wwPDB X-ray Structure Validation Summary Report (i)

May 18, 2020 – 01:57 pm BST

PDB ID	:	6B4I
Title	:	Crystal structure of human Gle1 CTD-Nup42 GBM-DDX19B(ADP) complex
Authors	:	Lin, D.H.; Correia, A.R.; Cai, S.W.; Huber, F.M.; Jette, C.A.; Hoelz, A.
Deposited on	:	2017-09-26
Resolution	:	3.62 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 following versions of software and data (see references (i)) were used in the production of this report:

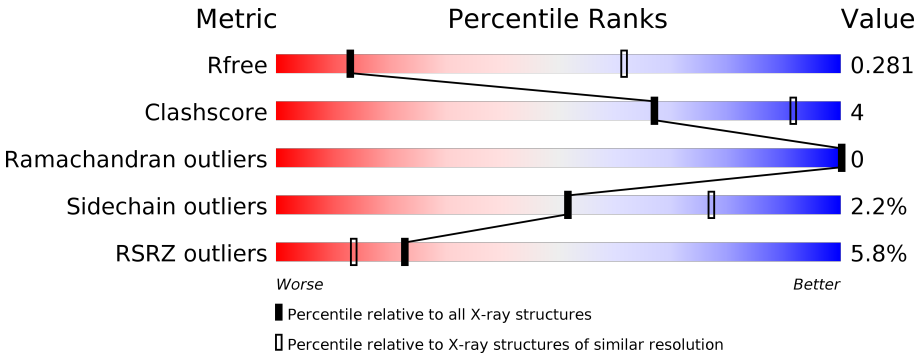
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (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 3.62 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	1290 (3.74-3.50)
Clashscore	141614	1387 (3.74-3.50)
Ramachandran outliers	138981	1339 (3.74-3.50)
Sidechain outliers	138945	1339 (3.74-3.50)
RSRZ outliers	127900	1191 (3.74-3.50)

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	317	3% 85% 13%
1	B	317	% 86% 12%
2	C	50	4% 84% 6% 10%
2	D	50	6% 68% 10% 22%
3	E	430	6% 89% 10%
3	F	430	11% 92% 6%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	PO4	A	701	-	-	-	X
4	PO4	E	502	-	-	X	-
4	PO4	E	503	-	-	X	-

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 25381 atoms, of which 12777 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 Nucleoporin GLE1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	B	315	Total	C	H	N	O	S	0	0	0
			5099	1640	2558	432	454	15			
1	A	314	Total	C	H	N	O	S	0	0	0
			5087	1637	2553	431	451	15			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	382	MET	-	initiating methionine	UNP Q53GS7
A	382	MET	-	initiating methionine	UNP Q53GS7

- Molecule 2 is a protein called Nucleoporin like 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	45	Total	C	H	N	O	0	0	0
			751	239	388	57	67			
2	D	39	Total	C	H	N	O	0	0	0
			663	212	344	50	57			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	374	GLY	-	expression tag	UNP Q3B7J4
C	375	PRO	-	expression tag	UNP Q3B7J4
C	376	SER	-	expression tag	UNP Q3B7J4
C	377	GLY	-	expression tag	UNP Q3B7J4
C	378	SER	-	expression tag	UNP Q3B7J4
C	379	ILE	-	expression tag	UNP Q3B7J4
C	380	ILE	-	expression tag	UNP Q3B7J4
D	374	GLY	-	expression tag	UNP Q3B7J4
D	375	PRO	-	expression tag	UNP Q3B7J4
D	376	SER	-	expression tag	UNP Q3B7J4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	377	GLY	-	expression tag	UNP Q3B7J4
D	378	SER	-	expression tag	UNP Q3B7J4
D	379	ILE	-	expression tag	UNP Q3B7J4
D	380	ILE	-	expression tag	UNP Q3B7J4

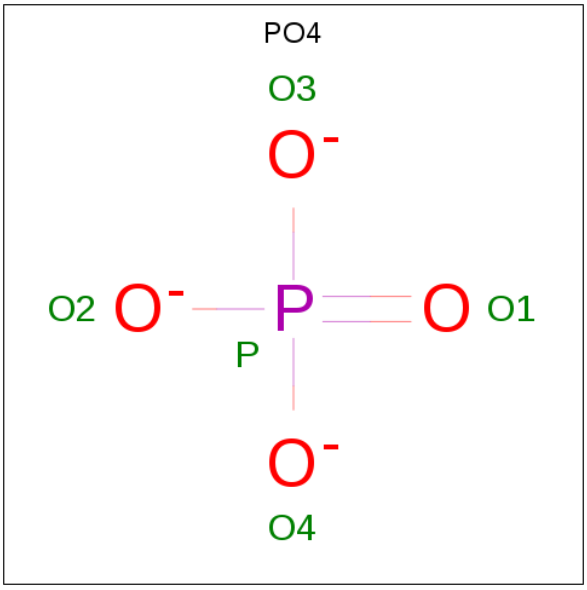
- Molecule 3 is a protein called ATP-dependent RNA helicase DDX19B.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	E	425	Total	C	H	N	O	S	0	2	0
			6827	2132	3450	592	633	20			
3	F	426	Total	C	H	N	O	S	0	2	0
			6841	2137	3455	593	636	20			

There are 8 discrepancies between the modelled and reference sequences:

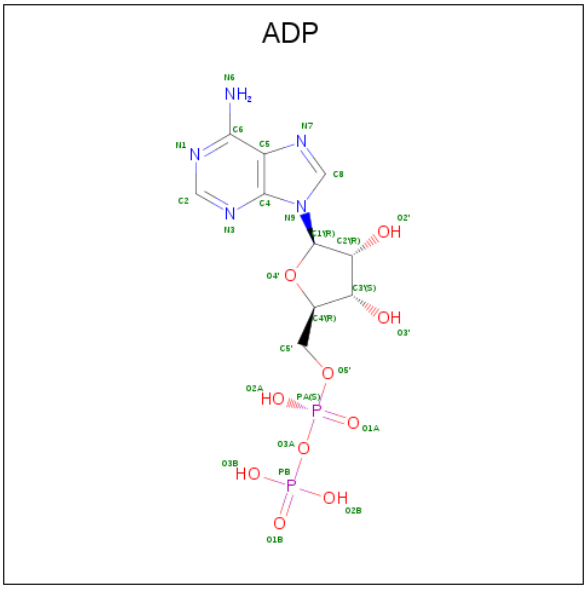
Chain	Residue	Modelled	Actual	Comment	Reference
E	50	GLY	-	expression tag	UNP Q9UMR2
E	51	PRO	-	expression tag	UNP Q9UMR2
E	52	HIS	-	expression tag	UNP Q9UMR2
E	53	MET	-	expression tag	UNP Q9UMR2
F	50	GLY	-	expression tag	UNP Q9UMR2
F	51	PRO	-	expression tag	UNP Q9UMR2
F	52	HIS	-	expression tag	UNP Q9UMR2
F	53	MET	-	expression tag	UNP Q9UMR2

- Molecule 4 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total O P 5 4 1	0	0
4	A	1	Total O P 5 4 1	0	0
4	E	1	Total O P 5 4 1	0	0
4	E	1	Total O P 5 4 1	0	0
4	F	1	Total O P 5 4 1	0	0
4	F	1	Total O P 5 4 1	0	0

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



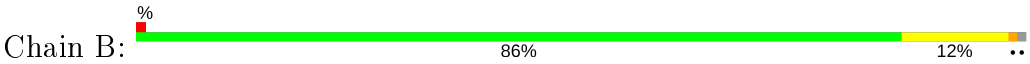
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	E	1	Total C H N O P 42 10 15 5 10 2	0	0
5	F	1	Total C H N O P 41 10 14 5 10 2	0	0

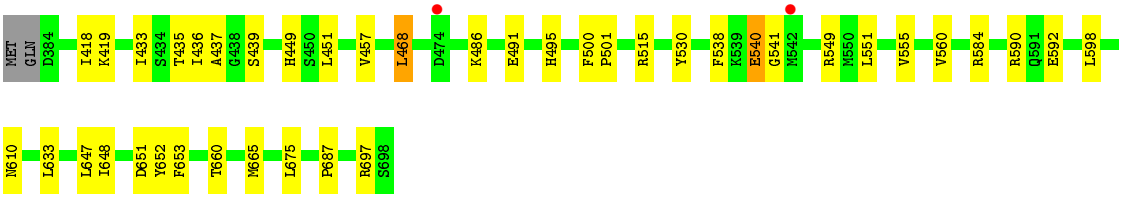
3 Residue-property plots (i)

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: Nucleoporin GLE1

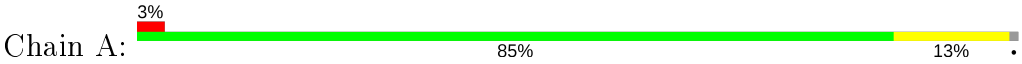
Chain B:

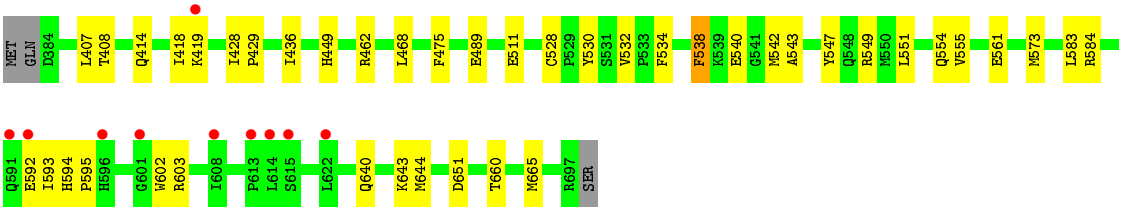




- Molecule 1: Nucleoporin GLE1

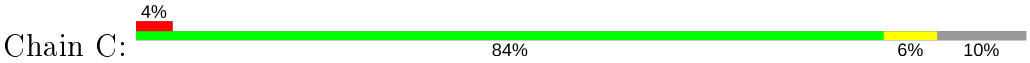
Chain A:

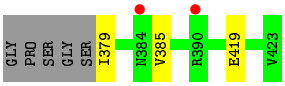




- Molecule 2: Nucleoporin like 2

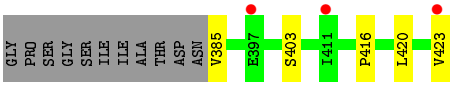
Chain C:





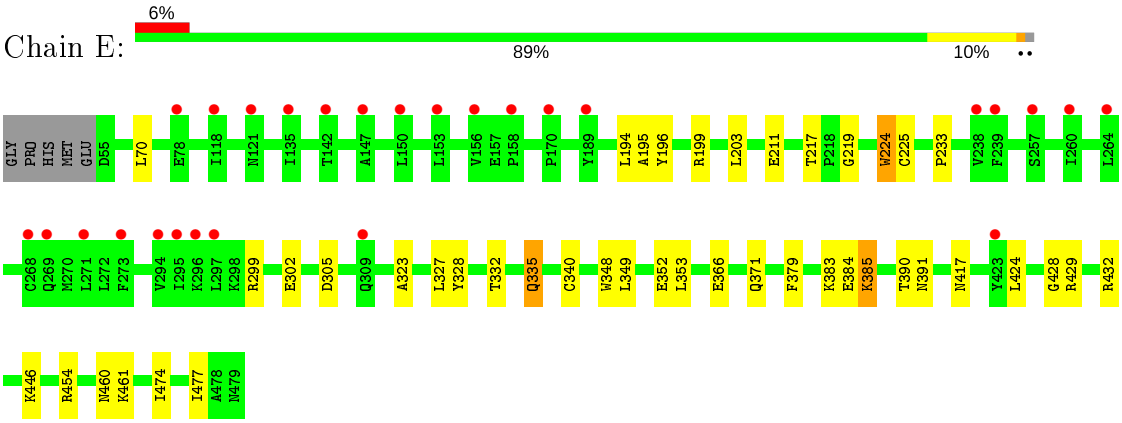
- Molecule 2: Nucleoporin like 2

Chain D:



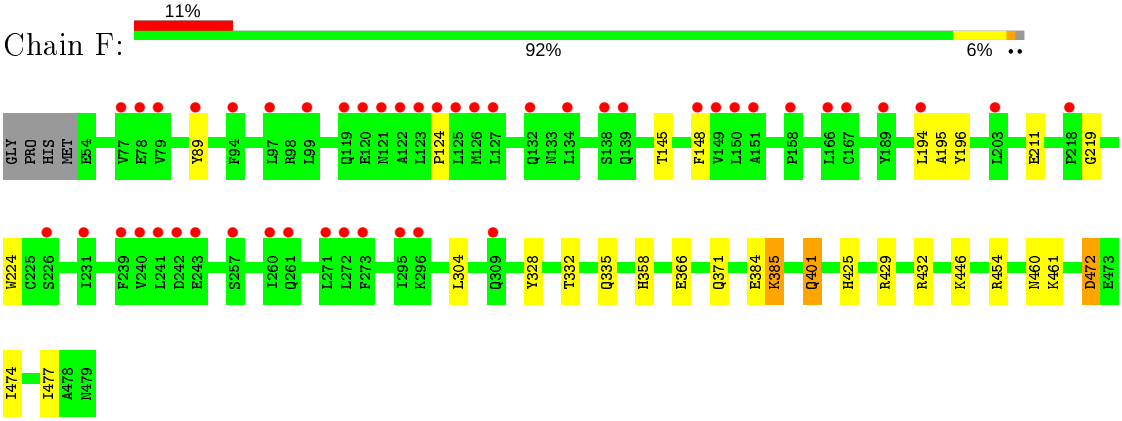
- Molecule 3: ATP-dependent RNA helicase DDX19B

Chain E:



- Molecule 3: ATP-dependent RNA helicase DDX19B

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	87.74Å 74.65Å 146.83Å 90.00° 94.85° 90.00°	Depositor
Resolution (Å)	46.09 – 3.62 48.77 – 3.62	Depositor EDS
% Data completeness (in resolution range)	86.5 (46.09-3.62) 98.6 (48.77-3.62)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.16	Depositor
$< I/\sigma(I) >$ [1]	1.06 (at 3.57Å)	Xtriage
Refinement program	PHENIX (dev_2006: ???)	Depositor
R, R_{free}	0.234 , 0.281 0.239 , 0.281	Depositor DCC
R_{free} test set	2002 reflections (9.23%)	wwPDB-VP
Wilson B-factor (Å^2)	156.6	Xtriage
Anisotropy	0.372	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 85.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	25381	wwPDB-VP
Average B, all atoms (Å^2)	153.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.01% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: PO4, 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	0.24	0/2597	0.40	0/3506
1	B	0.25	0/2604	0.41	0/3514
2	C	0.27	0/370	0.43	0/501
2	D	0.24	0/326	0.41	0/440
3	E	0.24	0/3440	0.41	0/4643
3	F	0.24	0/3449	0.41	0/4655
All	All	0.24	0/12786	0.41	0/17259

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

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	2534	2553	2553	31	0
1	B	2541	2558	2558	24	1
2	C	363	388	388	2	1
2	D	319	344	344	6	0
3	E	3377	3450	3442	29	0
3	F	3386	3455	3448	20	0
4	A	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	5	0	0	0	0
4	E	10	0	0	5	0
4	F	10	0	0	2	0
5	E	27	15	12	1	0
5	F	27	14	12	0	0
All	All	12604	12777	12757	100	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 4.

The worst 5 of 100 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:584:ARG:NH2	1:A:592:GLU:O	2.26	0.68
3:F:335:GLN:OE1	3:F:401:GLN:NE2	2.26	0.67
3:F:366:GLU:OE1	3:F:366:GLU:N	2.27	0.67
1:A:640:GLN:HE21	2:D:423:VAL:HG21	1.60	0.66
1:A:418:ILE:HG22	1:A:468:LEU:HD21	1.81	0.63

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:B:610:ASN:O	2:C:379:ILE:N[2_756]	2.10	0.10

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	312/317 (98%)	296 (95%)	16 (5%)	0	100	100
1	B	313/317 (99%)	296 (95%)	17 (5%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	43/50 (86%)	43 (100%)	0	0	100	100
2	D	37/50 (74%)	36 (97%)	1 (3%)	0	100	100
3	E	425/430 (99%)	411 (97%)	14 (3%)	0	100	100
3	F	426/430 (99%)	411 (96%)	15 (4%)	0	100	100
All	All	1556/1594 (98%)	1493 (96%)	63 (4%)	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	276/279 (99%)	271 (98%)	5 (2%)	59	81
1	B	277/279 (99%)	271 (98%)	6 (2%)	52	77
2	C	43/46 (94%)	43 (100%)	0	100	100
2	D	38/46 (83%)	38 (100%)	0	100	100
3	E	376/378 (100%)	368 (98%)	8 (2%)	53	78
3	F	377/378 (100%)	366 (97%)	11 (3%)	42	71
All	All	1387/1406 (99%)	1357 (98%)	30 (2%)	52	77

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	E	328	TYR
3	E	385	LYS
3	F	454	ARG
3	E	384	GLU
3	E	446	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	640	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

8 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	$\#\|Z\| > 2$
4	PO4	E	503	-	4,4,4	0.91	0	6,6,6	0.60	0
4	PO4	F	503	-	4,4,4	0.91	0	6,6,6	0.38	0
5	ADP	E	501	-	24,29,29	0.65	0	29,45,45	2.41	5 (17%)
5	ADP	F	501	-	24,29,29	0.66	0	29,45,45	2.49	4 (13%)
4	PO4	E	502	-	4,4,4	0.89	0	6,6,6	0.53	0
4	PO4	A	701	-	4,4,4	0.92	0	6,6,6	0.45	0
4	PO4	B	701	-	4,4,4	0.90	0	6,6,6	0.44	0
4	PO4	F	502	-	4,4,4	0.90	0	6,6,6	0.44	0

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
5	ADP	F	501	-	-	2/12/32/32	0/3/3/3
5	ADP	E	501	-	-	4/12/32/32	0/3/3/3

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	F	501	ADP	O5'-PA-O1A	-9.22	73.05	109.07
5	E	501	ADP	O5'-PA-O1A	-8.93	74.18	109.07
5	F	501	ADP	O2A-PA-O1A	-8.72	69.14	112.24
5	E	501	ADP	O2A-PA-O1A	-8.35	70.94	112.24
5	F	501	ADP	O2A-PA-O5'	2.62	119.93	107.75

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	F	501	ADP	C5'-O5'-PA-O3A
5	E	501	ADP	C3'-C4'-C5'-O5'
5	E	501	ADP	O4'-C4'-C5'-O5'
5	E	501	ADP	C5'-O5'-PA-O1A
5	F	501	ADP	C5'-O5'-PA-O1A

There are no ring outliers.

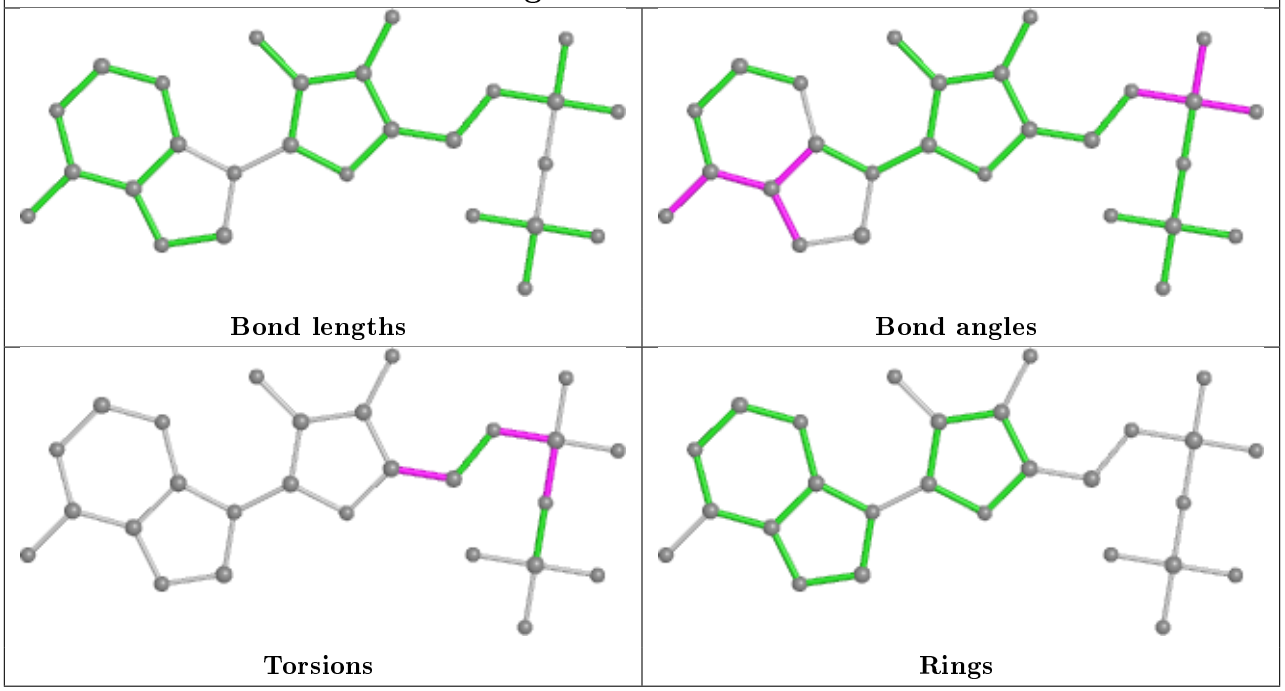
5 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	E	503	PO4	3	0
4	F	503	PO4	1	0
5	E	501	ADP	1	0
4	E	502	PO4	2	0
4	F	502	PO4	1	0

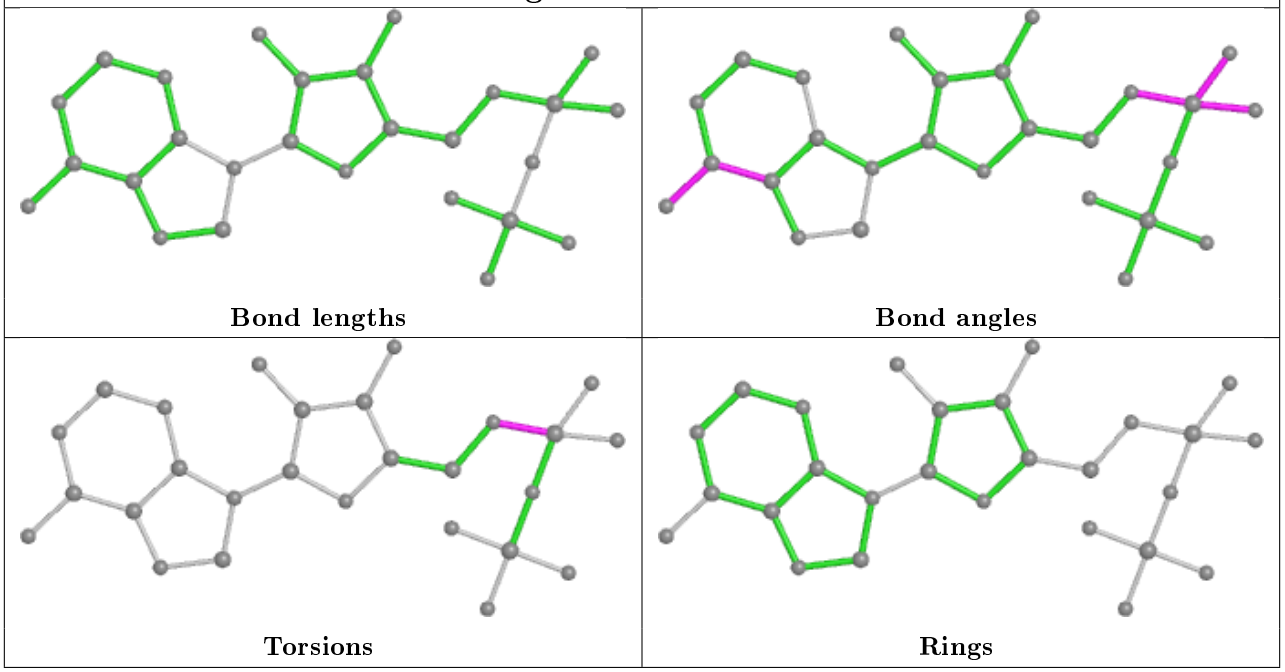
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. 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 E 501



Ligand ADP F 501



5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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, 95^{th} 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	314/317 (99%)	0.21	10 (3%) 47 32	107, 149, 181, 189	0
1	B	315/317 (99%)	0.02	2 (0%) 89 81	72, 118, 166, 188	0
2	C	45/50 (90%)	0.01	2 (4%) 34 22	94, 129, 156, 169	0
2	D	39/50 (78%)	0.41	3 (7%) 13 8	162, 180, 195, 211	0
3	E	425/430 (98%)	0.38	27 (6%) 19 11	83, 136, 169, 213	0
3	F	426/430 (99%)	0.50	47 (11%) 5 3	83, 145, 184, 218	0
All	All	1564/1594 (98%)	0.30	91 (5%) 23 14	72, 139, 180, 218	0

The worst 5 of 91 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	E	295	ILE	4.7
3	F	121	ASN	4.7
2	C	384	ASN	4.1
1	A	615	SER	4.1
3	E	296	LYS	3.9

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

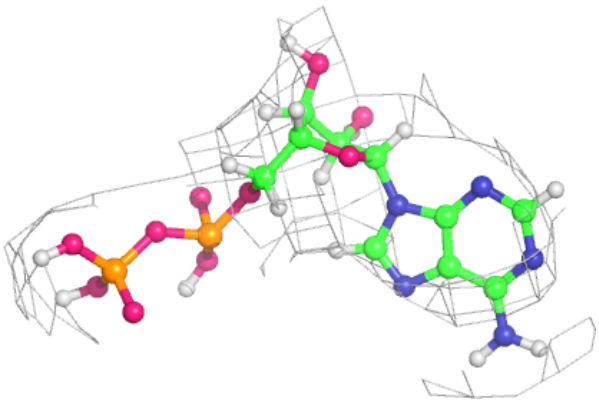
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

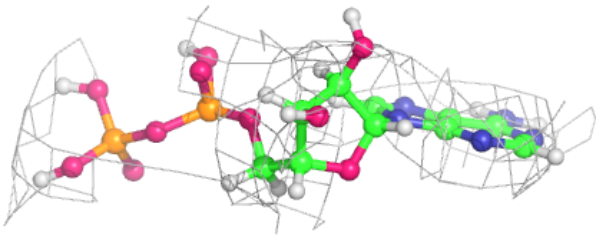
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	PO4	A	701	5/5	0.56	0.41	159,170,184,218	0
4	PO4	E	502	5/5	0.71	0.21	164,164,171,183	0
4	PO4	F	503	5/5	0.77	0.12	139,151,156,158	0
4	PO4	B	701	5/5	0.82	0.21	116,122,136,140	0
5	ADP	E	501	27/27	0.85	0.22	92,109,143,154	0
5	ADP	F	501	27/27	0.86	0.20	117,138,170,188	0
4	PO4	F	502	5/5	0.88	0.18	164,166,171,178	0
4	PO4	E	503	5/5	0.95	0.10	139,146,152,155	0

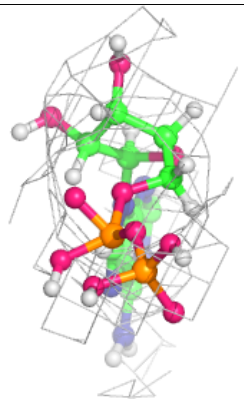
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around ADP E 501:

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

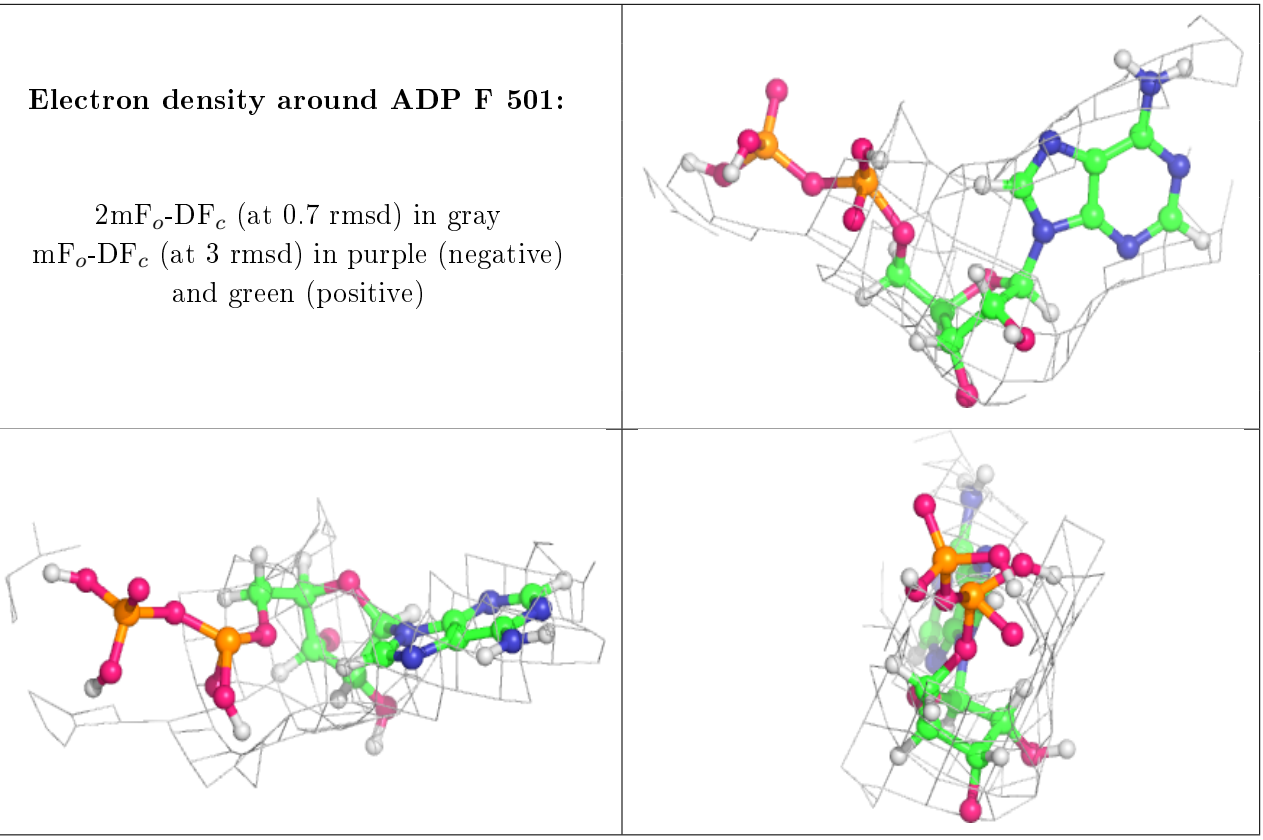






Electron density around ADP F 501:

$2mF_o$-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

There are no such residues in this entry.