



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 03:39 PM GMT

PDB ID : 4CYJ
Title : Chaetomium thermophilum Pan2:Pan3 complex
Authors : Wolf, J.; Valkov, E.; Allen, M.D.; Meineke, B.; Gordiyenko, Y.; McLaughlin, S.H.; Olsen, T.M.; Robinson, C.V.; Bycroft, M.; Stewart, M.; Passmore, L.A.
Deposited on : 2014-04-11
Resolution : 2.59 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 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
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) : trunk26865

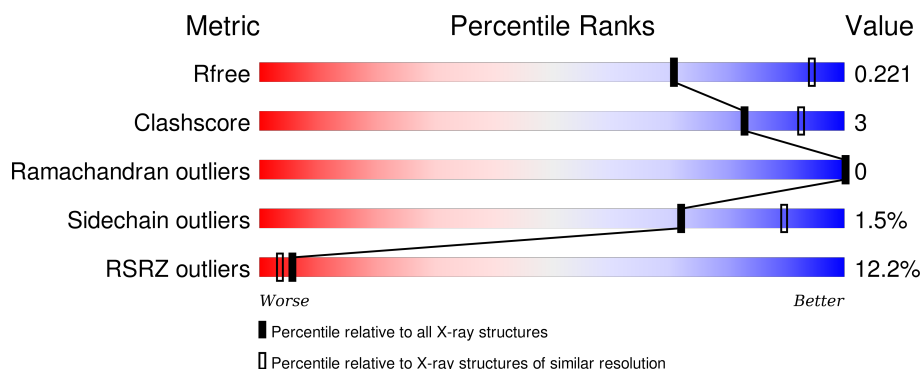
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.59 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}	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	438	<div> <div>11%</div> <div>85%</div> <div>7%</div> <div>8%</div> </div>
1	B	438	<div> <div>11%</div> <div>83%</div> <div>11%</div> <div>6%</div> </div>
1	C	438	<div> <div>7%</div> <div>88%</div> <div>•</div> <div>8%</div> </div>
1	D	438	<div> <div>11%</div> <div>87%</div> <div>6%</div> <div>• 6%</div> </div>
2	E	119	<div> <div>22%</div> <div>38%</div> <div>6%</div> <div>56%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	119	

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
3	ATP	B	900	-	-	-	X

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 28335 atoms, of which 14053 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 PAB-DEPENDENT POLY(A)-SPECIFIC RIBONUCLEASE SUBUNIT PAN3-LIKE PROTEIN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	403	Total	C	H	N	O	S	0	0	0
			6534	2074	3273	576	597	14			
1	B	412	Total	C	H	N	O	S	0	0	0
			6652	2120	3323	583	612	14			
1	C	404	Total	C	H	N	O	S	0	0	0
			6543	2075	3281	577	596	14			
1	D	411	Total	C	H	N	O	S	0	0	0
			6622	2109	3309	579	611	14			

- Molecule 2 is a protein called PAN2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	52	Total	C	H	N	O	0	0	0
			812	262	402	71	77			
2	F	54	Total	C	H	N	O	0	0	0
			850	272	421	77	80			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	340	GLY	-	EXPRESSION TAG	UNP G0SAK8
E	341	SER	-	EXPRESSION TAG	UNP G0SAK8
E	342	MET	-	EXPRESSION TAG	UNP G0SAK8
F	340	GLY	-	EXPRESSION TAG	UNP G0SAK8
F	341	SER	-	EXPRESSION TAG	UNP G0SAK8
F	342	MET	-	EXPRESSION TAG	UNP G0SAK8

- Molecule 3 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: C₁₀H₁₆N₅O₁₃P₃).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	C	H	N	O	P	0	0
			42	10	11	5	13	3		
3	B	1	Total	C	H	N	O	P	0	0
			42	10	11	5	13	3		
3	C	1	Total	C	H	N	O	P	0	0
			42	10	11	5	13	3		
3	D	1	Total	C	H	N	O	P	0	0
			42	10	11	5	13	3		

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		
4	A	1	Total	Mg	0	0
			1	1		
4	D	1	Total	Mg	0	0
			1	1		
4	C	1	Total	Mg	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	41	Total	O	0	0
			41	41		

Continued on next page...

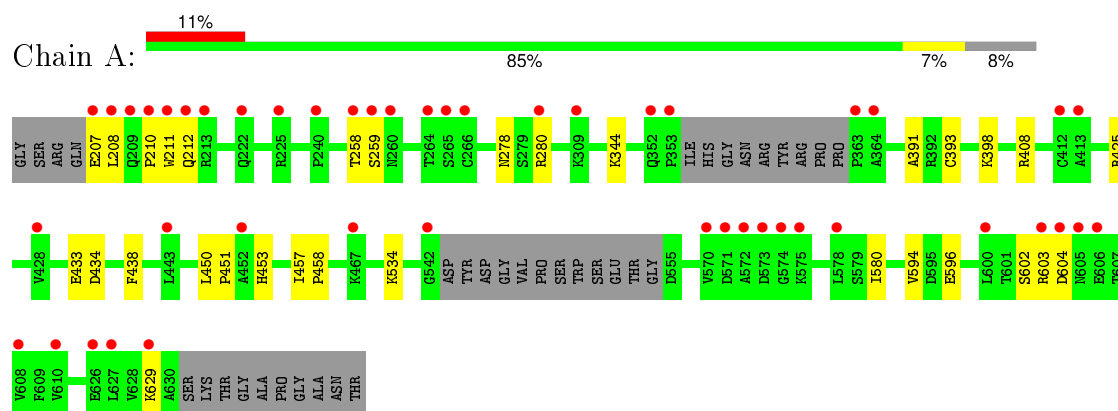
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	35	Total 35	O 35	0	0
5	C	42	Total 42	O 42	0	0
5	D	30	Total 30	O 30	0	0
5	F	2	Total 2	O 2	0	0

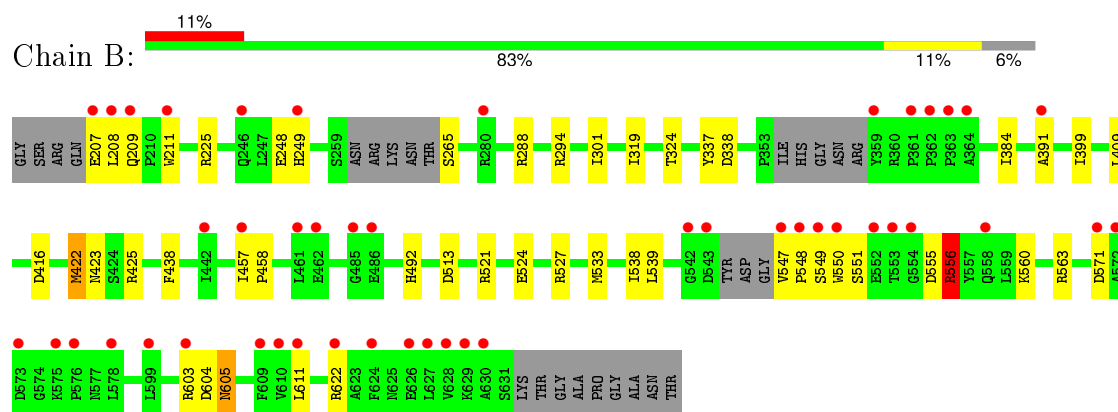
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

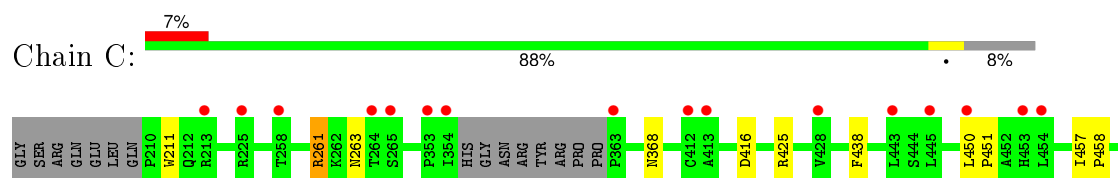
• Molecule 1: PAB-DEPENDENT POLY(A)-SPECIFIC RIBONUCLEASE SUBUNIT PAN3-LIKE PROTEIN

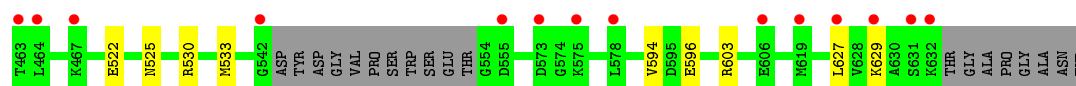


• Molecule 1: PAB-DEPENDENT POLY(A)-SPECIFIC RIBONUCLEASE SUBUNIT PAN3-LIKE PROTEIN

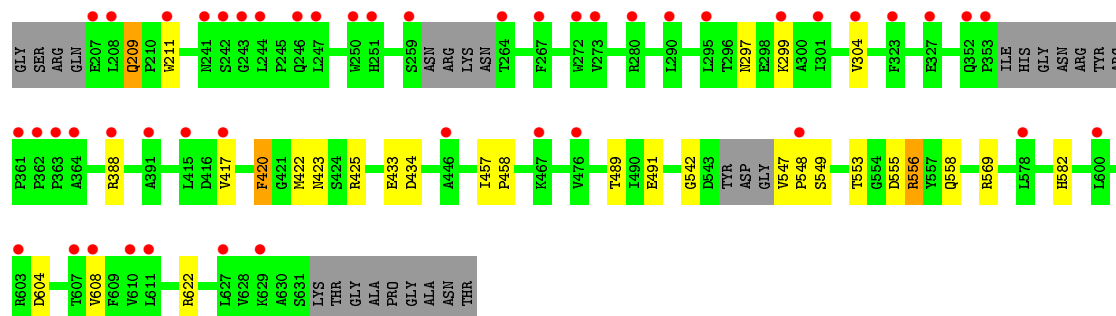
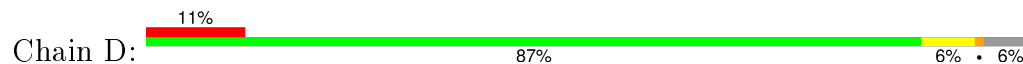


• Molecule 1: PAB-DEPENDENT POLY(A)-SPECIFIC RIBONUCLEASE SUBUNIT PAN3-LIKE PROTEIN

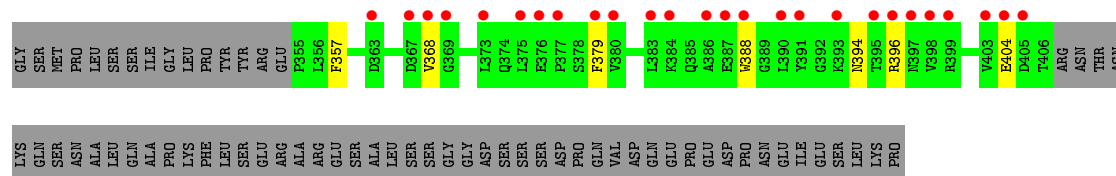
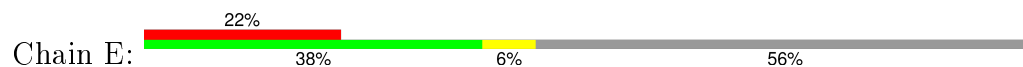




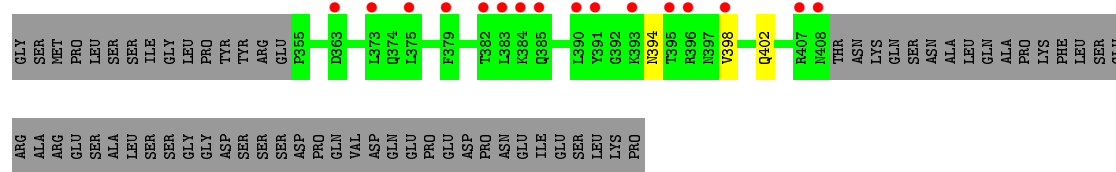
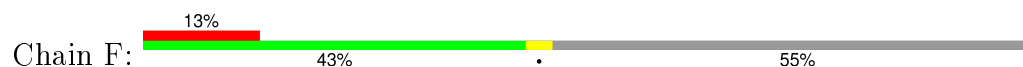
• Molecule 1: PAB-DEPENDENT POLY(A)-SPECIFIC RIBONUCLEASE SUBUNIT PAN3-LIKE PROTEIN



• Molecule 2: PAN2



• Molecule 2: PAN2



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	89.09Å 145.36Å 101.81Å 90.00° 94.86° 90.00°	Depositor
Resolution (Å)	47.89 – 2.59 47.89 – 2.59	Depositor EDS
% Data completeness (in resolution range)	99.8 (47.89-2.59) 99.8 (47.89-2.59)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.79 (at 2.58Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.183 , 0.221 0.180 , 0.221	Depositor DCC
R_{free} test set	4021 reflections (5.02%)	DCC
Wilson B-factor (Å ²)	55.0	Xtriage
Anisotropy	0.250	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.41 , 52.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 80129 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	28335	wwPDB-VP
Average B, all atoms (Å ²)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.74% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: MG, 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	0.33	0/3329	0.57	0/4501
1	B	0.34	0/3402	0.59	2/4605 (0.0%)
1	C	0.35	0/3330	0.56	0/4500
1	D	0.33	0/3385	0.53	0/4582
2	E	0.29	0/421	0.51	0/575
2	F	0.31	0/440	0.56	0/600
All	All	0.33	0/14307	0.56	2/19363 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	556	ARG	NE-CZ-NH1	5.89	123.25	120.30
1	B	571	ASP	CB-CG-OD2	5.17	122.95	118.30

There are no chirality outliers.

There are no planarity outliers.

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	3261	3273	3260	19	0
1	B	3329	3323	3310	34	2
1	C	3262	3281	3268	12	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	3313	3309	3296	23	2
2	E	410	402	402	6	0
2	F	429	421	421	2	0
3	A	31	11	12	1	0
3	B	31	11	12	2	0
3	C	31	11	12	0	0
3	D	31	11	12	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
5	A	41	0	0	2	0
5	B	35	0	0	4	0
5	C	42	0	0	0	0
5	D	30	0	0	2	0
5	F	2	0	0	0	0
All	All	14282	14053	14005	88	2

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 3.

The worst 5 of 88 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:D:547:VAL:N	5:D:2026:HOH:O	1.96	0.97
1:D:425:ARG:NH1	1:D:433:GLU:OE2	2.11	0.83
1:D:553:THR:O	1:D:556:ARG:NH1	2.13	0.82
1:D:297:ASN:OD1	1:D:299:LYS:N	2.18	0.75
1:D:555:ASP:OD2	1:D:622:ARG:NH1	2.20	0.74

All (2) 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:422:MET:CE	1:D:388:ARG:NH2[2_658]	2.05	0.15
1:B:422:MET:SD	1:D:388:ARG:NH2[2_658]	2.16	0.04

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	397/438 (91%)	395 (100%)	2 (0%)	0	100	100
1	B	404/438 (92%)	401 (99%)	3 (1%)	0	100	100
1	C	398/438 (91%)	394 (99%)	4 (1%)	0	100	100
1	D	403/438 (92%)	402 (100%)	1 (0%)	0	100	100
2	E	50/119 (42%)	50 (100%)	0	0	100	100
2	F	52/119 (44%)	52 (100%)	0	0	100	100
All	All	1704/1990 (86%)	1694 (99%)	10 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	358/385 (93%)	355 (99%)	3 (1%)	86	95
1	B	366/385 (95%)	357 (98%)	9 (2%)	55	81
1	C	358/385 (93%)	353 (99%)	5 (1%)	74	90
1	D	365/385 (95%)	359 (98%)	6 (2%)	70	89
2	E	44/103 (43%)	44 (100%)	0	100	100
2	F	46/103 (45%)	46 (100%)	0	100	100
All	All	1537/1746 (88%)	1514 (98%)	23 (2%)	72	90

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	556	ARG
1	C	261	ARG
1	D	558	GLN
1	B	605	ASN
1	C	416	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	249	HIS
1	B	605	ASN
1	C	249	HIS
1	C	349	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](#)

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 for Mogul analysis.

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	900	4	24,33,33	0.88	1 (4%)	31,52,52	2.01	7 (22%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	ATP	B	900	4	24,33,33	0.91	1 (4%)	31,52,52	2.07	6 (19%)
3	ATP	C	900	4	24,33,33	0.91	1 (4%)	31,52,52	1.75	6 (19%)
3	ATP	D	900	4	24,33,33	0.88	1 (4%)	31,52,52	1.87	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	900	4	-	0/18/38/38	0/3/3/3
3	ATP	B	900	4	-	0/18/38/38	0/3/3/3
3	ATP	C	900	4	-	0/18/38/38	0/3/3/3
3	ATP	D	900	4	-	0/18/38/38	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	900	ATP	C5-C4	2.65	1.46	1.40
3	C	900	ATP	C5-C4	2.67	1.46	1.40
3	A	900	ATP	C5-C4	2.81	1.46	1.40
3	B	900	ATP	C5-C4	2.90	1.47	1.40

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	900	ATP	N3-C2-N1	-7.79	122.93	128.89
3	A	900	ATP	N3-C2-N1	-7.46	123.18	128.89
3	D	900	ATP	N3-C2-N1	-7.42	123.21	128.89
3	C	900	ATP	N3-C2-N1	-6.60	123.84	128.89
3	B	900	ATP	C2'-C1'-N9	-4.19	107.89	114.29

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	900	ATP	1	0
3	B	900	ATP	2	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	403/438 (92%)	0.89	46 (11%) 7 4	33, 58, 121, 157	0
1	B	412/438 (94%)	0.84	47 (11%) 7 4	33, 62, 116, 140	0
1	C	404/438 (92%)	0.75	30 (7%) 17 12	35, 56, 109, 125	0
1	D	411/438 (93%)	0.97	47 (11%) 7 4	34, 65, 112, 138	0
2	E	52/119 (43%)	2.03	26 (50%) 0 0	67, 105, 128, 137	0
2	F	54/119 (45%)	1.44	16 (29%) 1 0	50, 92, 122, 127	0
All	All	1736/1990 (87%)	0.91	212 (12%) 5 3	33, 62, 118, 157	0

The worst 5 of 212 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	361	PRO	10.7
1	D	362	PRO	9.5
1	C	354	ILE	9.2
1	B	629	LYS	7.8
1	D	363	PRO	6.7

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(Å ²)	Q<0.9
3	ATP	B	900	31/31	0.79	0.31	2.82	51,100,164,182	0
3	ATP	C	900	31/31	0.95	0.23	0.50	40,63,125,132	0
3	ATP	D	900	31/31	0.88	0.20	-0.12	57,80,119,174	0
3	ATP	A	900	31/31	0.96	0.18	-1.19	34,55,116,144	0
4	MG	A	1000	1/1	0.73	0.42	-	71,71,71,71	0
4	MG	D	1000	1/1	0.96	0.06	-	76,76,76,76	0
4	MG	C	1000	1/1	0.54	0.33	-	75,75,75,75	0
4	MG	B	1000	1/1	0.71	0.29	-	87,87,87,87	0

6.5 Other polymers [i](#)

There are no such residues in this entry.