



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 23, 2016 – 02:57 PM EDT

PDB ID : 4H4K  
Title : Structure of the Cmr2-Cmr3 subcomplex of the Cmr RNA-silencing complex  
Authors : Shao, Y.; Cocozaki, A.I.; Ramia, N.F.; Terns, R.M.; Terns, M.P.; Li, H.  
Deposited on : 2012-09-17  
Resolution : 2.80 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  
<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 : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027107  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0122  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027107

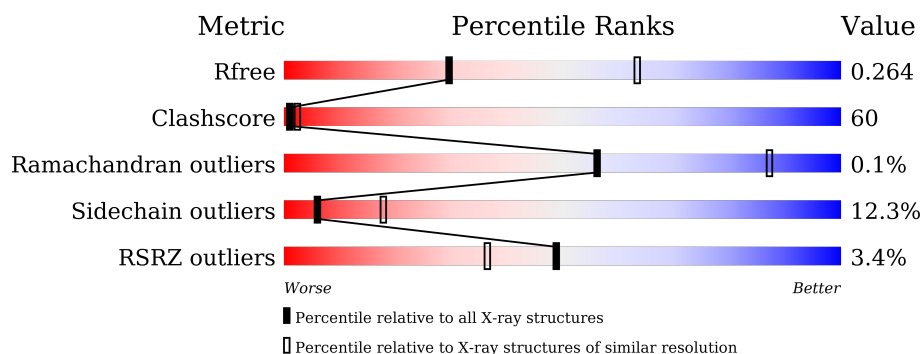
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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. 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	322	<div> <div>3%</div> <div>29% 43% 9% 20%</div> </div>
2	C	696	<div> <div>3%</div> <div>35% 45% 7% 13%</div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7018 atoms, of which 4 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 CRISPR system Cmr subunit Cmr3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	259	Total	C	N	O	S	0	0	0
			2066	1366	327	371	2			

- Molecule 2 is a protein called CRISPR system Cmr subunit Cmr2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	604	Total	C	N	O	S	0	0	0
			4913	3186	812	902	13			

There are 39 discrepancies between the modelled and reference sequences:

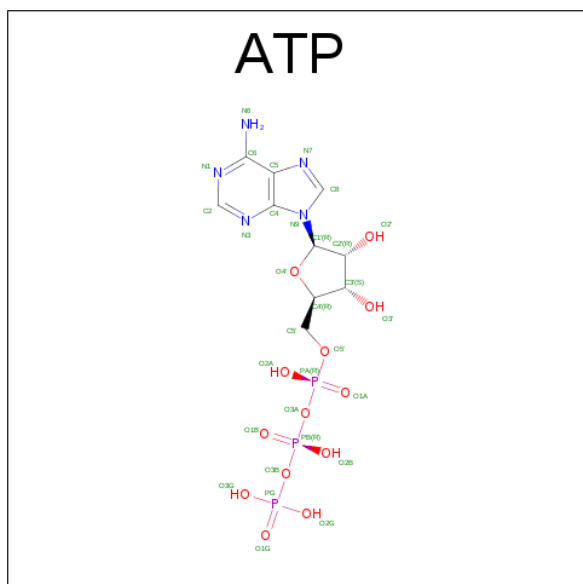
Chain	Residue	Modelled	Actual	Comment	Reference
C	176	MET	-	INITIATING METHIONINE	UNP Q8U1S6
C	177	ARG	-	EXPRESSION TAG	UNP Q8U1S6
C	178	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	179	SER	-	EXPRESSION TAG	UNP Q8U1S6
C	180	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	181	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	182	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	183	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	184	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	185	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	186	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	187	MET	-	EXPRESSION TAG	UNP Q8U1S6
C	188	ALA	-	EXPRESSION TAG	UNP Q8U1S6
C	189	SER	-	EXPRESSION TAG	UNP Q8U1S6
C	190	MET	-	EXPRESSION TAG	UNP Q8U1S6
C	191	THR	-	EXPRESSION TAG	UNP Q8U1S6
C	192	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	193	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	194	GLN	-	EXPRESSION TAG	UNP Q8U1S6
C	195	GLN	-	EXPRESSION TAG	UNP Q8U1S6

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	196	MET	-	EXPRESSION TAG	UNP Q8U1S6
C	197	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	198	ARG	-	EXPRESSION TAG	UNP Q8U1S6
C	199	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	200	LEU	-	EXPRESSION TAG	UNP Q8U1S6
C	201	TYR	-	EXPRESSION TAG	UNP Q8U1S6
C	202	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	203	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	204	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	205	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	206	LYS	-	EXPRESSION TAG	UNP Q8U1S6
C	207	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	208	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	209	PRO	-	EXPRESSION TAG	UNP Q8U1S6
C	210	PHE	-	EXPRESSION TAG	UNP Q8U1S6
C	211	THR	-	EXPRESSION TAG	UNP Q8U1S6
C	212	MET	-	EXPRESSION TAG	UNP Q8U1S6
C	213	ALA	-	EXPRESSION TAG	UNP Q8U1S6
C	214	LEU	-	EXPRESSION TAG	UNP Q8U1S6

- 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
3	C	1	Total	C	H	N	O	P	
			35	10	4	5	13	3	

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	1	Total	Zn	0	0
			1	1		

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	2	Total	Na	0	0
			2	2		

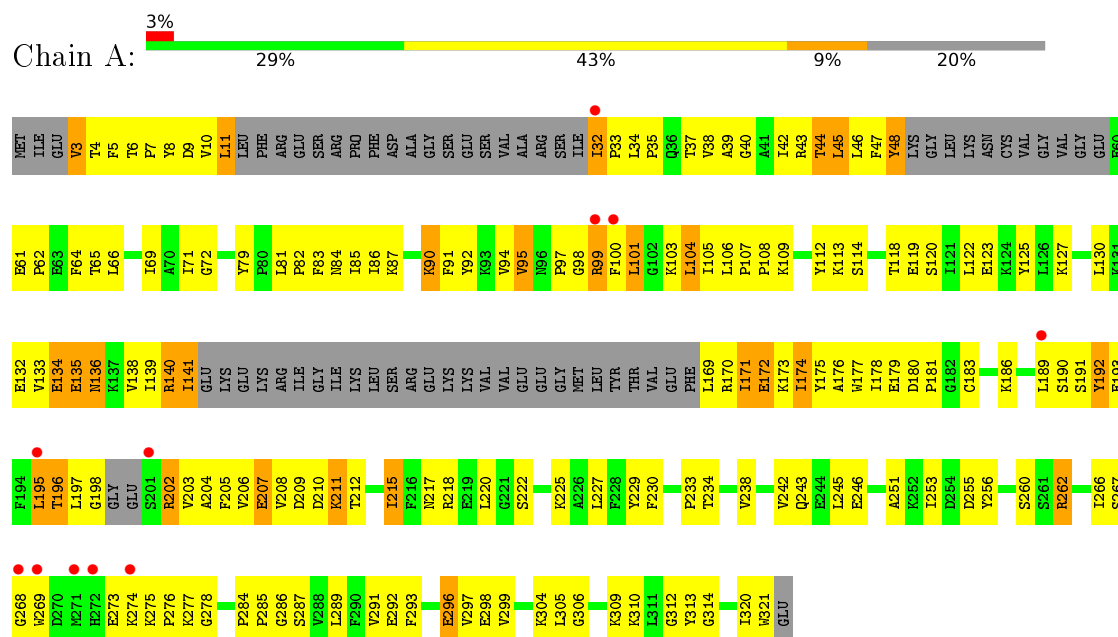
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	1	Total	O	0	0
			1	1		

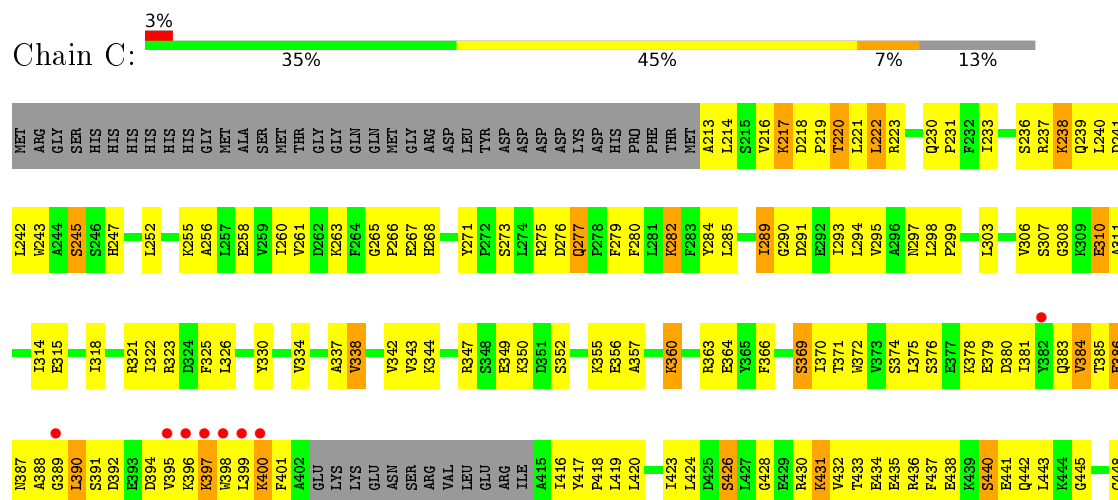
### 3 Residue-property plots

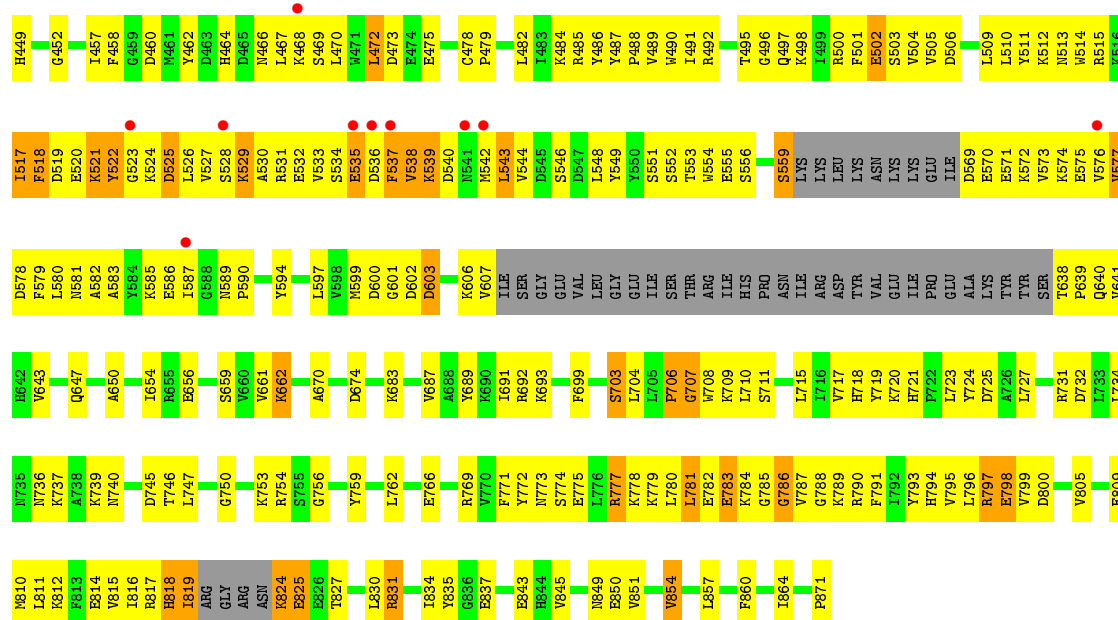
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: CRISPR system Cmr subunit Cmr3



#### • Molecule 2: CRISPR system Cmr subunit Cmr2





## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	102.52Å 135.96Å 189.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.83 – 2.80 46.50 – 2.80	Depositor EDS
% Data completeness (in resolution range)	95.0 (36.83-2.80) 94.1 (46.50-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.14 (at 2.81Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, $R_{free}$	0.214 , 0.284 0.227 , 0.264	Depositor DCC
$R_{free}$ test set	1884 reflections (6.09%)	DCC
Wilson B-factor (Å <sup>2</sup> )	65.7	Xtriage
Anisotropy	0.294	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 56.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 32761 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	7018	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.08% 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, 0.375 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: NA, ZN, 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	0.26	0/2113	0.51	0/2851
2	C	0.26	0/5012	0.55	2/6757 (0.0%)
All	All	0.26	0/7125	0.54	2/9608 (0.0%)

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
2	C	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	C	707	GLY	N-CA-C	-7.51	94.32	113.10
2	C	786	GLY	N-CA-C	6.48	129.29	113.10

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	706	PRO	Peptide

### 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	2066	0	2146	295	1
2	C	4913	0	5009	585	1
3	C	31	4	12	1	0
4	C	1	0	0	0	0
5	C	2	0	0	0	0
6	C	1	0	0	0	0
All	All	7014	4	7167	849	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 60.

The worst 5 of 849 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:C:783:GLU:O	2:C:786:GLY:HA2	1.21	1.32
1:A:140:ARG:O	1:A:170:ARG:NH2	1.71	1.23
2:C:827:THR:HG22	2:C:831:ARG:NH1	1.55	1.21
2:C:370:ILE:O	2:C:431:LYS:NZ	1.75	1.19
2:C:783:GLU:O	2:C:786:GLY:CA	1.89	1.19

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:A:192:TYR:OH	2:C:720:LYS:NZ[2_645]	2.15	0.05

## 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	249/322 (77%)	242 (97%)	7 (3%)	0	100	100
2	C	594/696 (85%)	560 (94%)	33 (6%)	1 (0%)	52	84
All	All	843/1018 (83%)	802 (95%)	40 (5%)	1 (0%)	56	87

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	577	VAL

### 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	225/280 (80%)	191 (85%)	34 (15%)	3	10
2	C	537/619 (87%)	477 (89%)	60 (11%)	7	22
All	All	762/899 (85%)	668 (88%)	94 (12%)	6	18

5 of 94 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	307	SER
2	C	390	LEU
2	C	805	VAL
2	C	310	GLU
2	C	369	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	268	HIS
2	C	464	HIS
2	C	497	GLN
2	C	581	ASN
2	C	794	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 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	C	901	5	26,33,33	0.98	2 (7%)	26,52,52	2.25	4 (15%)

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	C	901	5	-	0/18/38/38	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	901	ATP	C2-N3	2.14	1.35	1.32
3	C	901	ATP	C5-C4	2.85	1.46	1.40

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
3	C	901	ATP	N3-C2-N1	-9.39	121.50	128.87
3	C	901	ATP	C2'-C1'-N9	-3.61	103.81	113.47
3	C	901	ATP	C2-N1-C6	2.17	122.64	118.77
3	C	901	ATP	N6-C6-N1	2.81	123.24	118.52

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	901	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, 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	259/322 (80%)	0.08	11 (4%) 40 28	45, 72, 108, 123	0
2	C	604/696 (86%)	0.18	18 (2%) 54 41	46, 74, 126, 149	0
All	All	863/1018 (84%)	0.15	29 (3%) 49 36	45, 73, 123, 149	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	400	LYS	5.0
1	A	189	LEU	4.9
2	C	587	ILE	4.8
2	C	382	TYR	4.7
2	C	536	ASP	4.5

### 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, 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ATP	C	901	31/31	0.91	0.24	-0.22	51,69,119,130	0
4	ZN	C	902	1/1	0.96	0.15	-1.00	62,62,62,62	0
5	NA	C	904	1/1	0.62	0.63	-	81,81,81,81	0
5	NA	C	903	1/1	0.94	0.56	-	87,87,87,87	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.