W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 04:25 PM GMT

PDB ID	:	4ETP
Title	:	C-terminal motor and motor homology domain of Kar3Vik1 fused to a synthetic heterodimeric coiled coil
Authors	:	Rank, K.C.; Chen, C.J.; Cope, J.; Porche, K.; Hoenger, A.; Gilbert, S.P.; Rayment, I.
Deposited on	:	2012-04-24
Resolution	:	2.30 Å(reported)

The following versions of software and data (see references (1)) 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 2.30 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		0.246
Clashscore		4
Ramachandran outliers		0.3%
Sidechain outliers		1.8%
RSRZ outliers		22.2%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
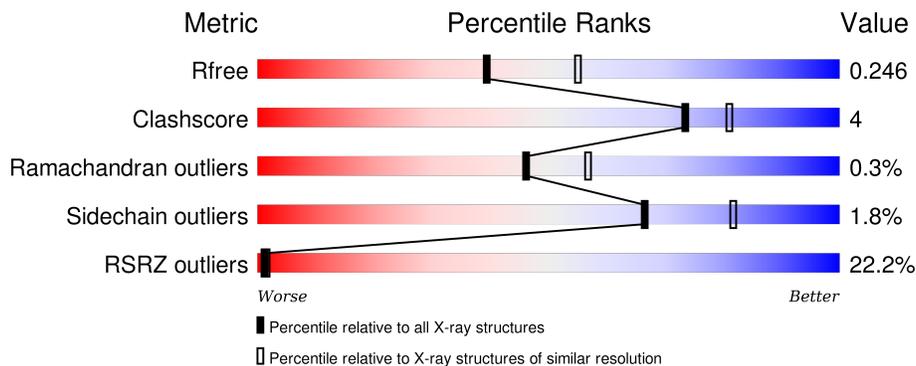
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	91344	3852 (2.30-2.30)
Clashscore	102246	4452 (2.30-2.30)
Ramachandran outliers	100387	4410 (2.30-2.30)
Sidechain outliers	100360	4409 (2.30-2.30)
RSRZ outliers	91569	3857 (2.30-2.30)

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 <=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	403	12% — 84% 10% 6%
2	B	333	30% — 77% 9% 14%

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
5	GOL	A	804	-	-	X	X
5	GOL	A	805	-	-	X	X
5	GOL	A	806	-	-	-	X
6	EBC	B	701	-	-	-	X

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 5681 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 Kinesin-like protein KAR3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	379	3010	1885	530	590	5	0	3	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	327	GLY	-	SEE REMARK 999	UNP P17119
A	328	ALA	-	SEE REMARK 999	UNP P17119
A	329	SER	-	SEE REMARK 999	UNP P17119
A	330	LYS	-	SEE REMARK 999	UNP P17119
A	331	ILE	-	SEE REMARK 999	UNP P17119
A	332	ALA	-	SEE REMARK 999	UNP P17119
A	333	ALA	-	SEE REMARK 999	UNP P17119
A	334	LEU	-	SEE REMARK 999	UNP P17119
A	335	LYS	-	SEE REMARK 999	UNP P17119
A	336	GLU	-	SEE REMARK 999	UNP P17119
A	337	LYS	-	SEE REMARK 999	UNP P17119
A	338	ILE	-	SEE REMARK 999	UNP P17119
A	339	ALA	-	SEE REMARK 999	UNP P17119
A	340	ALA	-	SEE REMARK 999	UNP P17119
A	341	LEU	-	SEE REMARK 999	UNP P17119
A	342	LYS	-	SEE REMARK 999	UNP P17119
A	343	GLU	-	SEE REMARK 999	UNP P17119
A	344	LYS	-	SEE REMARK 999	UNP P17119
A	345	ILE	-	SEE REMARK 999	UNP P17119
A	346	ALA	-	SEE REMARK 999	UNP P17119
A	347	ALA	-	SEE REMARK 999	UNP P17119
A	348	LEU	-	SEE REMARK 999	UNP P17119
A	349	LYS	-	SEE REMARK 999	UNP P17119
A	350	GLU	-	SEE REMARK 999	UNP P17119
A	351	LYS	-	SEE REMARK 999	UNP P17119
A	391	LEU	CYS	ENGINEERED MUTATION	UNP P17119
A	469	ALA	CYS	ENGINEERED MUTATION	UNP P17119

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	517	ALA	CYS	ENGINEERED MUTATION	UNP P17119
A	566	VAL	CYS	ENGINEERED MUTATION	UNP P17119
A	655	ALA	CYS	ENGINEERED MUTATION	UNP P17119

- Molecule 2 is a protein called Spindle pole body-associated protein VIK1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	287	2351	1513	375	456	7	0	0	0

There are 33 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	315	GLY	-	SEE REMARK 999	UNP Q12045
B	316	ALA	-	SEE REMARK 999	UNP Q12045
B	317	SER	-	SEE REMARK 999	UNP Q12045
B	318	GLU	-	SEE REMARK 999	UNP Q12045
B	319	ILE	-	SEE REMARK 999	UNP Q12045
B	320	ALA	-	SEE REMARK 999	UNP Q12045
B	321	ALA	-	SEE REMARK 999	UNP Q12045
B	322	LEU	-	SEE REMARK 999	UNP Q12045
B	323	GLU	-	SEE REMARK 999	UNP Q12045
B	324	LYS	-	SEE REMARK 999	UNP Q12045
B	325	GLU	-	SEE REMARK 999	UNP Q12045
B	326	ILE	-	SEE REMARK 999	UNP Q12045
B	327	ALA	-	SEE REMARK 999	UNP Q12045
B	328	ALA	-	SEE REMARK 999	UNP Q12045
B	329	LEU	-	SEE REMARK 999	UNP Q12045
B	330	GLU	-	SEE REMARK 999	UNP Q12045
B	331	LYS	-	SEE REMARK 999	UNP Q12045
B	332	GLU	-	SEE REMARK 999	UNP Q12045
B	333	ILE	-	SEE REMARK 999	UNP Q12045
B	334	ALA	-	SEE REMARK 999	UNP Q12045
B	335	ALA	-	SEE REMARK 999	UNP Q12045
B	336	LEU	-	SEE REMARK 999	UNP Q12045
B	337	GLU	-	SEE REMARK 999	UNP Q12045
B	338	LYS	-	SEE REMARK 999	UNP Q12045
B	339	GLU	-	SEE REMARK 999	UNP Q12045
B	340	ILE	-	SEE REMARK 999	UNP Q12045
B	355	CYS	GLU	ENGINEERED MUTATION	UNP Q12045
B	377	VAL	CYS	ENGINEERED MUTATION	UNP Q12045
B	423	CYS	LYS	ENGINEERED MUTATION	UNP Q12045

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	436	ALA	CYS	ENGINEERED MUTATION	UNP Q12045
B	536	ALA	CYS	ENGINEERED MUTATION	UNP Q12045
B	596	ALA	CYS	ENGINEERED MUTATION	UNP Q12045
B	640	ALA	CYS	ENGINEERED MUTATION	UNP Q12045

- Molecule 3 is ADENOSINE-5'-DIPHOSPHATE (three-letter code: ADP) (formula: $C_{10}H_{15}N_5O_{10}P_2$).
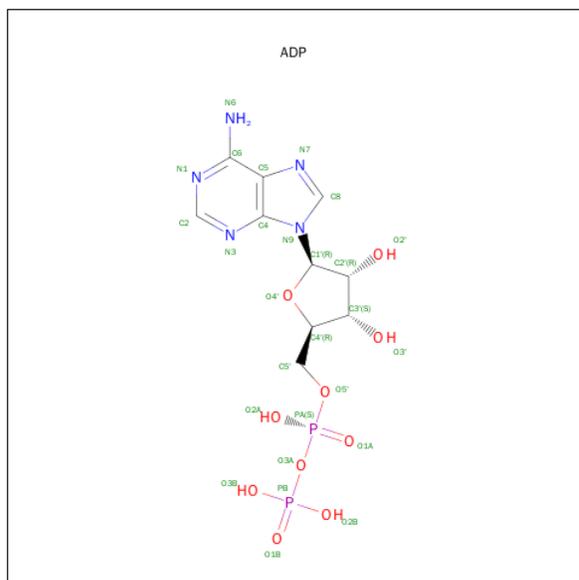
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
3	A	1	27	10	5	10	2	0	0

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

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

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).
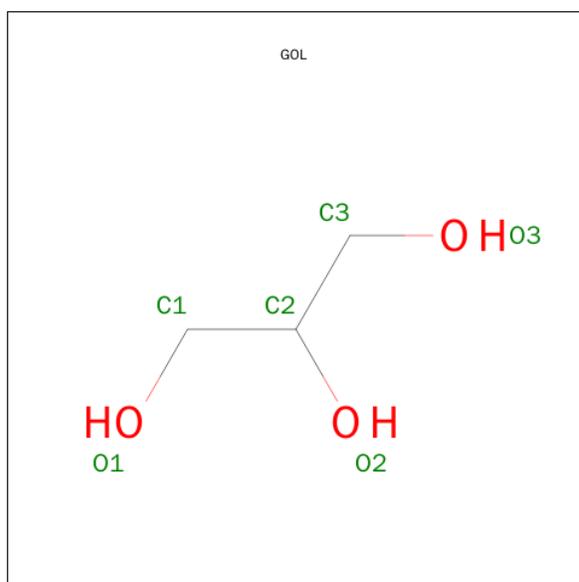
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total 6	C 3	O 3	0	0
5	A	1	Total 6	C 3	O 3	0	0
5	A	1	Total 6	C 3	O 3	0	0
5	A	1	Total 6	C 3	O 3	0	0
5	A	1	Total 6	C 3	O 3	0	0
5	B	1	Total 6	C 3	O 3	0	0

- Molecule 6 is N,N'-ETHANE-1,2-DIYLBIS(2-IODOACETAMIDE) (three-letter code: EBC) (formula: $C_6H_{10}I_2N_2O_2$).

EBC
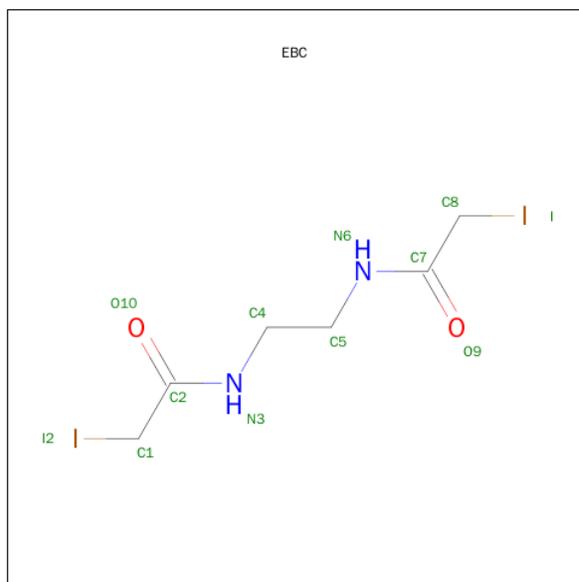
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	B	1	Total	C	N	O	0	0
			10	6	2	2		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	197	Total	O	0	0
			197	197		
7	B	49	Total	O	0	0
			49	49		

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: Kinesin-like protein KAR3
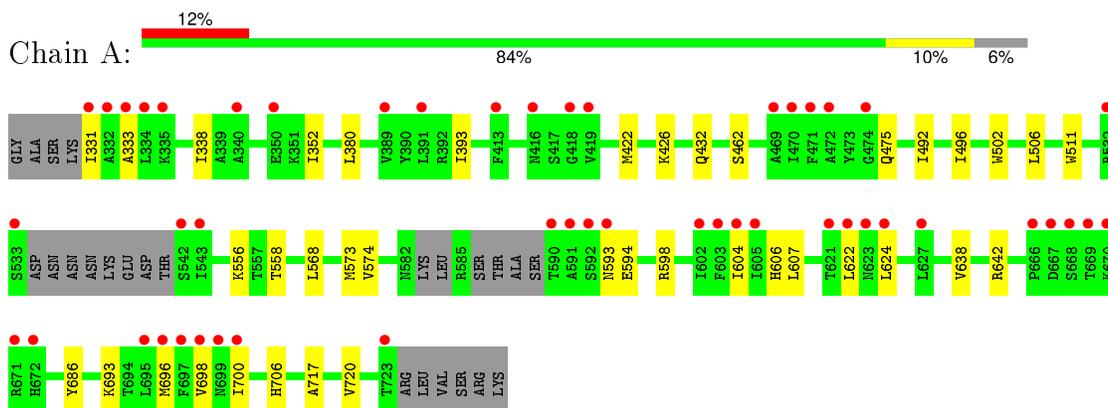
- Molecule 2: Spindle pole body-associated protein VIK1
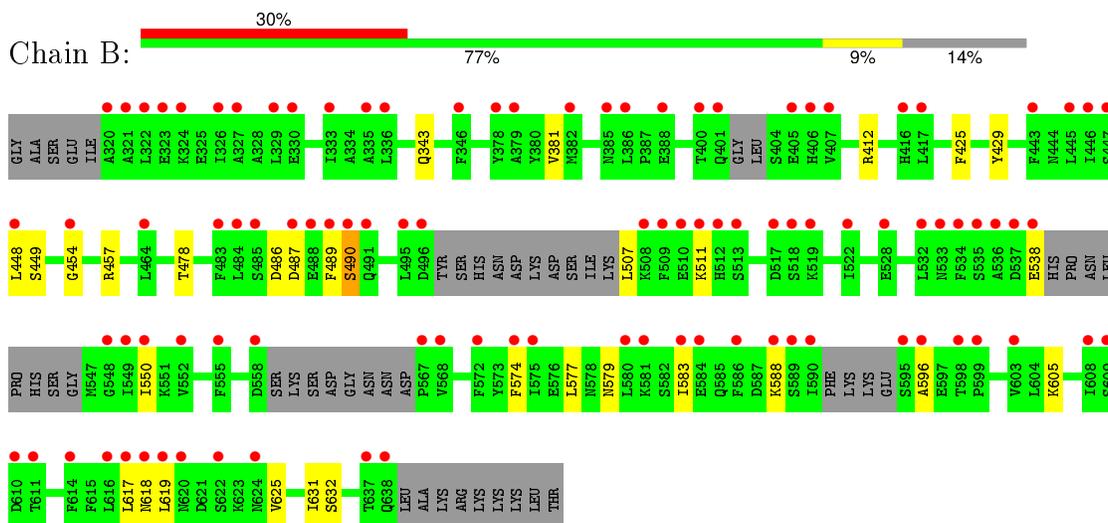
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	88.06Å 94.94Å 114.59Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.97 – 2.30 24.97 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.0 (24.97-2.30) 99.1 (24.97-2.30)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	3.26 (at 2.31Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.207 , 0.245 0.208 , 0.246	Depositor DCC
R_{free} test set	2178 reflections (5.35%)	DCC
Wilson B-factor (Å^2)	49.5	Xtriage
Anisotropy	0.048	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 46.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Outliers	0 of 42865 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5681	wwPDB-VP
Average B, all atoms (Å^2)	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.92% 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, 0.375 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: GOL, MG, ADP, EBC

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	0.39	1/3055 (0.0%)	0.47	0/4118
2	B	0.33	0/2388	0.46	0/3213
All	All	0.36	1/5443 (0.0%)	0.47	0/7331

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	502	TRP	CD2-CE2	5.03	1.47	1.41

There are no bond angle outliers.

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	3010	0	3044	22	0
2	B	2351	0	2353	19	0
3	A	27	0	12	0	0
4	A	1	0	0	0	0
5	A	30	0	40	11	0
5	B	6	0	8	0	0
6	B	10	0	6	0	0
7	A	197	0	0	0	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	B	49	0	0	0	0
All	All	5681	0	5463	45	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 45 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:B:489:PHE:HA	2:B:490:SER:CB	1.87	1.03
5:A:804:GOL:C3	5:A:805:GOL:H31	1.88	1.02
2:B:489:PHE:HA	2:B:490:SER:HB3	1.06	1.02
2:B:489:PHE:CA	2:B:490:SER:HB3	1.97	0.94
2:B:617:LEU:HD11	2:B:632:SER:HB3	1.52	0.92

There are no symmetry-related clashes.

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	375/403 (93%)	373 (100%)	2 (0%)	0	100	100
2	B	275/333 (83%)	266 (97%)	7 (2%)	2 (1%)	26	31
All	All	650/736 (88%)	639 (98%)	9 (1%)	2 (0%)	46	57

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	490	SER
2	B	511	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	343/361 (95%)	337 (98%)	6 (2%)	68	83
2	B	270/310 (87%)	265 (98%)	5 (2%)	65	81
All	All	613/671 (91%)	602 (98%)	11 (2%)	66	82

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	622	LEU
1	A	693	LYS
2	B	507	LEU
1	A	593	ASN
2	B	478	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	401	GLN
2	B	638	GLN
2	B	579	ASN
2	B	343	GLN
2	B	618	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)

Of 9 ligands modelled in this entry, 1 is monoatomic - leaving 8 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	ADP	A	801	4	22,29,29	1.02	1 (4%)	27,45,45	1.90	3 (11%)
5	GOL	A	803	-	5,5,5	0.16	0	5,5,5	0.42	0
5	GOL	A	804	-	5,5,5	0.13	0	5,5,5	0.24	0
5	GOL	A	805	-	5,5,5	0.17	0	5,5,5	0.19	0
5	GOL	A	806	-	5,5,5	0.23	0	5,5,5	0.20	0
5	GOL	A	807	-	5,5,5	0.22	0	5,5,5	0.25	0
6	EBC	B	701	2	9,9,11	0.21	0	10,10,12	0.47	0
5	GOL	B	702	-	5,5,5	0.26	0	5,5,5	0.23	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
3	ADP	A	801	4	-	0/12/32/32	0/3/3/3
5	GOL	A	803	-	-	0/4/4/4	0/0/0/0
5	GOL	A	804	-	-	0/4/4/4	0/0/0/0
5	GOL	A	805	-	-	0/4/4/4	0/0/0/0
5	GOL	A	806	-	-	0/4/4/4	0/0/0/0
5	GOL	A	807	-	-	0/4/4/4	0/0/0/0
6	EBC	B	701	2	-	0/7/7/11	0/0/0/0
5	GOL	B	702	-	-	0/4/4/4	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	801	ADP	C5-C4	3.07	1.47	1.40

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	801	ADP	N3-C2-N1	-7.96	122.80	128.89
3	A	801	ADP	C4-C5-N7	-2.81	106.89	109.48
3	A	801	ADP	C2-N1-C6	2.02	122.38	118.77

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	804	GOL	8	0
5	A	805	GOL	6	0
5	A	806	GOL	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(Å²)	Q<0.9
1	A	379/403 (94%)	0.56	49 (12%)	5	7	26, 45, 98, 163	18 (4%)
2	B	287/333 (86%)	1.55	99 (34%)	0	0	37, 92, 140, 198	44 (15%)
All	All	666/736 (90%)	0.98	148 (22%)	1	1	26, 58, 132, 198	62 (9%)

The worst 5 of 148 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	599	PRO	8.4
2	B	495	LEU	8.0
2	B	320	ALA	8.0
1	A	670	LYS	7.8
1	A	671	ARG	7.7

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
5	GOL	A	806	6/6	0.90	0.33	7.02	31,39,43,51	6
5	GOL	A	804	6/6	0.89	0.56	4.63	24,28,32,34	6
5	GOL	A	805	6/6	0.91	0.41	3.95	19,25,30,34	6
6	EBC	B	701	10/12	0.89	0.23	2.75	71,88,92,103	0
5	GOL	A	803	6/6	0.92	0.21	1.39	53,64,67,68	0
3	ADP	A	801	27/27	0.98	0.11	-0.91	28,36,42,46	0
5	GOL	B	702	6/6	0.61	0.26	-	79,92,103,107	0
5	GOL	A	807	6/6	0.84	0.16	-	74,86,89,91	0
4	MG	A	802	1/1	0.99	0.12	-	33,33,33,33	0

6.5 Other polymers (i)

There are no such residues in this entry.