



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 11:06 am BST

PDB ID : 5YH2  
Title : The structure of DrFam20C1 and hFam20A complex  
Authors : Zhang, H.; Xiao, J.  
Deposited on : 2017-09-27  
Resolution : 3.55 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](mailto: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  
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

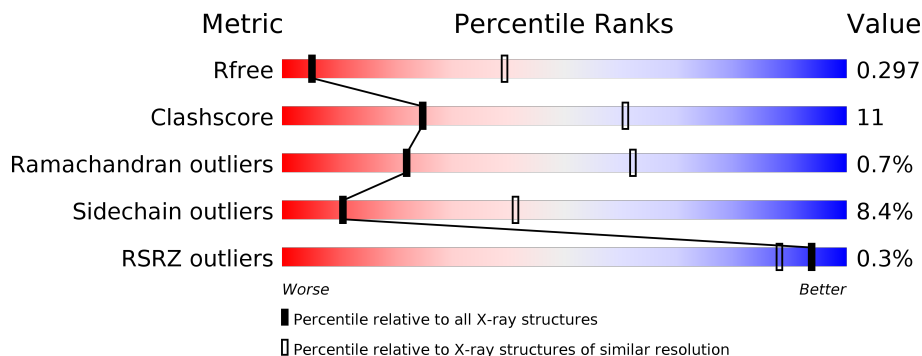
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.55 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	1020 (3.62-3.50)
Clashscore	141614	1100 (3.62-3.50)
Ramachandran outliers	138981	1065 (3.62-3.50)
Sidechain outliers	138945	1066 (3.62-3.50)
RSRZ outliers	127900	1009 (3.64-3.48)

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  $\geq 3$ , 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	467	
1	B	467	
2	C	560	
2	D	560	

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 13493 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 Pseudokinase FAM20A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	428	3459	2214	608	617	20	0	0	0
1	B	428	3463	2219	608	616	20	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	332	LYS	ASN	variant	UNP Q96MK3
B	332	LYS	ASN	variant	UNP Q96MK3

- Molecule 2 is a protein called Family with sequence similarity 20, member Ca.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	395	3258	2078	572	590	18	0	0	0
2	D	394	3251	2074	571	588	18	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

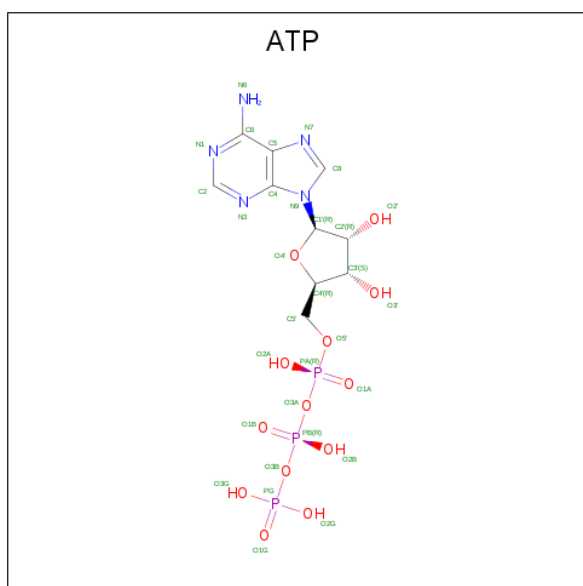
Chain	Residue	Modelled	Actual	Comment	Reference
C	223	ALA	THR	engineered mutation	UNP E7FBB8
C	421	MET	LYS	engineered mutation	UNP E7FBB8
C	422	ASP	ALA	engineered mutation	UNP E7FBB8
C	423	MET	LEU	engineered mutation	UNP E7FBB8
C	425	ILE	HIS	engineered mutation	UNP E7FBB8
C	426	PHE	TYR	engineered mutation	UNP E7FBB8
C	427	ASP	SER	engineered mutation	UNP E7FBB8
C	428	PHE	LEU	engineered mutation	UNP E7FBB8
C	429	LEU	LYS	engineered mutation	UNP E7FBB8
C	430	MET	THR	engineered mutation	UNP E7FBB8

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	223	ALA	THR	engineered mutation	UNP E7FBB8
D	421	MET	LYS	engineered mutation	UNP E7FBB8
D	422	ASP	ALA	engineered mutation	UNP E7FBB8
D	423	MET	LEU	engineered mutation	UNP E7FBB8
D	425	ILE	HIS	engineered mutation	UNP E7FBB8
D	426	PHE	TYR	engineered mutation	UNP E7FBB8
D	427	ASP	SER	engineered mutation	UNP E7FBB8
D	428	PHE	LEU	engineered mutation	UNP E7FBB8
D	429	LEU	LYS	engineered mutation	UNP E7FBB8
D	430	MET	THR	engineered mutation	UNP E7FBB8

- Molecule 3 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula:  $C_{10}H_{16}N_5O_{13}P_3$ ).

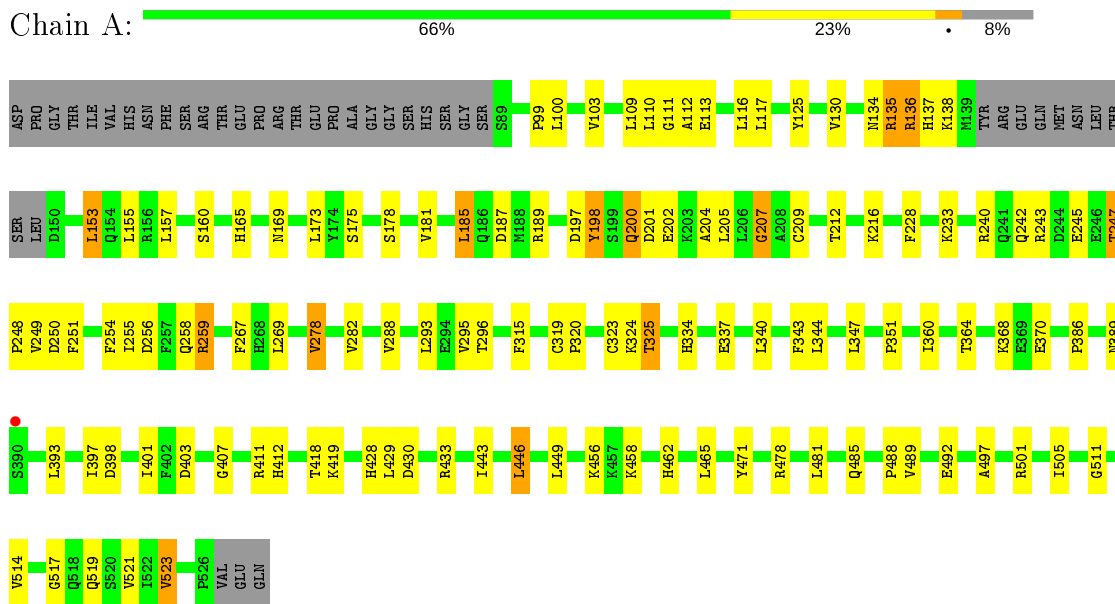


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
			Total	C	N	O			P
3	A	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
3	B	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

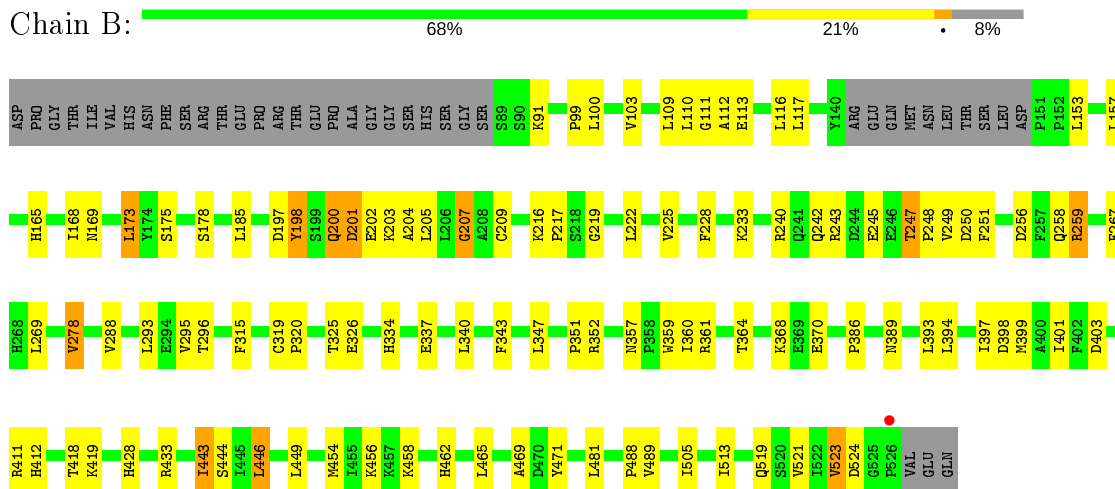
### 3 Residue-property plots

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: Pseudokinase FAM20A



- Molecule 1: Pseudokinase FAM20A



- Molecule 2: Family with sequence similarity 20, member Ca



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	110.06Å 110.06Å 461.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.94 – 3.55 49.67 – 3.55	Depositor EDS
% Data completeness (in resolution range)	87.7 (48.94-3.55) 80.7 (49.67-3.55)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.08 (at 3.57Å)	Xtrriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, $R_{free}$	0.245 , 0.298 0.246 , 0.297	Depositor DCC
$R_{free}$ test set	1560 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	55.5	Xtrriage
Anisotropy	0.231	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 2.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	13493	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.88% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical 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

### 5.1 Standard geometry

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

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	0.30	0/3545	0.48	0/4803
1	B	0.30	0/3550	0.47	0/4809
2	C	0.31	0/3337	0.49	0/4508
2	D	0.30	0/3330	0.49	0/4498
All	All	0.30	0/13762	0.48	0/18618

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.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	201	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	3459	0	3447	66	0
1	B	3463	0	3453	58	0
2	C	3258	0	3218	88	0
2	D	3251	0	3211	88	0
3	A	31	0	12	2	0
3	B	31	0	12	2	0
All	All	13493	0	13353	291	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 11.

The worst 5 of 291 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)
2:D:414:GLY:HA3	2:D:416:ARG:H	1.33	0.92
2:C:414:GLY:HA3	2:C:416:ARG:H	1.42	0.84
1:A:185:LEU:HD23	1:A:282:VAL:HG12	1.64	0.80
2:D:309:LEU:HD12	2:D:509:ARG:HE	1.50	0.77
2:C:506:LEU:HD12	2:C:511:VAL:HG21	1.68	0.76

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	424/467 (91%)	390 (92%)	30 (7%)	4 (1%)	17	57
1	B	424/467 (91%)	390 (92%)	32 (8%)	2 (0%)	29	67
2	C	391/560 (70%)	348 (89%)	41 (10%)	2 (0%)	29	67
2	D	390/560 (70%)	352 (90%)	35 (9%)	3 (1%)	19	59
All	All	1629/2054 (79%)	1480 (91%)	138 (8%)	11 (1%)	22	62

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	136	ARG
2	C	161	LEU
2	D	161	LEU
1	A	112	ALA
1	A	207	GLY

### 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	384/418 (92%)	353 (92%)	31 (8%)	11	42
1	B	384/418 (92%)	350 (91%)	34 (9%)	9	39
2	C	359/495 (72%)	328 (91%)	31 (9%)	10	41
2	D	358/495 (72%)	330 (92%)	28 (8%)	12	43
All	All	1485/1826 (81%)	1361 (92%)	124 (8%)	11	41

5 of 124 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	484	TYR
1	B	201	ASP
2	D	436	HIS
2	C	490	LEU
1	B	110	LEU

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

Mol	Chain	Res	Type
1	A	134	ASN
2	C	142	HIS
2	D	142	HIS

### 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](#)

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
3	ATP	A	701	-	26,33,33	0.92	1 (3%)	31,52,52	1.59	5 (16%)
3	ATP	B	701	-	26,33,33	0.90	1 (3%)	31,52,52	1.58	5 (16%)

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	ATP	A	701	-	-	6/18/38/38	0/3/3/3
3	ATP	B	701	-	-	9/18/38/38	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	701	ATP	C5-C4	2.43	1.47	1.40

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	701	ATP	C5-C4	2.38	1.47	1.40

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	701	ATP	PA-O3A-PB	-3.98	119.17	132.83
3	A	701	ATP	PA-O3A-PB	-3.98	119.17	132.83
3	A	701	ATP	C3'-C2'-C1'	3.63	106.45	100.98
3	A	701	ATP	N3-C2-N1	-3.34	123.46	128.68
3	B	701	ATP	N3-C2-N1	-3.31	123.51	128.68

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

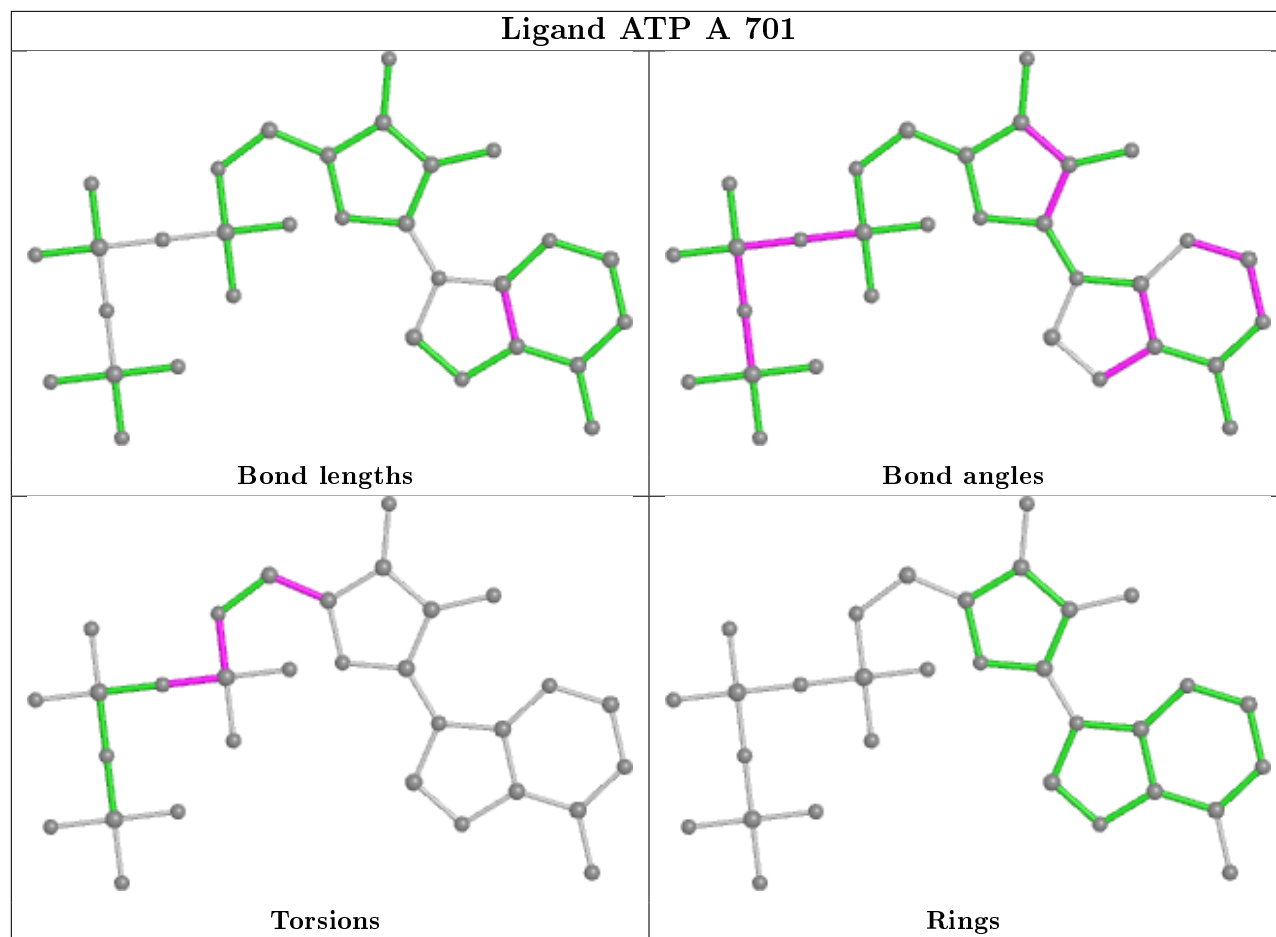
Mol	Chain	Res	Type	Atoms
3	A	701	ATP	C5'-O5'-PA-O2A
3	A	701	ATP	O4'-C4'-C5'-O5'
3	B	701	ATP	C5'-O5'-PA-O1A
3	B	701	ATP	C5'-O5'-PA-O2A
3	B	701	ATP	O4'-C4'-C5'-O5'

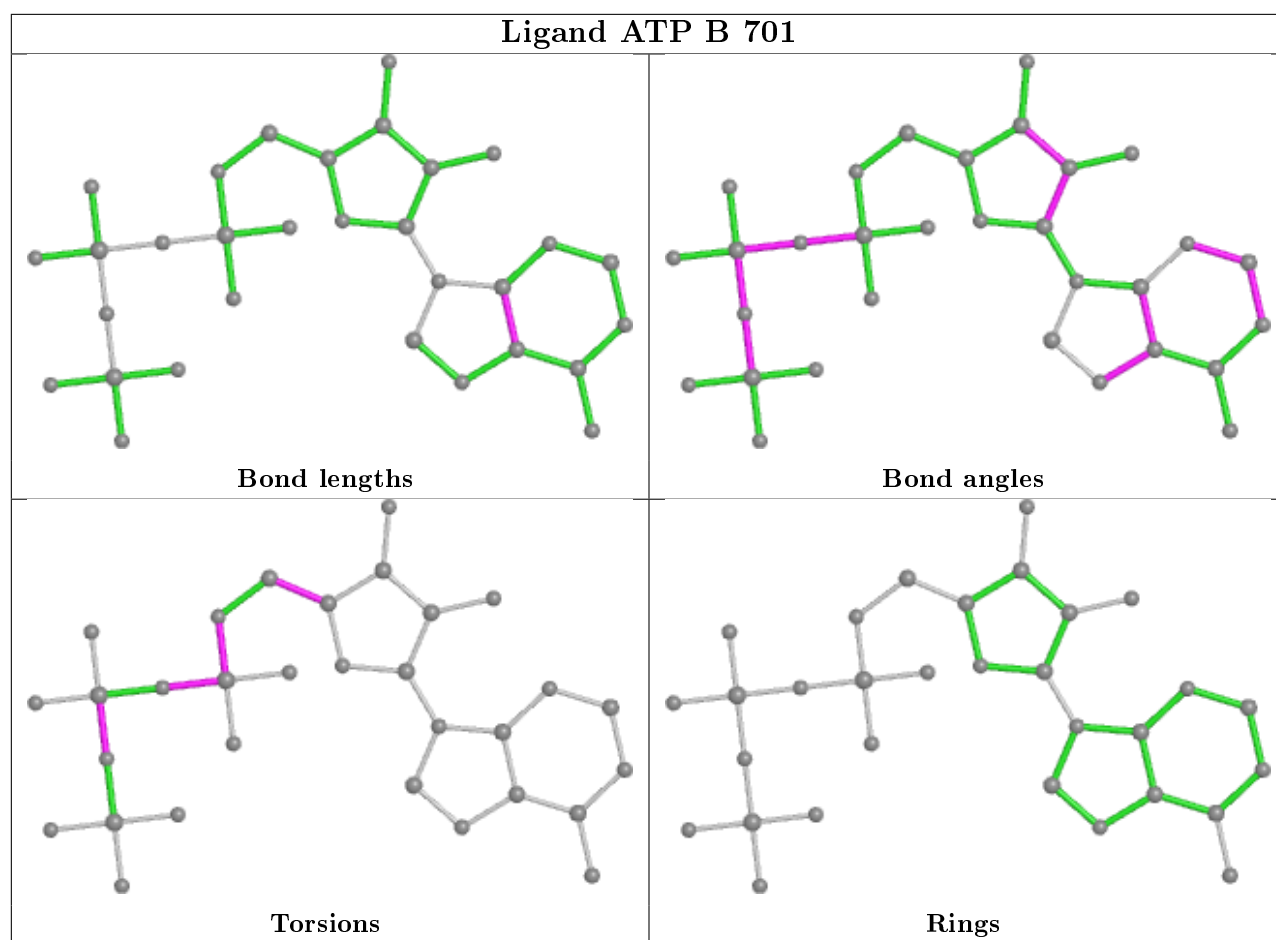
There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	701	ATP	2	0
3	B	701	ATP	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 than 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.





## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	428/467 (91%)	-0.23	1 (0%) 95 91	38, 52, 85, 114	0
1	B	428/467 (91%)	-0.31	1 (0%) 95 91	38, 49, 81, 116	0
2	C	395/560 (70%)	-0.15	1 (0%) 94 88	39, 61, 86, 107	0
2	D	394/560 (70%)	-0.19	2 (0%) 91 83	42, 61, 85, 110	0
All	All	1645/2054 (80%)	-0.22	5 (0%) 94 88	38, 56, 85, 116	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	535	ASP	2.6
1	A	390	SER	2.4
2	D	443	LYS	2.3
2	D	242	GLY	2.2
1	B	526	PRO	2.2

### 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.

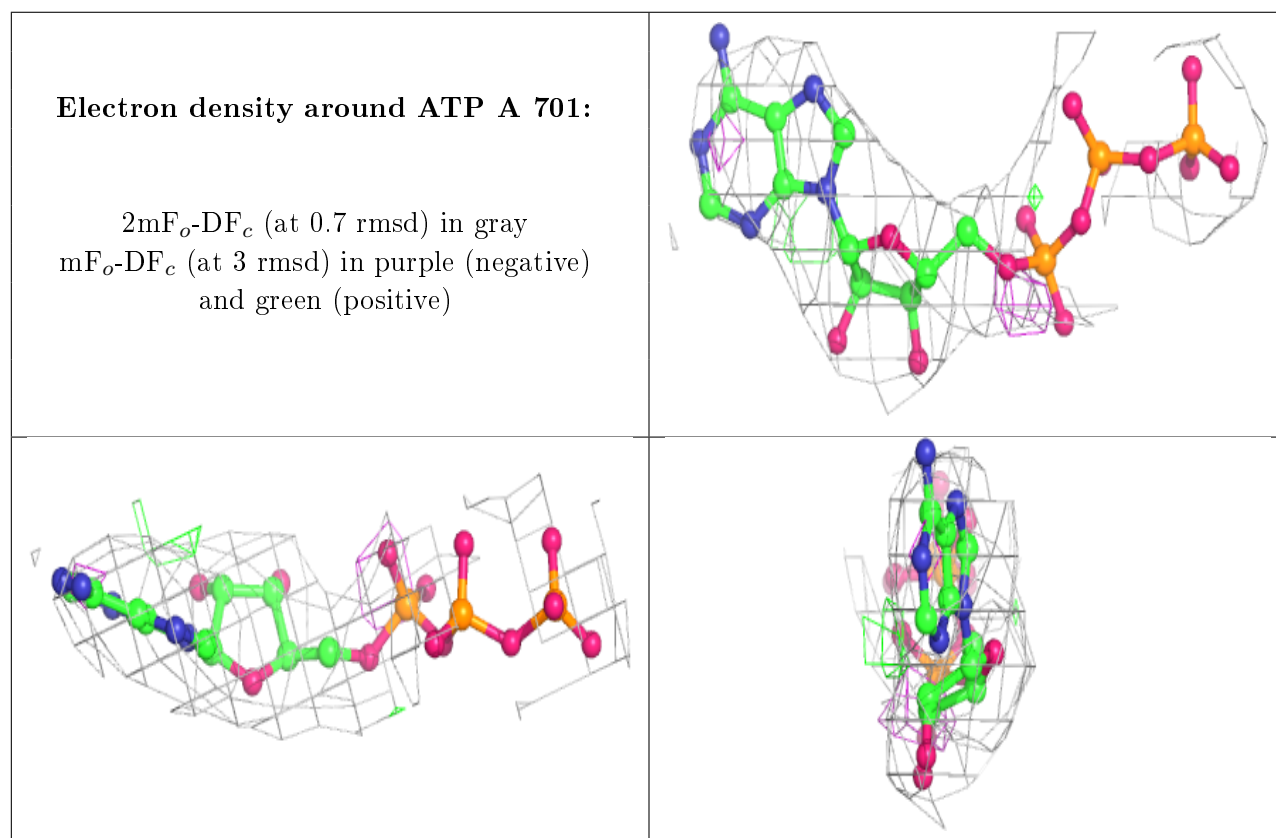
### 6.4 Ligands [i](#)

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,

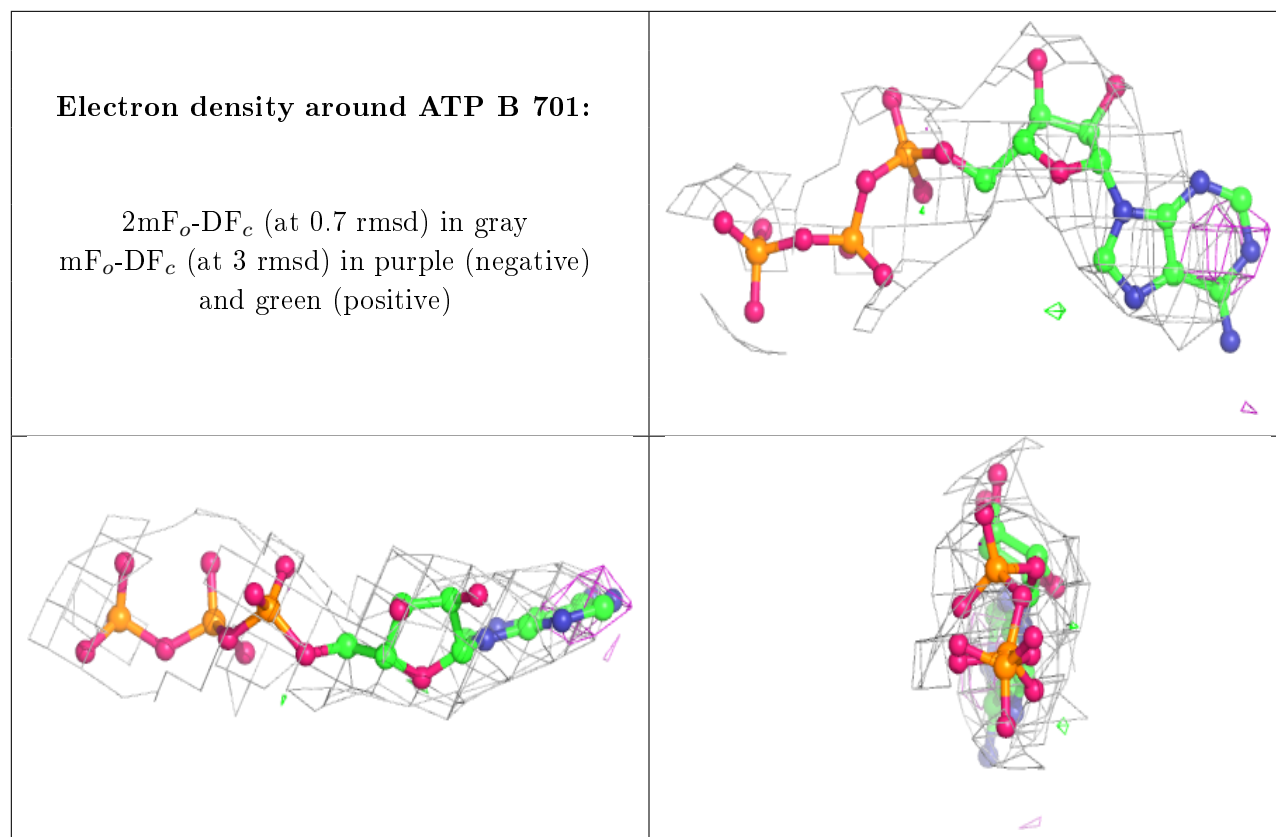
median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
3	ATP	A	701	31/31	0.91	0.19	49,68,93,101	0
3	ATP	B	701	31/31	0.92	0.21	45,61,78,88	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.







## 6.5 Other polymers [i](#)

There are no such residues in this entry.