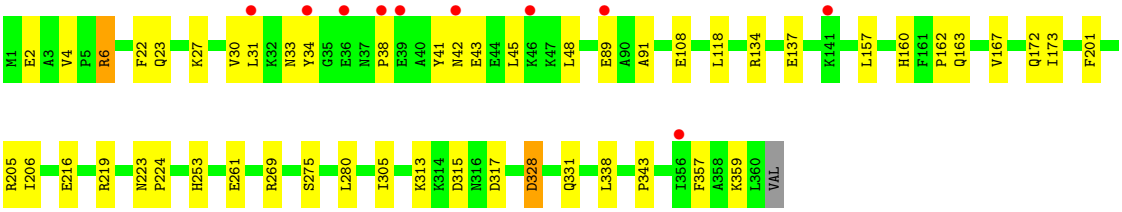
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	10	10	10	0	0
2	B	11	11	11	0	0
2	C	13	13	13	0	0
2	D	10	10	10	0	0

3 Residue-property plots (i)

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.

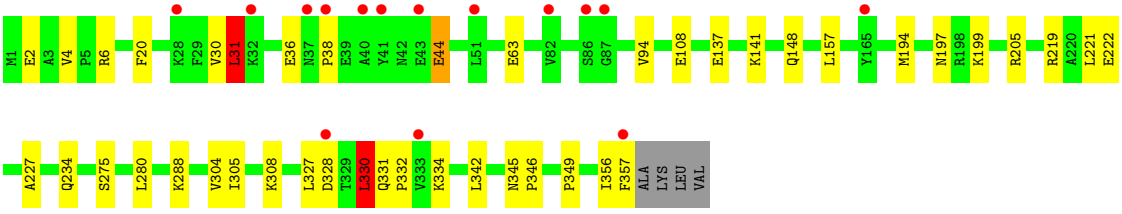
• Molecule 1: Tyrosine-protein phosphatase non-receptor type 23

Chain A:



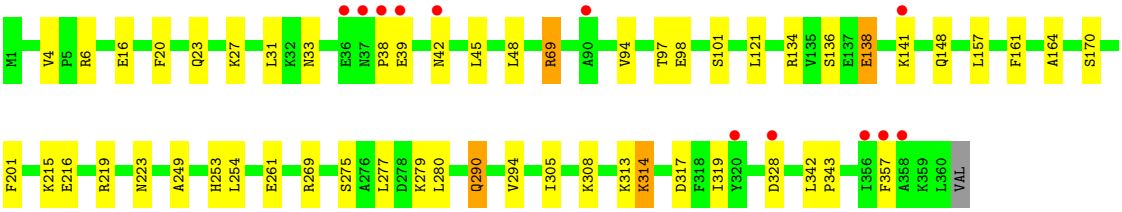
• Molecule 1: Tyrosine-protein phosphatase non-receptor type 23

Chain B:



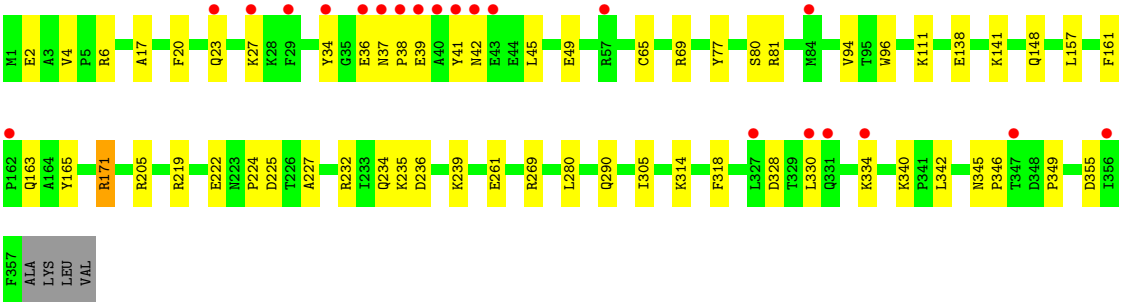
• Molecule 1: Tyrosine-protein phosphatase non-receptor type 23

Chain C:



• Molecule 1: Tyrosine-protein phosphatase non-receptor type 23

Chain D:


M1
E2
A3
V4
P5
R6
A17
F20
Q23
K27
K28
F29
Y34
G35
E36
N37
P38
E39
A40
Y41
N42
E43
E44
L45
E49
R57
C65
R69
Y77
S80
R81
M84
V94
T95
W96
K111
E138
K141
Q148
L157
F161
P162
Q163
A164
Y165
R171
R205
R219
E222
N223
P224
D225
T226
A227
R232
I233
Q234
K235
D236
K239
E261
R269
L280
Q290
I305
K314
F318
L327
D328
T329
L330
Q331
K334
K340
P341
L342
N345
P346
T347
D348
P349
D355
I356
F357
ALA
LYS
LEU
VAL

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	64.99Å 66.87Å 82.42Å 89.95° 89.94° 89.84°	Depositor
Resolution (Å)	30.98 – 2.40 30.99 – 2.36	Depositor EDS
% Data completeness (in resolution range)	94.0 (30.98-2.40) 91.4 (30.99-2.36)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.70 (at 2.36Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.220 , 0.259 0.222 , 0.259	Depositor DCC
R_{free} test set	2056 reflections (3.93%)	wwPDB-VP
Wilson B-factor (Å^2)	38.7	Xtriage
Anisotropy	0.236	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 20.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.000 for k,-h,l 0.000 for -k,h,l 0.458 for h,-k,-l 0.027 for -h,k,-l 0.026 for -h,-k,l 0.000 for k,h,-l 0.000 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11411	wwPDB-VP
Average B, all atoms (Å^2)	48.0	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 23.90 % 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 4.2583e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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 (i)

5.1 Standard geometry (i)

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	0.26	0/2920	0.47	0/3943
1	B	0.30	0/2898	0.54	2/3915 (0.1%)
1	C	0.30	0/2920	0.53	1/3943 (0.0%)
1	D	0.31	0/2890	0.54	0/3904
All	All	0.29	0/11628	0.52	3/15705 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	290	GLN	CA-CB-CG	7.65	130.23	113.40
1	B	31	LEU	CA-CB-CG	5.62	128.22	115.30
1	B	330	LEU	CB-CG-CD2	5.17	119.80	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	330	LEU	Peptide

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	2853	0	2869	34	1
1	B	2833	0	2846	36	0
1	C	2853	0	2869	35	1
1	D	2828	0	2835	49	0
2	A	10	0	0	1	0
2	B	11	0	0	1	0
2	C	13	0	0	1	0
2	D	10	0	0	0	0
All	All	11411	0	11419	138	1

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 6.

The worst 5 of 138 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:199:LYS:HD2	1:B:199:LYS:H	1.28	0.96
1:A:27:LYS:NZ	1:A:42:ASN:OD1	2.02	0.92
1:D:219:ARG:NH1	1:D:222:GLU:OE2	2.06	0.88
1:C:27:LYS:NZ	1:C:42:ASN:OD1	2.11	0.83
1:C:33:ASN:OD1	1:D:205:ARG:NH1	2.14	0.80

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:137:GLU:OE2	1:C:215:LYS:NZ[1_564]	2.10	0.10

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	359/361 (99%)	354 (99%)	5 (1%)	0	100	100
1	B	356/361 (99%)	348 (98%)	8 (2%)	0	100	100
1	C	359/361 (99%)	354 (99%)	5 (1%)	0	100	100
1	D	355/361 (98%)	350 (99%)	5 (1%)	0	100	100
All	All	1429/1444 (99%)	1406 (98%)	23 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	304/304 (100%)	301 (99%)	3 (1%)	76	88
1	B	302/304 (99%)	295 (98%)	7 (2%)	50	70
1	C	304/304 (100%)	299 (98%)	5 (2%)	62	79
1	D	301/304 (99%)	300 (100%)	1 (0%)	92	97
All	All	1211/1216 (100%)	1195 (99%)	16 (1%)	71	84

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	319	ILE
1	C	314	LYS
1	B	288	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	141	LYS
1	B	221[B]	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	33	ASN
1	D	163	GLN
1	D	290	GLN

5.3.3 RNA (i)

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	360/361 (99%)	0.14	10 (2%)	53	51	25, 44, 74, 100	0
1	B	357/361 (98%)	0.26	15 (4%)	36	35	25, 45, 84, 101	0
1	C	360/361 (99%)	0.20	12 (3%)	46	45	25, 44, 75, 101	0
1	D	357/361 (98%)	0.33	21 (5%)	22	21	25, 45, 80, 105	0
All	All	1434/1444 (99%)	0.23	58 (4%)	38	37	25, 45, 77, 105	0

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	39	GLU	7.8
1	D	331	GLN	6.5
1	B	82	VAL	5.5
1	B	43	GLU	5.5
1	C	357	PHE	5.1

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.