wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 01:41 PM GMT

PDB ID	:	3UNG
Title	:	Structure of the Cmr2 subunit of the CRISPR RNA silencing complex
Authors	:	Cocozaki, A.I.; Ramia, N.F.; Shao, Y.; Hale, C.R.; Terns, R.M.; Terns, M.P.; Li, H.
Deposited on	:	2011-11-15
Resolution	:	2.31 Å(reported)

The following versions of software and data (see references (i)) 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 2.31 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.

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution
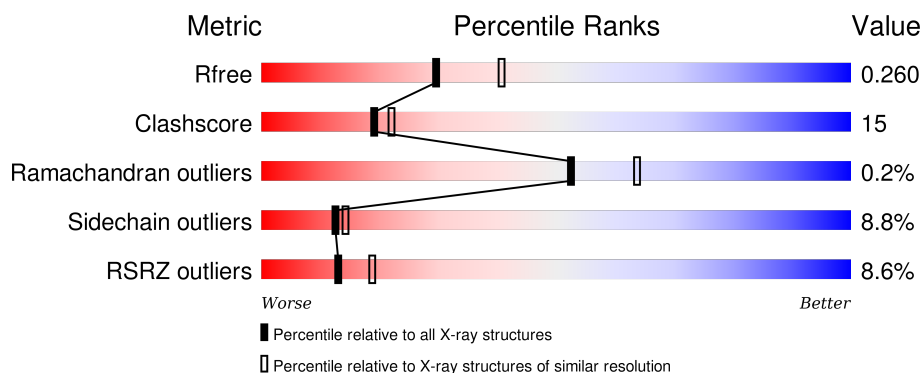
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	4425 (2.34-2.30)
Clashscore	102246	5057 (2.34-2.30)
Ramachandran outliers	100387	5008 (2.34-2.30)
Sidechain outliers	100360	5007 (2.34-2.30)
RSRZ outliers	91569	4432 (2.34-2.30)

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	C	693	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4534 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 Cmr2dHD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	C	548	4469	2905	740	811	13	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	179	MET	-	EXPRESSION TAG	UNP Q8U1S6
C	180	ARG	-	EXPRESSION TAG	UNP Q8U1S6
C	181	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	182	SER	-	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	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	187	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	188	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	189	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	190	MET	-	EXPRESSION TAG	UNP Q8U1S6
C	191	ALA	-	EXPRESSION TAG	UNP Q8U1S6
C	192	SER	-	EXPRESSION TAG	UNP Q8U1S6
C	193	MET	-	EXPRESSION TAG	UNP Q8U1S6
C	194	THR	-	EXPRESSION TAG	UNP Q8U1S6
C	195	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	196	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	197	GLN	-	EXPRESSION TAG	UNP Q8U1S6
C	198	GLN	-	EXPRESSION TAG	UNP Q8U1S6
C	199	MET	-	EXPRESSION TAG	UNP Q8U1S6
C	200	GLY	-	EXPRESSION TAG	UNP Q8U1S6
C	201	ARG	-	EXPRESSION TAG	UNP Q8U1S6
C	202	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	203	LEU	-	EXPRESSION TAG	UNP Q8U1S6
C	204	TYR	-	EXPRESSION TAG	UNP Q8U1S6
C	205	ASP	-	EXPRESSION TAG	UNP Q8U1S6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	206	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	207	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	208	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	209	LYS	-	EXPRESSION TAG	UNP Q8U1S6
C	210	ASP	-	EXPRESSION TAG	UNP Q8U1S6
C	211	HIS	-	EXPRESSION TAG	UNP Q8U1S6
C	212	PRO	-	EXPRESSION TAG	UNP Q8U1S6
C	213	PHE	-	EXPRESSION TAG	UNP Q8U1S6
C	214	THR	-	EXPRESSION TAG	UNP Q8U1S6

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	2	Total 2	Ca 2	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 4 is A D E N O S I N E - 5'- D I P H O S P H A T E (three-letter code: ADP) (formula: $C_{10}H_{15}N_5O_{10}P_2$).
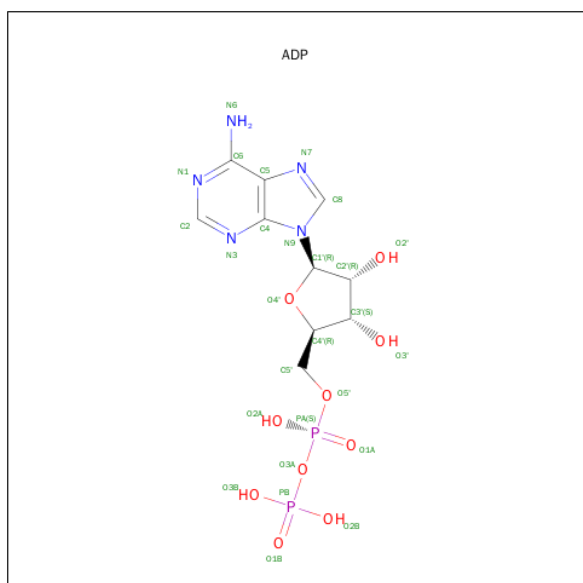
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	C	1	Total 27	C 10	N 5	O 10	P 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	35	Total 35	O 35	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: Cmr2dHD
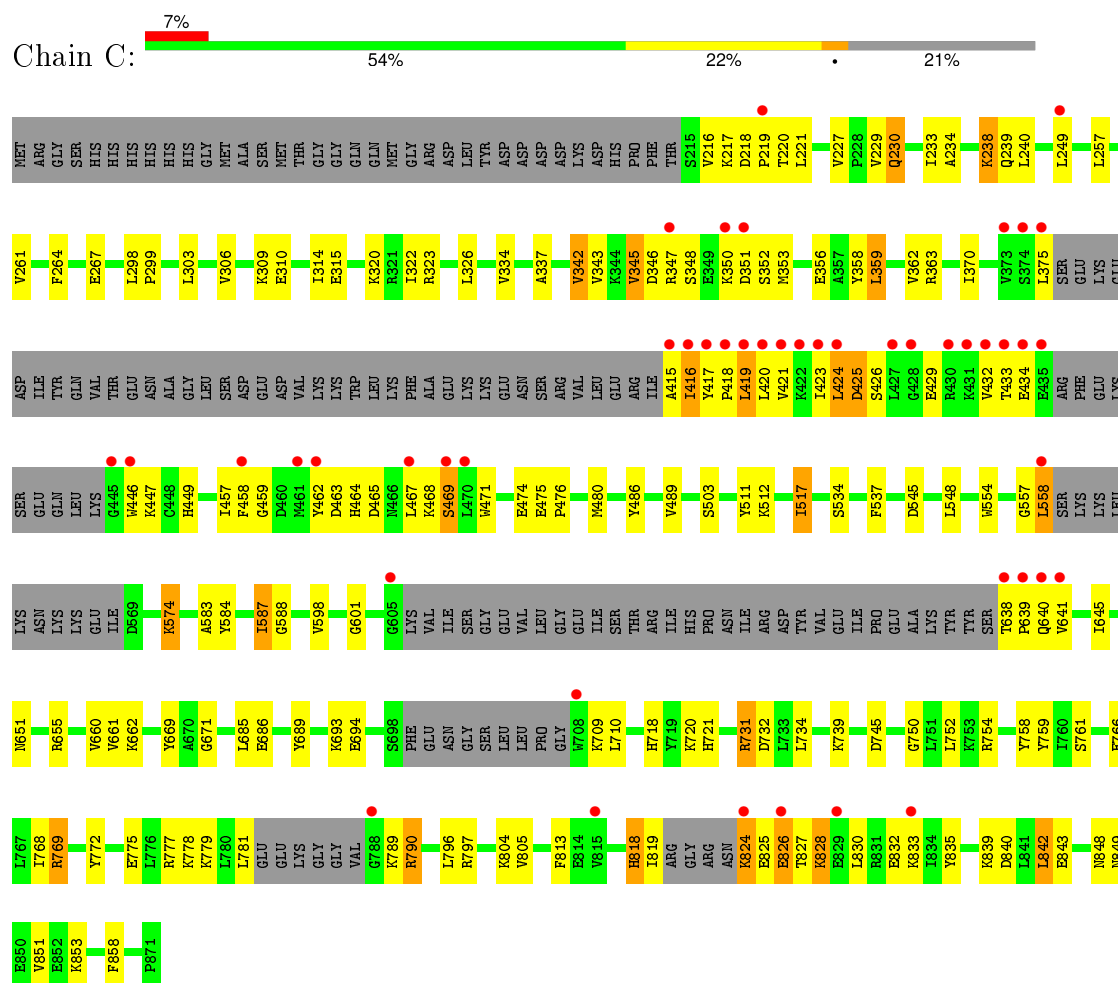
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	73.17Å 85.07Å 140.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.61 – 2.31 43.49 – 2.31	Depositor EDS
% Data completeness (in resolution range)	79.9 (29.61-2.31) 78.6 (43.49-2.31)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.21 (at 2.32Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, R_{free}	0.203 , 0.264 0.195 , 0.260	Depositor DCC
R_{free} test set	1778 reflections (5.69%)	DCC
Wilson B-factor (Å^2)	52.5	Xtriage
Anisotropy	0.522	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 67.1	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 35242 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4534	wwPDB-VP
Average B, all atoms (Å^2)	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.76% 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: ZN, CA, 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	C	0.43	0/4558	0.61	0/6145

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 (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	C	4469	0	4567	138	0
2	C	2	0	0	0	0
3	C	1	0	0	0	0
4	C	27	0	12	1	0
5	C	35	0	0	1	0
All	All	4534	0	4579	138	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 15.

The worst 5 of 138 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:C:517:ILE:HG13	1:C:587:ILE:HD11	1.23	1.10
1:C:230:GLN:O	1:C:234:ALA:HB2	1.55	1.05
1:C:219:PRO:HB2	1:C:375:LEU:HD12	1.48	0.92
1:C:731:ARG:HH11	1:C:731:ARG:HG2	1.34	0.90
1:C:362:VAL:HA	1:C:480:MET:CE	2.03	0.89

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	C	532/693 (77%)	511 (96%)	20 (4%)	1 (0%)	52 64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	416	ILE

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	C	489/617 (79%)	446 (91%)	43 (9%)	12 14

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	558	LEU
1	C	686	GLU
1	C	832	GLU
1	C	574	LYS
1	C	587	ILE

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

Mol	Chain	Res	Type
1	C	235	ASN
1	C	718	HIS
1	C	740	ASN
1	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	$\#|Z| > 2$
4	ADP	C	904	2	22,29,29	1.09	1 (4%)	27,45,45	1.96	6 (22%)

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
4	ADP	C	904	2	-	0/12/32/32	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	904	ADP	C5-C4	3.57	1.48	1.40

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	C	904	ADP	N3-C2-N1	-6.86	123.64	128.89
4	C	904	ADP	C4-C5-N7	-3.24	106.50	109.48
4	C	904	ADP	PA-O3A-PB	-3.11	122.24	132.67
4	C	904	ADP	C2'-C1'-N9	-2.74	110.10	114.29
4	C	904	ADP	C2-N1-C6	2.26	122.80	118.77

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
4	C	904	ADP	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	C	548/693 (79%)	0.60	47 (8%)	13	19	44, 75, 138, 184	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	375	LEU	9.4
1	C	416	ILE	9.2
1	C	420	LEU	8.3
1	C	421	VAL	7.1
1	C	415	ALA	7.0

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	ZN	C	903	1/1	0.99	0.14	1.20	79,79,79,79	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors($Å^2$)	Q<0.9
2	CA	C	902	1/1	0.95	0.14	0.46	86,86,86,86	0
4	ADP	C	904	27/27	0.89	0.18	0.19	55,97,173,263	0
2	CA	C	901	1/1	0.92	0.17	-	85,85,85,85	0

6.5 Other polymers (i)

There are no such residues in this entry.