



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 03:04 PM GMT

PDB ID : 4BE7
Title : MUTANT (K220R) OF THE HSDR SUBUNIT OF THE ECOR124I RESTRICTION ENZYME IN COMPLEX WITH ATP
Authors : Csefalvay, E.; Lapkouski, M.; Guzanova, A.; Csefalvay, L.; Baikova, T.; Shevelev, I.; Janscak, P.; Smatanova, I.K.; Panjikar, S.; Carey, J.; Weiserova, M.; Ettrich, R.
Deposited on : 2013-03-06
Resolution : 2.74 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

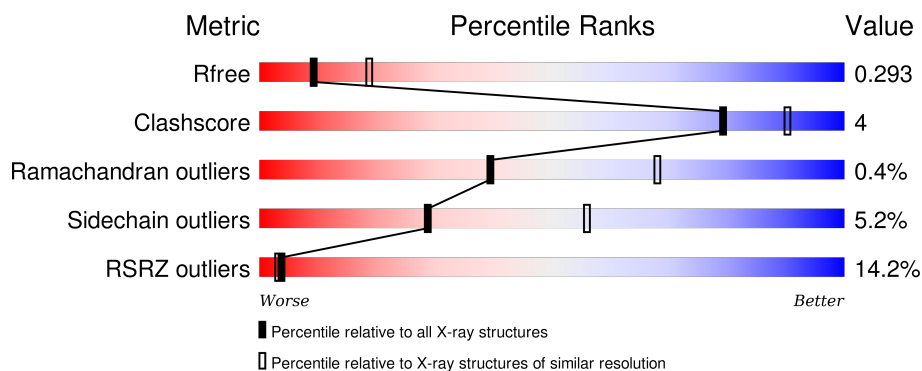
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.74 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	3050 (2.78-2.70)
Clashscore	102246	3424 (2.78-2.70)
Ramachandran outliers	100387	3367 (2.78-2.70)
Sidechain outliers	100360	3368 (2.78-2.70)
RSRZ outliers	91569	3055 (2.78-2.70)

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	B	1038	<div> <div>10%</div> <div>71%</div> <div>8%</div> <div>19%</div> </div>
1	D	1038	<div> <div>13%</div> <div>70%</div> <div>9%</div> <div>20%</div> </div>

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 13387 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.

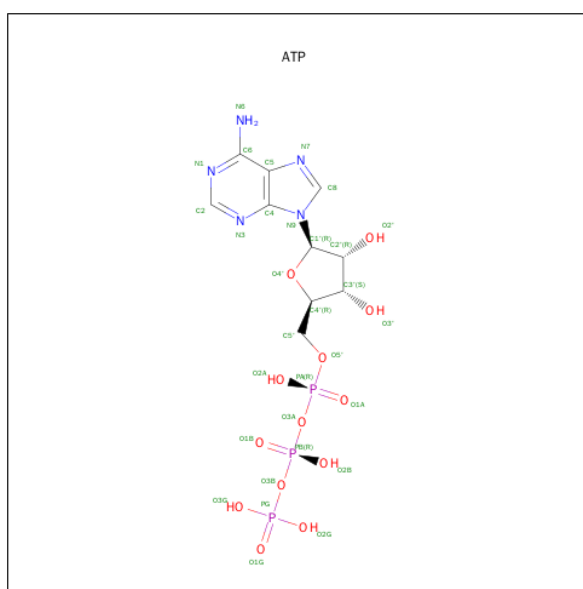
- Molecule 1 is a protein called TYPE I RESTRICTION ENZYME ECOR124II R PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	837	Total	C	N	O	S	0	0	0
			6631	4226	1112	1277	16			
1	D	835	Total	C	N	O	S	0	0	0
			6594	4204	1106	1268	16			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	220	ARG	LYS	ENGINEERED MUTATION	UNP Q304R3
D	220	ARG	LYS	ENGINEERED MUTATION	UNP Q304R3

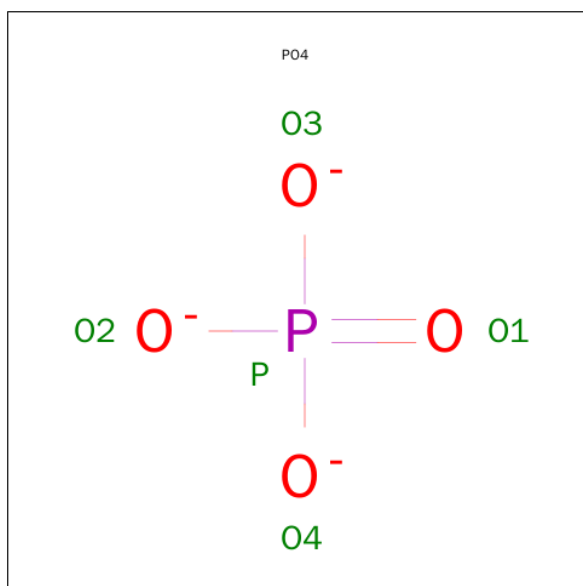
- Molecule 2 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: $C_{10}H_{16}N_5O_{13}P_3$).



Continued from previous page...

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

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O₄P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	O	P	0	0
			5	4	1		
3	B	1	Total	O	P	0	0
			5	4	1		
3	D	1	Total	O	P	0	0
			5	4	1		
3	D	1	Total	O	P	0	0
			5	4	1		

- 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	D	1	Total	Mg	0	0
			1	1		

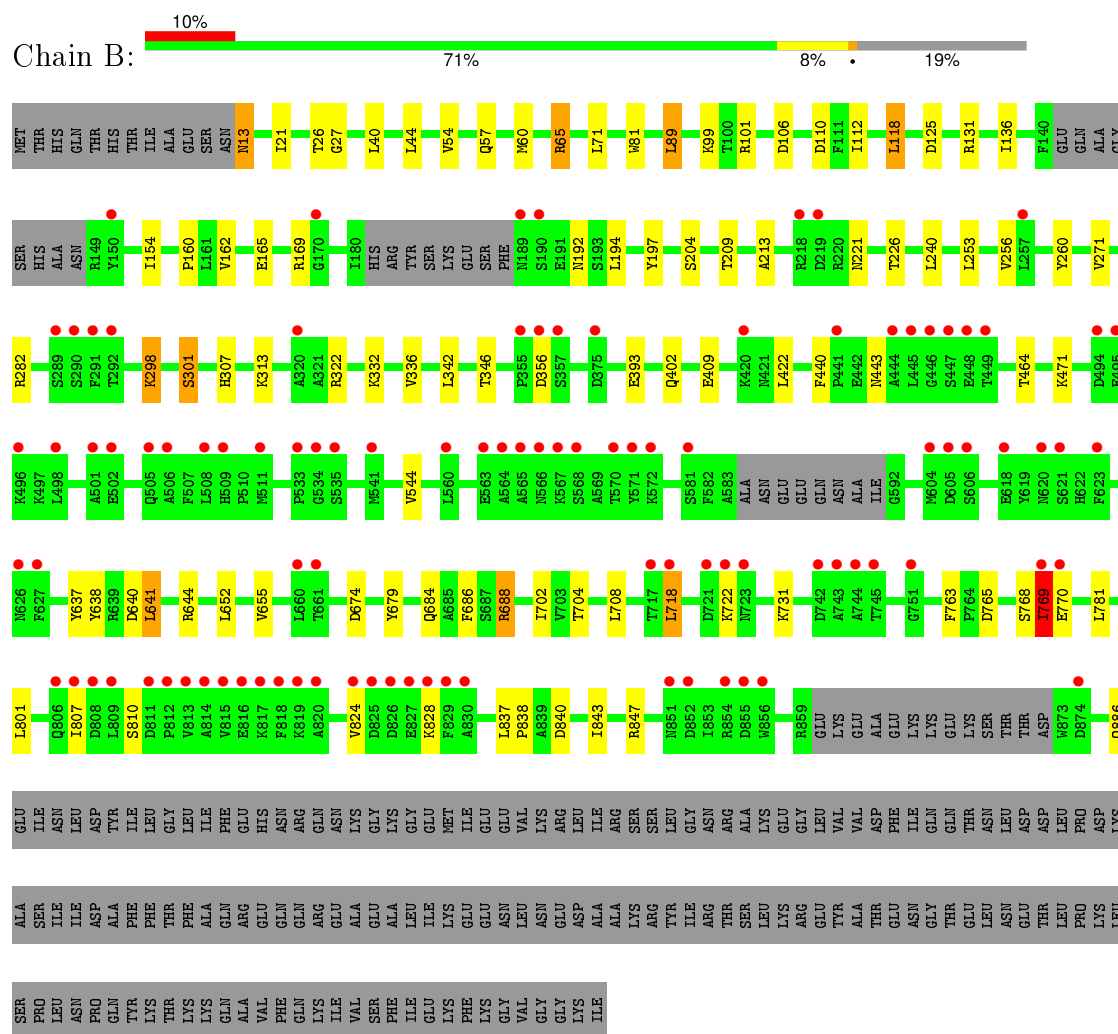
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	45	Total 45	O 45	0	0
5	D	33	Total 33	O 33	0	0

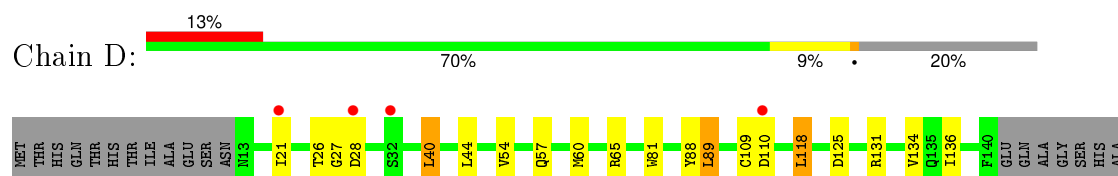
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: TYPE I RESTRICTION ENZYME ECOR124II R PROTEIN



• Molecule 1: TYPE I RESTRICTION ENZYME ECOR124II R PROTEIN



LYS	ILE	ARG	GLN	E821	S687	A569	T449	T309	ASN
ILE	VAL	GLU	ASN	E822	R688	T570	S452	K313	R149
SER	GLU	ALA	LYS	V824	I702	K571	S452	I154	I154
PHE	LEU	ALA	LYS	D826	R710	K572	D465	A321	L159
ILE	ILE	LEU	GLU	E827	S711	F582	T464	R322	P160
GLU	LYS	ILE	MET	K828	D714	A583	R468	T325	L161
LYS	LYS	LYS	ILE	L832	I717	ALA	K476	D328	V162
PHE	GLU	GLU	GLU	I717	L718	ASN	V477	F329	I164
LYS	LYS	ASN	GLU	L718	D721	GLU	D478	V336	R169
VAL	LEU	LEU	VAL	D721	K722	GLN	Y479	Y334	G170
ASN	ASN	ARG	LYS	A839	N723	ASN	L489	E490	A176
GLU	GLU	ILE	LEU	I843	K731	ALA	E490	S354	I180
LYS	ALA	ALA	ARG	I850	E760	G592	T493	P355	HIS
LYS	SER	SER	SER	I851	F763	E597	D494	D356	ARG
ARG	LEU	ARG	SER	D852	P764	T598	E495	S357	TYR
ILE	TYR	ILE	LEU	A743	A744	TS98	K496	T384	SER
ARG	ARG	ARG	GLY	T745	T746	T601	K497	A385	LYS
THR	THR	THR	ASN	T746	E747	S602	L498	R374	GLU
SER	SER	SER	ALA	D855	E747	A603	S499	D375	SER
LEU	LEU	LEU	LYS	W856	E760	M604	A500	E501	GLU
LYS	LYS	LYS	GLU	E859	E760	D605	E502	E502	PHE
ARG	ARG	ARG	GLY	GLU	GLU	S606	Q805	I378	N189
LYS	GLU	GLU	LEU	LYS	LYS	S607	A506	N387	S190
VAL	VAL	VAL	VAL	ALA	ALA	A608	F507	N388	E191
ALA	ALA	ALA	ASP	ALA	ALA	A614	L508	E393	N192
THR	THR	THR	PHE	GLU	GLU	E618	M511	S394	T203
GLU	GLU	GLU	ILE	LYS	LYS	S621	E515	Q402	S204
ASN	ASN	ASN	GLN	LYS	LYS	H622	T528	I406	T209
LEU	LEU	LEU	THR	THR	THR	F623	R530	E409	A213
ASN	ASN	ASN	THR	THR	THR	K624	T531	O410	K217
GLU	GLU	GLU	ASP	ASP	ASP	T625	F832	G414	D224
TRP	TRP	TRP	TRP	TRP	TRP	N626	G534	G416	L240
D874	D874	D874	D874	D874	D874	Y637	K536	N421	L253
S885	S885	S885	S885	S885	S885	L641	M541	L422	V256
GLN	GLN	GLN	GLN	GLN	GLN	N647	R559	K425	Y260
ILE	ILE	ILE	ILE	ILE	ILE	Q648	L560	F440	V271
LEU	LEU	LEU	LEU	LEU	LEU	D649	Q561	P441	W285
ASP	ASP	ASP	ASP	ASP	ASP	I650	E562	E442	A293
TYR	TYR	TYR	ALA	ALA	ALA	D651	E563	N443	K294
ILE	ILE	ILE	PHE	PHE	PHE	L652	A564	A444	W295
PHE	PHE	PHE	GLN	GLN	GLN	V655	A565	L445	H307
LYS	LYS	LYS	THR	THR	THR	R678	M566	G446	T308
LYS	LYS	LYS	THR	THR	THR	Y679	S447	E448	
LYS	LYS	LYS	ALA	ALA	ALA	Q684			
GLN	GLN	GLN	ARG	ARG	ARG	A685			
ALA	ALA	ALA	GLU	GLU	GLU	F686			
VAL	VAL	VAL	GLU	GLU	GLU				
PHE	PHE	PHE	ASN	ASN	ASN				
GLN	GLN	GLN	GLN	GLN	GLN				
ARG	ARG	ARG	GLN	GLN	GLN				
VAL	VAL	VAL	GLN	GLN	GLN				
PHE	PHE	PHE	GLN	GLN	GLN				
GLN	GLN	GLN	GLN	GLN	GLN				

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	87.05Å 124.35Å 128.01Å 90.00° 108.86° 90.00°	Depositor
Resolution (Å)	32.36 – 2.74 33.73 – 2.74	Depositor EDS
% Data completeness (in resolution range)	93.2 (32.36-2.74) 93.5 (33.73-2.74)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.83 (at 2.76Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.249 , 0.292 0.249 , 0.293	Depositor DCC
R_{free} test set	3192 reflections (5.07%)	DCC
Wilson B-factor (Å ²)	42.1	Xtriage
Anisotropy	0.071	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 60.6	EDS
Estimated twinning fraction	0.018 for h,-k,-h-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 63146 reflections	Xtriage
F_o, F_c correlation	0.87	EDS
Total number of atoms	13387	wwPDB-VP
Average B, all atoms (Å ²)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.14% 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, PO4, 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	B	0.27	0/6760	0.52	1/9146 (0.0%)
1	D	0.27	0/6721	0.53	3/9097 (0.0%)
All	All	0.27	0/13481	0.53	4/18243 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	441	PRO	C-N-CA	6.15	137.08	121.70
1	D	441	PRO	CA-C-N	5.29	128.83	117.20
1	B	769	ILE	N-CA-C	-5.26	96.79	111.00
1	D	826	ASP	C-N-CA	5.03	134.27	121.70

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	B	6631	0	6305	44	0
1	D	6594	0	6268	47	0
2	B	31	0	12	1	0
2	D	31	0	12	1	0
3	B	10	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	10	0	0	0	0
4	B	1	0	0	0	0
4	D	1	0	0	0	0
5	B	45	0	0	1	0
5	D	33	0	0	1	0
All	All	13387	0	12597	91	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 4.

The worst 5 of 91 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:824:VAL:HG13	1:D:828:LYS:HB3	1.54	0.90
1:D:27:GLY:HA2	1:D:28:ASP:HB2	1.59	0.85
1:D:169:ARG:HG3	1:D:169:ARG:HH11	1.52	0.73
1:D:465:ASP:OD1	1:D:468:ARG:NH2	2.22	0.71
1:D:322:ARG:NH2	5:D:2021:HOH:O	2.25	0.69

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	B	827/1038 (80%)	798 (96%)	26 (3%)	3 (0%)	39	68
1	D	825/1038 (80%)	785 (95%)	37 (4%)	3 (0%)	39	68
All	All	1652/2076 (80%)	1583 (96%)	63 (4%)	6 (0%)	39	68

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	810	SER
1	B	807	ILE
1	B	810	SER
1	D	807	ILE
1	B	769	ILE

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	B	690/927 (74%)	656 (95%)	34 (5%)	31	59
1	D	684/927 (74%)	647 (95%)	37 (5%)	27	54
All	All	1374/1854 (74%)	1303 (95%)	71 (5%)	29	56

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	781	LEU
1	D	134	VAL
1	D	718	LEU
1	B	840	ASP
1	D	40	LEU

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

Mol	Chain	Res	Type
1	B	91	ASN

5.3.3 RNA ⓘ

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, 2 are monoatomic - leaving 6 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$
2	ATP	B	1887	4	24,33,33	0.94	1 (4%)	31,52,52	1.91	5 (16%)
3	PO4	B	1888	-	4,4,4	0.45	0	6,6,6	0.27	0
3	PO4	B	1889	-	4,4,4	0.47	0	6,6,6	0.27	0
2	ATP	D	1887	4	24,33,33	0.94	1 (4%)	31,52,52	1.93	5 (16%)
3	PO4	D	1889	-	4,4,4	0.43	0	6,6,6	0.27	0
3	PO4	D	1890	-	4,4,4	0.48	0	6,6,6	0.27	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
2	ATP	B	1887	4	-	0/18/38/38	0/3/3/3
3	PO4	B	1888	-	-	0/0/0/0	0/0/0/0
3	PO4	B	1889	-	-	0/0/0/0	0/0/0/0
2	ATP	D	1887	4	-	0/18/38/38	0/3/3/3
3	PO4	D	1889	-	-	0/0/0/0	0/0/0/0
3	PO4	D	1890	-	-	0/0/0/0	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	1887	ATP	C5-C4	2.99	1.47	1.40
2	B	1887	ATP	C5-C4	2.99	1.47	1.40

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	1887	ATP	N3-C2-N1	-6.84	123.66	128.89
2	B	1887	ATP	N3-C2-N1	-6.60	123.84	128.89
2	B	1887	ATP	PA-O3A-PB	-4.43	120.28	132.73
2	D	1887	ATP	PA-O3A-PB	-4.21	120.91	132.73
2	D	1887	ATP	C2'-C1'-N9	-3.39	109.12	114.29

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	1887	ATP	1	0
2	D	1887	ATP	1	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	B	837/1038 (80%)	0.83	100 (11%) 6 5	13, 33, 71, 121	0
1	D	835/1038 (80%)	0.98	137 (16%) 2 2	17, 40, 84, 114	0
All	All	1672/2076 (80%)	0.90	237 (14%) 4 3	13, 36, 78, 121	0

The worst 5 of 237 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	809	LEU	14.3
1	B	808	ASP	8.4
1	B	807	ILE	8.0
1	B	813	VAL	8.0
1	D	604	MET	7.8

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
3	PO4	D	1889	5/5	0.86	0.17	-0.41	36,53,55,82	0
3	PO4	B	1889	5/5	0.90	0.16	-0.61	52,55,69,83	0
3	PO4	D	1890	5/5	0.98	0.14	-1.03	16,17,21,23	0
2	ATP	D	1887	31/31	0.97	0.17	-1.09	13,21,27,33	0
2	ATP	B	1887	31/31	0.97	0.16	-1.33	11,18,23,25	0
3	PO4	B	1888	5/5	0.98	0.13	-1.37	24,24,27,32	0
4	MG	B	1890	1/1	0.92	0.34	-	23,23,23,23	0
4	MG	D	1886	1/1	0.87	0.27	-	25,25,25,25	0

6.5 Other polymers [i](#)

There are no such residues in this entry.