Full wwPDB X-ray Structure Validation Report (i)

May 22, 2020 – 06:51 am BST

PDB ID	:	5K25
Title	:	Crystal structure of human phosphatase PRL-2 in complex with the ADP-bound Bateman domain of human magnesium transporter CNNM3
Authors	:	Gulerez, I.; Kozlov, G.; Gehring, K.
Deposited on	:	2016-05-18
Resolution	:	3.05 Å(reported)

The following versions of software and data (see references (i)) 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
buster-report	:	1.1.7 (2018)
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 (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.05 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	Percentile Ranks	Value
Rfree		0.283
Clashscore		7
Ramachandran outliers		0.7%
Sidechain outliers		18.2%
RSRZ outliers		1.0%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
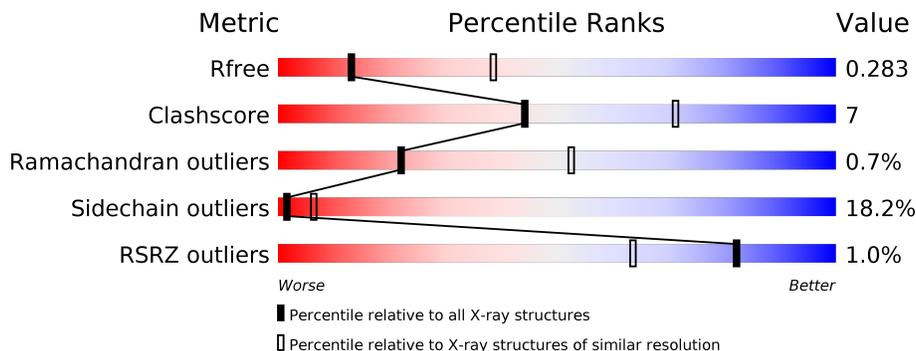
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	1754 (3.10-3.02)
Clashscore	141614	1864 (3.10-3.02)
Ramachandran outliers	138981	1794 (3.10-3.02)
Sidechain outliers	138945	1793 (3.10-3.02)
RSRZ outliers	127900	1713 (3.10-3.02)

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 <=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	190	2% ▮ 54% 23% 6% • 17%
2	C	156	66% 24% 6% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2448 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 type IVA 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	157	1235	791	214	221	9	0	0	0

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-22	MET	-	initiating methionine	UNP Q12974
A	-21	SER	-	expression tag	UNP Q12974
A	-20	TYR	-	expression tag	UNP Q12974
A	-19	TYR	-	expression tag	UNP Q12974
A	-18	HIS	-	expression tag	UNP Q12974
A	-17	HIS	-	expression tag	UNP Q12974
A	-16	HIS	-	expression tag	UNP Q12974
A	-15	HIS	-	expression tag	UNP Q12974
A	-14	HIS	-	expression tag	UNP Q12974
A	-13	HIS	-	expression tag	UNP Q12974
A	-12	LEU	-	expression tag	UNP Q12974
A	-11	GLU	-	expression tag	UNP Q12974
A	-10	SER	-	expression tag	UNP Q12974
A	-9	THR	-	expression tag	UNP Q12974
A	-8	SER	-	expression tag	UNP Q12974
A	-7	LEU	-	expression tag	UNP Q12974
A	-6	TYR	-	expression tag	UNP Q12974
A	-5	LYS	-	expression tag	UNP Q12974
A	-4	LYS	-	expression tag	UNP Q12974
A	-3	ALA	-	expression tag	UNP Q12974
A	-2	GLY	-	expression tag	UNP Q12974
A	-1	PHE	-	expression tag	UNP Q12974
A	0	ALA	-	expression tag	UNP Q12974

- Molecule 2 is a protein called Metal transporter CNNM3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	150	1183	755	189	232	7	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	297	GLN	-	expression tag	UNP Q8NE01
C	298	GLY	-	expression tag	UNP Q8NE01
C	299	PRO	-	expression tag	UNP Q8NE01
C	300	LEU	-	expression tag	UNP Q8NE01
C	301	ASN	-	expression tag	UNP Q8NE01
C	302	MET	-	expression tag	UNP Q8NE01
C	303	ILE	-	expression tag	UNP Q8NE01
C	304	GLN	-	expression tag	UNP Q8NE01
C	305	GLY	-	expression tag	UNP Q8NE01
C	306	VAL	-	expression tag	UNP Q8NE01
C	307	LEU	-	expression tag	UNP Q8NE01
C	308	GLU	-	expression tag	UNP Q8NE01

- Molecule 3 is ADENOSINE-5'-DIPHOSPHATE (three-letter code: ADP) (formula: $C_{10}H_{15}N_5O_{10}P_2$).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
3	C	1	27	10	5	10	2	0	0

- Molecule 4 is water.
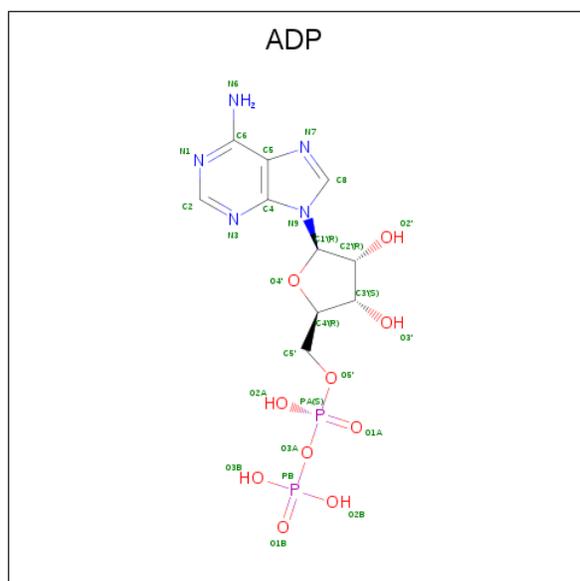
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	3	Total 3	O 3	0	0

3 Residue-property plots (i)

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 (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: Protein tyrosine phosphatase type IVA 2

Chain A:
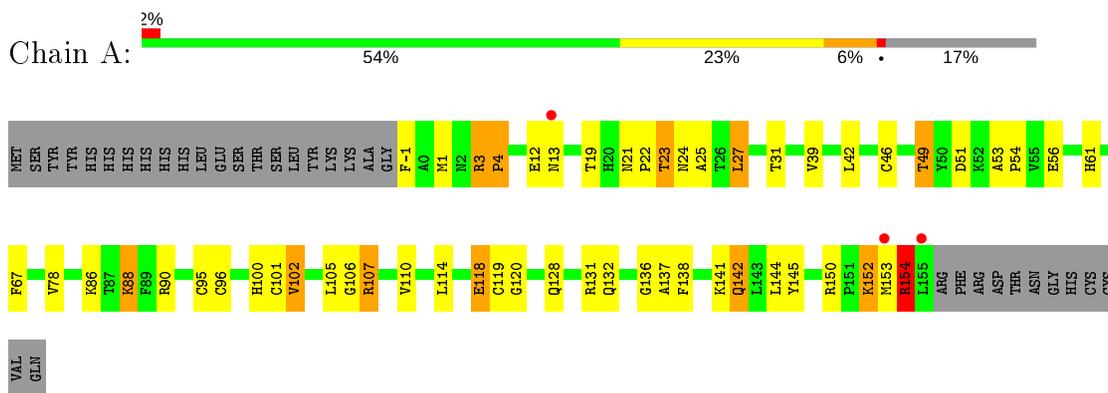
- Molecule 2: Metal transporter CNNM3

Chain C:
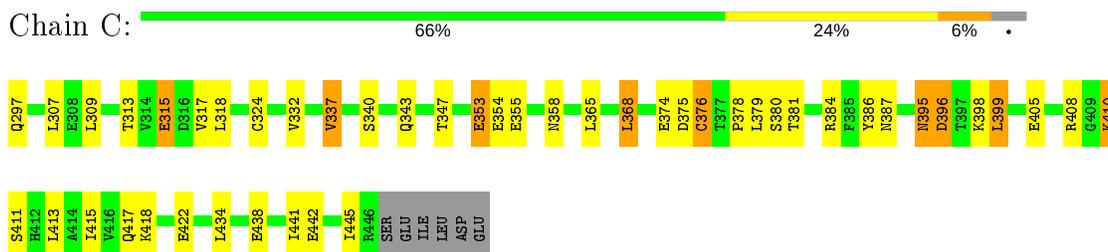
4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	52.15Å 124.40Å 164.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.71 – 3.05 49.71 – 3.05	Depositor EDS
% Data completeness (in resolution range)	99.7 (49.71-3.05) 99.7 (49.71-3.05)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	5.07 (at 3.07Å)	Xtriage
Refinement program	REFMAC 5.8.0069	Depositor
R, R_{free}	0.232 , 0.283 0.232 , 0.283	Depositor DCC
R_{free} test set	503 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å^2)	94.3	Xtriage
Anisotropy	0.018	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 58.3	EDS
L-test for twinning[2]	$<|L|> = 0.50$, $<L^2> = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2448	wwPDB-VP
Average B, all atoms (Å^2)	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.96% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[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)

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

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	0.28	0/1265	0.45	0/1722
2	C	0.28	0/1205	0.46	0/1639
All	All	0.28	0/2470	0.45	0/3361

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	8
2	C	0	12
All	All	0	20

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (20) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	120	GLY	Peptide
1	A	136	GLY	Peptide
1	A	152	LYS	Peptide
1	A	153	MET	Peptide
1	A	154	ARG	Peptide
1	A	22	PRO	Peptide
1	A	4	PRO	Peptide
1	A	49	THR	Peptide
2	C	297	GLN	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
2	C	353	GLU	Peptide
2	C	368	LEU	Peptide
2	C	374	GLU	Peptide
2	C	375	ASP	Peptide
2	C	386	TYR	Peptide
2	C	387	ASN	Peptide
2	C	396	ASP	Peptide
2	C	408	ARG	Peptide
2	C	417	GLN	Peptide
2	C	422	GLU	Peptide
2	C	445	ILE	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	1235	0	1215	18	0
2	C	1183	0	1154	15	0
3	C	27	0	12	2	0
4	C	3	0	0	0	0
All	All	2448	0	2381	32	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 7.

All (32) 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:101:CYS:SG	1:A:102:VAL:N	2.75	0.59
2:C:332:VAL:O	2:C:337:VAL:HG11	2.03	0.59
1:A:53:ALA:N	1:A:54:PRO:HD2	2.19	0.58
2:C:324:CYS:HA	3:C:501:ADP:C2	2.40	0.57
2:C:413:LEU:HD21	2:C:434:LEU:HD22	1.88	0.55
2:C:324:CYS:HA	3:C:501:ADP:H2	1.72	0.55
1:A:131:ARG:HE	1:A:137:ALA:H	1.56	0.54
1:A:24:ASN:HA	1:A:27:LEU:HD12	1.89	0.53

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1:MET:HG3	2:C:405:GLU:HG2	1.90	0.53
1:A:118:GLU:HG2	1:A:154:ARG:H	1.75	0.52
1:A:3:ARG:HG3	1:A:4:PRO:HD2	1.92	0.52
1:A:61:HIS:HD2	1:A:88:LYS:HE2	1.75	0.51
1:A:39:VAL:HG12	1:A:96:CYS:HB3	1.93	0.51
2:C:380:SER:O	2:C:384:ARG:HB2	2.11	0.50
2:C:410:LYS:HG3	2:C:411:SER:N	2.28	0.49
2:C:332:VAL:HA	2:C:378:PRO:HA	1.97	0.47
1:A:23:THR:HB	1:A:25:ALA:H	1.79	0.47
2:C:315:GLU:HG2	2:C:395:ASN:HB2	1.96	0.47
1:A:114:LEU:HD11	1:A:145:TYR:HE2	1.82	0.45
2:C:313:THR:HG22	2:C:398:LYS:HD3	1.99	0.45
2:C:395:ASN:H	2:C:395:ASN:HD22	1.65	0.44
1:A:12:GLU:HB3	1:A:13:ASN:H	1.61	0.43
2:C:438:GLU:HA	2:C:441:ILE:HD12	2.00	0.43
2:C:317:VAL:HG21	2:C:399:LEU:HD21	2.01	0.43
2:C:332:VAL:CG1	2:C:376:CYS:HB3	2.49	0.43
1:A:142:GLN:H	1:A:142:GLN:HG2	1.64	0.42
1:A:106:GLY:O	1:A:110:VAL:HG23	2.19	0.41
1:A:131:ARG:HH21	1:A:138:PHE:H	1.69	0.41
1:A:67:PHE:CZ	1:A:107:ARG:HB2	2.56	0.41
1:A:46:CYS:SG	1:A:107:ARG:CZ	3.10	0.40
1:A:86:LYS:O	1:A:90:ARG:HB2	2.22	0.40
2:C:353:GLU:O	2:C:358:ASN:HB3	2.21	0.40

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	155/190 (82%)	144 (93%)	10 (6%)	1 (1%)	25	55

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	148/156 (95%)	139 (94%)	8 (5%)	1 (1%)	22	52
All	All	303/346 (88%)	283 (93%)	18 (6%)	2 (1%)	22	52

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	376	CYS
1	A	152	LYS

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	130/166 (78%)	103 (79%)	27 (21%)	1	4
2	C	134/143 (94%)	113 (84%)	21 (16%)	2	10
All	All	264/309 (85%)	216 (82%)	48 (18%)	1	6

All (48) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	-1	PHE
1	A	3	ARG
1	A	19	THR
1	A	21	ASN
1	A	23	THR
1	A	27	LEU
1	A	31	THR
1	A	42	LEU
1	A	49	THR
1	A	51	ASP
1	A	56	GLU
1	A	78	VAL
1	A	88	LYS
1	A	95	CYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	100	HIS
1	A	102	VAL
1	A	105	LEU
1	A	107	ARG
1	A	118	GLU
1	A	119	CYS
1	A	128	GLN
1	A	132	GLN
1	A	141	LYS
1	A	142	GLN
1	A	144	LEU
1	A	150	ARG
1	A	154	ARG
2	C	307	LEU
2	C	309	LEU
2	C	315	GLU
2	C	318	LEU
2	C	337	VAL
2	C	340	SER
2	C	343	GLN
2	C	347	THR
2	C	354	GLU
2	C	355	GLU
2	C	365	LEU
2	C	368	LEU
2	C	379	LEU
2	C	381	THR
2	C	395	ASN
2	C	396	ASP
2	C	399	LEU
2	C	410	LYS
2	C	415	ILE
2	C	418	LYS
2	C	442	GLU

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

Mol	Chain	Res	Type
1	A	61	HIS
2	C	395	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

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	#$|Z| > 2$
3	ADP	C	501	-	24,29,29	0.99	1 (4%)	29,45,45	1.33	3 (10%)

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
3	ADP	C	501	-	-	4/12/32/32	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	501	ADP	C5-C4	2.60	1.47	1.40

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	501	ADP	PA-O3A-PB	-3.20	121.86	132.83
3	C	501	ADP	N3-C2-N1	-3.02	123.96	128.68
3	C	501	ADP	C4-C5-N7	-2.82	106.46	109.40

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	501	ADP	C5'-O5'-PA-O3A
3	C	501	ADP	C5'-O5'-PA-O1A
3	C	501	ADP	O4'-C4'-C5'-O5'
3	C	501	ADP	C3'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	501	ADP	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand ADP C 501

Bond lengths

Bond angles

Torsions

Rings
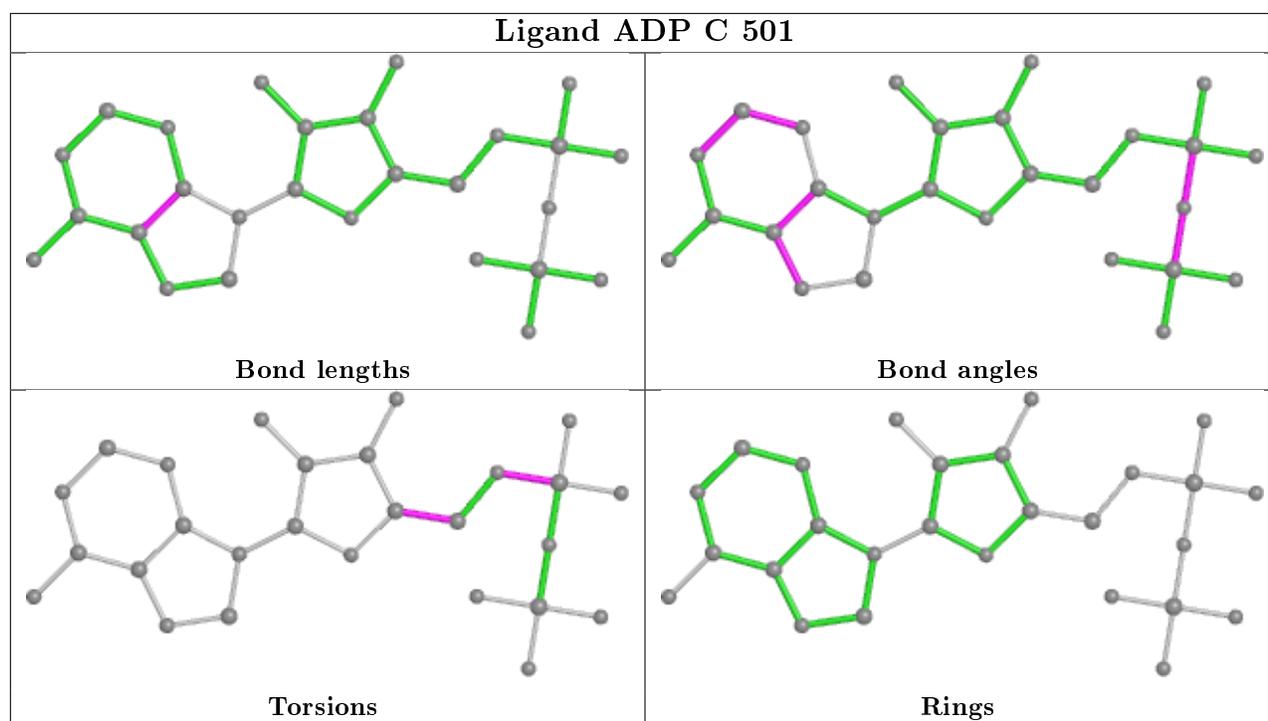
5.7 Other polymers (i)

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 ⓘ

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	157/190 (82%)	0.35	3 (1%)	66	43	71, 99, 128, 148	0
2	C	150/156 (96%)	0.26	0	100	100	65, 79, 107, 120	0
All	All	307/346 (88%)	0.31	3 (0%)	82	63	65, 89, 123, 148	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	13	ASN	3.8
1	A	153	MET	2.5
1	A	155	LEU	2.2

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. 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(Å^2)	Q<0.9
3	ADP	C	501	27/27	0.94	0.20	87,92,108,111	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around ADP C 501:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

6.5 Other polymers (i)

There are no such residues in this entry.