



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 16, 2018 – 06:59 PM EST

PDB ID : 4HXZ
Title : Pyrrolopyrimidine inhibitors of dna gyrase b and topoisomerase iv, part i: structure guided discovery and optimization of dual targeting agents with potent, broad-spectrum enzymatic activity.
Authors : Bensen, D.C.; Creighton, C.J.; Kwan, B.; Tari, L.W.
Deposited on : 2012-11-12
Resolution : 2.70 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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030736
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
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) : rb-20030736

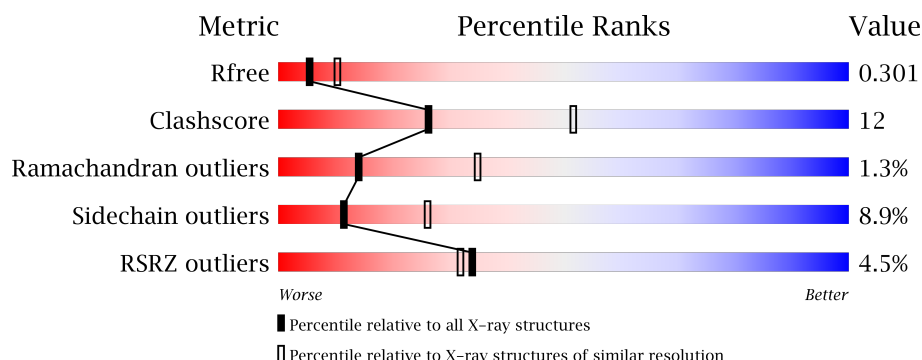
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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}	100719	2259 (2.70-2.70)
Clashscore	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)
RSRZ outliers	101464	2275 (2.70-2.70)

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	390	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, green 72%, yellow 18%, orange 6%);"></div> <div style="display: flex; justify-content: space-between; padding: 0 5px;"> 1% 72% 18% 6% </div> </div>
1	B	390	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 6%, green 48%, yellow 22%, orange 2%, grey 24%);"></div> <div style="display: flex; justify-content: space-between; padding: 0 5px;"> 6% 48% 22% • 27% </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5360 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.

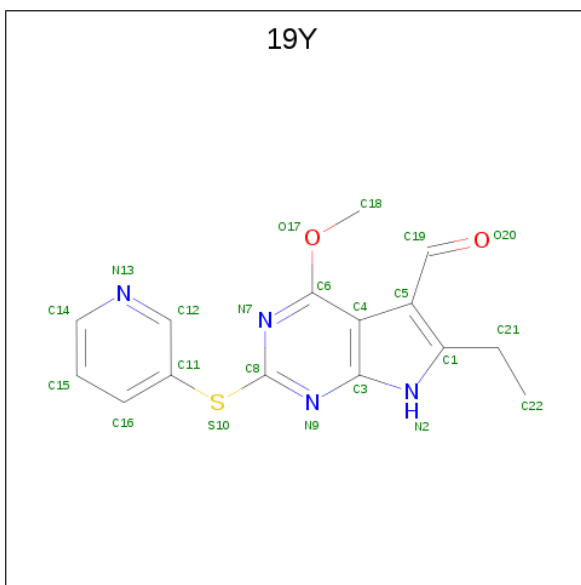
- Molecule 1 is a protein called Topoisomerase IV, subunit B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	366	Total	C	N	O	S	0	0	0
			2891	1842	485	557	7			
1	B	284	Total	C	N	O	S	0	0	0
			2253	1440	378	429	6			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	258	PRO	LEU	SEE REMARK 999	UNP Q2A1P5
A	383	LEU	-	EXPRESSION TAG	UNP Q2A1P5
A	384	GLU	-	EXPRESSION TAG	UNP Q2A1P5
A	385	HIS	-	EXPRESSION TAG	UNP Q2A1P5
A	386	HIS	-	EXPRESSION TAG	UNP Q2A1P5
A	387	HIS	-	EXPRESSION TAG	UNP Q2A1P5
A	388	HIS	-	EXPRESSION TAG	UNP Q2A1P5
A	389	HIS	-	EXPRESSION TAG	UNP Q2A1P5
A	390	HIS	-	EXPRESSION TAG	UNP Q2A1P5
B	258	PRO	LEU	SEE REMARK 999	UNP Q2A1P5
B	383	LEU	-	EXPRESSION TAG	UNP Q2A1P5
B	384	GLU	-	EXPRESSION TAG	UNP Q2A1P5
B	385	HIS	-	EXPRESSION TAG	UNP Q2A1P5
B	386	HIS	-	EXPRESSION TAG	UNP Q2A1P5
B	387	HIS	-	EXPRESSION TAG	UNP Q2A1P5
B	388	HIS	-	EXPRESSION TAG	UNP Q2A1P5
B	389	HIS	-	EXPRESSION TAG	UNP Q2A1P5
B	390	HIS	-	EXPRESSION TAG	UNP Q2A1P5

- Molecule 2 is 6-ethyl-4-methoxy-2-(pyridin-3-ylsulfanyl)-7H-pyrrolo[2,3-d]pyrimidine-5-carbaldehyde (three-letter code: 19Y) (formula: C₁₅H₁₄N₄O₂S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			22	15	4	2	1		
2	B	1	Total	C	N	O	S	0	0
			22	15	4	2	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	110	Total	O	0	0
			110	110		
3	B	62	Total	O	0	0
			62	62		

- Molecule 1: Topoisomerase IV, subunit B



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	44.12Å 162.87Å 77.41Å 90.00° 92.71° 90.00°	Depositor
Resolution (Å)	40.72 – 2.70 40.72 – 2.70	Depositor EDS
% Data completeness (in resolution range)	97.0 (40.72-2.70) 97.0 (40.72-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.20 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.240 , 0.303 0.239 , 0.301	Depositor DCC
R_{free} test set	1466 reflections (5.31%)	DCC
Wilson B-factor (Å ²)	40.7	Xtriage
Anisotropy	0.135	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 43.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	0.095 for h,-k,-l	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	5360	wwPDB-VP
Average B, all atoms (Å ²)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.20% 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: 19Y

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	0.59	2/2944 (0.1%)	0.72	0/3982
1	B	0.57	3/2294 (0.1%)	0.66	0/3094
All	All	0.58	5/5238 (0.1%)	0.69	0/7076

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	A	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	212	TRP	CD2-CE2	5.37	1.47	1.41
1	B	254	TRP	CD2-CE2	5.35	1.47	1.41
1	B	167	TRP	CD2-CE2	5.33	1.47	1.41
1	A	356	TRP	CD2-CE2	5.16	1.47	1.41
1	A	254	TRP	CD2-CE2	5.13	1.47	1.41

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	257	ASP	Peptide

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	2891	0	2912	60	0
1	B	2253	0	2262	60	0
2	A	22	0	14	2	0
2	B	22	0	14	5	0
3	A	110	0	0	4	0
3	B	62	0	0	6	0
All	All	5360	0	5202	121	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 12.

The worst 5 of 121 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:252:PHE:HD2	1:B:315:LEU:HD11	1.21	1.03
1:B:73:MET:HB2	1:B:129:ILE:CD1	1.94	0.98
1:A:245:ASP:N	1:A:246:SER:HA	1.79	0.97
1:B:254:TRP:HE3	3:B:1549:HOH:O	1.56	0.89
1:A:245:ASP:H	1:A:246:SER:HA	1.39	0.86

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	362/390 (93%)	325 (90%)	31 (9%)	6 (2%)	11 27

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	272/390 (70%)	243 (89%)	27 (10%)	2 (1%)	25	53
All	All	634/780 (81%)	568 (90%)	58 (9%)	8 (1%)	14	35

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	257	ASP
1	A	258	PRO
1	A	288	ASP
1	A	340	ASP
1	B	338	ASN

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	323/344 (94%)	299 (93%)	24 (7%)	16	37
1	B	253/344 (74%)	226 (89%)	27 (11%)	8	18
All	All	576/688 (84%)	525 (91%)	51 (9%)	11	26

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	383	LEU
1	B	34	LEU
1	B	340	ASP
1	B	12	LEU
1	B	58	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	327	GLN
1	A	331	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	185	ASN
1	A	316	ASN
1	A	380	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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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	19Y	A	1401	-	22,24,24	2.55	10 (45%)	19,33,33	3.26	7 (36%)
2	19Y	B	1401	-	22,24,24	2.85	8 (36%)	19,33,33	3.02	7 (36%)

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	19Y	A	1401	-	-	0/7/10/10	0/3/3/3

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	19Y	B	1401	-	-	0/7/10/10	0/3/3/3

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1401	19Y	C8-S10	-4.18	1.70	1.75
2	B	1401	19Y	C11-S10	-2.96	1.71	1.77
2	A	1401	19Y	C11-S10	-2.81	1.71	1.77
2	A	1401	19Y	C8-S10	-2.35	1.72	1.75
2	A	1401	19Y	C4-C3	2.04	1.48	1.43

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1401	19Y	C4-C6-N7	-9.14	117.33	124.62
2	B	1401	19Y	C4-C6-N7	-8.28	118.02	124.62
2	B	1401	19Y	C5-C4-C3	-4.92	103.61	107.54
2	A	1401	19Y	C5-C4-C3	-4.50	103.94	107.54
2	A	1401	19Y	N9-C8-N7	-3.77	120.09	126.85

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1401	19Y	2	0
2	B	1401	19Y	5	0

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	366/390 (93%)	-0.13	5 (1%) 75 76	16, 33, 71, 105	0
1	B	284/390 (72%)	0.23	24 (8%) 11 9	22, 41, 94, 114	0
All	All	650/780 (83%)	0.03	29 (4%) 34 32	16, 38, 89, 114	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	316	ASN	7.0
1	B	222	LEU	5.0
1	B	337	SER	4.2
1	B	231	LEU	4.2
1	A	259	SER	4.0

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	19Y	B	1401	22/22	0.95	0.15	-0.68	31,41,45,47	0
2	19Y	A	1401	22/22	0.96	0.14	-1.12	22,26,31,32	0

6.5 Other polymers [i](#)

There are no such residues in this entry.