



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 28, 2014 – 01:46 PM GMT

PDB ID : 2XR1  
Title : DIMERIC ARCHAEL CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR WITH N-TERMINAL KH DOMAINS (KH-CPSF) FROM METHANOSARCINA MAZEI  
Authors : Mir-Montazeri, B.; Ammelburg, M.; Forouzan, D.; Lupas, A.N.; Hartmann, M.D.  
Deposited on : 2010-09-08  
Resolution : 2.59 Å(reported)

This is a full wwPDB validation 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/ValidationPDFNotes.html>

---

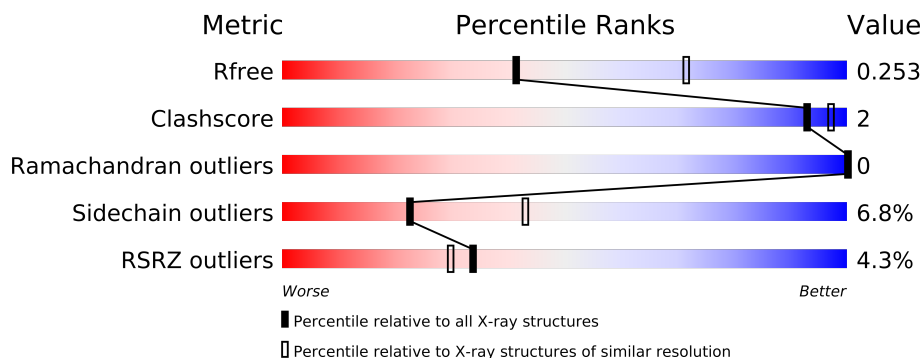
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

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))
$R_{free}$	66092	1718 (2.60-2.60)
Clashscore	79885	2154 (2.60-2.60)
Ramachandran outliers	78287	2113 (2.60-2.60)
Sidechain outliers	78261	2113 (2.60-2.60)
RSRZ outliers	66119	1718 (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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	640	
1	B	640	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9391 atoms, of which 0 are hydrogen and 0 are deuterium.

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 CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR 100 KD SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	605	Total	C	N	O	S	0	1	0
			4640	2964	791	864	21			
1	B	611	Total	C	N	O	S	0	1	0
			4705	3008	797	875	25			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP Q8PZ03
A	-1	SER	-	EXPRESSION TAG	UNP Q8PZ03
A	0	HIS	-	EXPRESSION TAG	UNP Q8PZ03
B	-2	GLY	-	EXPRESSION TAG	UNP Q8PZ03
B	-1	SER	-	EXPRESSION TAG	UNP Q8PZ03
B	0	HIS	-	EXPRESSION TAG	UNP Q8PZ03

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	2	Total	Zn	0	0
			2	2		
2	A	2	Total	Zn	0	0
			2	2		

- Molecule 3 is water.

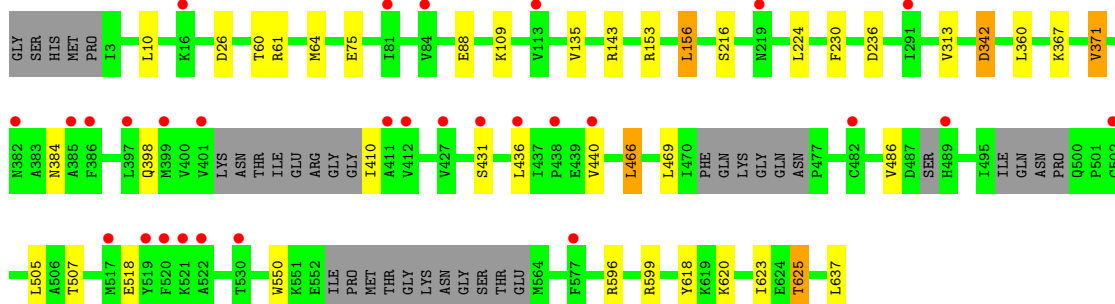
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	16	Total	O	0	0
			16	16		
3	B	26	Total	O	0	0
			26	26		

### 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.

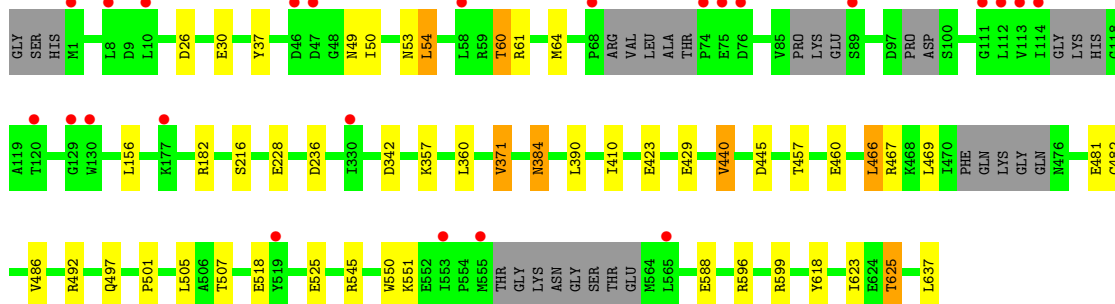
- Molecule 1: CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR 100 KD SUB-UNIT

Chain A: 



- Molecule 1: CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR 100 KD SUB-UNIT

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	77.66Å 97.72Å 90.72Å 90.00° 98.26° 90.00°	Depositor
Resolution (Å)	38.43 – 2.59 38.43 – 2.60	Depositor EDS
% Data completeness (in resolution range)	98.9 (38.43-2.59) 98.9 (38.43-2.60)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.19 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.209 , 0.267 0.204 , 0.253	Depositor DCC
$R_{free}$ test set	2023 reflections (5.17%)	DCC
Wilson B-factor (Å <sup>2</sup> )	54.2	Xtriage
Anisotropy	0.061	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 29.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 41153 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9391	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.78% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.67	0/4741	0.72	2/6450 (0.0%)
1	B	0.80	3/4807 (0.1%)	0.76	2/6529 (0.0%)
All	All	0.74	3/9548 (0.0%)	0.74	4/12979 (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
1	B	0	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	482	CYS	CB-SG	6.76	1.93	1.82
1	B	423	GLU	CG-CD	6.09	1.61	1.51
1	B	588	GLU	CG-CD	5.48	1.60	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	371	VAL	CB-CA-C	-6.41	99.23	111.40
1	A	342	ASP	CB-CG-OD1	-6.30	112.63	118.30
1	A	371	VAL	CB-CA-C	-6.15	99.71	111.40
1	B	182	ARG	NE-CZ-NH2	-5.21	117.70	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	384	ASN	Peptide

## 5.2 Close contacts ⓘ

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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4640	0	0	6	0
1	B	4705	0	0	10	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	16	0	0	0	0
3	B	26	0	0	0	0
All	All	9391	0	0	16	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 2.

All (16) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:360:LEU:CD2	1:A:466:LEU:CD1	2.82	0.56
1:B:30:GLU:OE2	1:B:37:TYR:OH	2.28	0.51
1:B:457:THR:CG2	1:B:457:THR:O	2.59	0.49
1:A:623:ILE:O	1:A:625:THR:CG2	2.61	0.49
1:A:224:LEU:O	1:A:230:PHE:CE1	2.66	0.49
1:B:49:ASN:O	1:B:53:ASN:N	2.45	0.48
1:A:156:LEU:CD2	1:A:156:LEU:O	2.63	0.47
1:B:360:LEU:CD2	1:B:466:LEU:CD1	2.93	0.46
1:B:228:GLU:OE1	1:B:228:GLU:N	2.49	0.46
1:B:445:ASP:OD2	1:B:492:ARG:NE	2.50	0.45
1:A:398:GLN:NE2	1:A:431:SER:OG	2.50	0.44
1:B:50:ILE:O	1:B:54:LEU:N	2.52	0.43
1:B:60:THR:OG1	1:B:61:ARG:N	2.52	0.42
1:B:623:ILE:O	1:B:625:THR:CG2	2.67	0.42
1:B:440:VAL:CG2	1:B:501:PRO:O	2.68	0.41
1:A:61:ARG:NE	1:A:313:VAL:O	2.54	0.41

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	594/640 (93%)	567 (96%)	27 (4%)	0	100	100
1	B	598/640 (93%)	570 (95%)	28 (5%)	0	100	100
All	All	1192/1280 (93%)	1137 (95%)	55 (5%)	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	487/560 (87%)	454 (93%)	33 (7%)	22	43
1	B	495/560 (88%)	461 (93%)	34 (7%)	22	42
All	All	982/1120 (88%)	915 (93%)	67 (7%)	22	43

All (67) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	10	LEU
1	A	26	ASP
1	A	60	THR
1	A	64	MET
1	A	75	GLU
1	A	88	GLU
1	A	109	LYS
1	A	135	VAL

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	143	ARG
1	A	153	ARG
1	A	156	LEU
1	A	216	SER
1	A	236	ASP
1	A	342	ASP
1	A	367	LYS
1	A	371	VAL
1	A	384	ASN
1	A	410	ILE
1	A	436	LEU
1	A	440	VAL
1	A	466	LEU
1	A	469	LEU
1	A	486	VAL
1	A	505	LEU
1	A	507	THR
1	A	518	GLU
1	A	550	TRP
1	A	596	ARG
1	A	599	ARG
1	A	618	TYR
1	A	620	LYS
1	A	625	THR
1	A	637	LEU
1	B	26	ASP
1	B	54	LEU
1	B	60	THR
1	B	64	MET
1	B	156	LEU
1	B	216	SER
1	B	236	ASP
1	B	342	ASP
1	B	357	LYS
1	B	371	VAL
1	B	384	ASN
1	B	390	LEU
1	B	410	ILE
1	B	429	GLU
1	B	440	VAL
1	B	460	GLU
1	B	466	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	467	ARG
1	B	469	LEU
1	B	481	GLU
1	B	486	VAL
1	B	497	GLN
1	B	505	LEU
1	B	507	THR
1	B	518	GLU
1	B	525	GLU
1	B	545	ARG
1	B	550	TRP
1	B	551	LYS
1	B	596	ARG
1	B	599	ARG
1	B	618	TYR
1	B	625	THR
1	B	637	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains 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, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

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	605/640 (94%)	0.42	29 (4%) 29 26	28, 60, 121, 158	0
1	B	611/640 (95%)	0.29	24 (3%) 37 33	22, 46, 93, 119	0
All	All	1216/1280 (95%)	0.35	53 (4%) 34 29	22, 55, 107, 158	0

All (53) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	520	PHE	5.4
1	B	555	MET	5.3
1	A	440	VAL	4.2
1	B	8	LEU	3.8
1	A	411	ALA	3.8
1	A	431	SER	3.6
1	A	81	ILE	3.6
1	B	112	LEU	3.6
1	A	530	THR	3.5
1	B	89	SER	3.5
1	B	114	ILE	3.4
1	A	517	MET	3.3
1	A	521	LYS	3.3
1	B	565	LEU	3.2
1	A	438	PRO	3.2
1	B	113	VAL	3.1
1	B	76	ASP	3.1
1	B	553	ILE	2.9
1	A	489	HIS	2.8
1	B	75	GLU	2.8
1	B	74	PRO	2.7
1	A	401	VAL	2.7
1	B	1	MET	2.7
1	B	120	THR	2.7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	482	CYS	2.6
1	B	68	PRO	2.6
1	A	436	LEU	2.6
1	A	399	MET	2.6
1	A	502	CYS	2.6
1	A	522	ALA	2.6
1	A	427	VAL	2.6
1	B	46	ASP	2.5
1	A	385	ALA	2.5
1	A	291	ILE	2.5
1	B	130	TRP	2.5
1	B	58	LEU	2.5
1	A	219	ASN	2.4
1	A	382	ASN	2.4
1	A	412	VAL	2.3
1	A	386	PHE	2.3
1	A	577	PHE	2.3
1	B	177	LYS	2.2
1	B	47	ASP	2.2
1	A	519	TYR	2.2
1	A	16	LYS	2.2
1	B	519	TYR	2.2
1	B	330	ILE	2.1
1	B	129	GLY	2.1
1	B	10	LEU	2.1
1	A	397	LEU	2.1
1	A	113	VAL	2.1
1	B	111	GLY	2.1
1	A	84	VAL	2.0

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

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

## 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 6.4 Ligands

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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	ZN	A	1638	1/1	0.20	-0.25	49,49,49,49	0
2	ZN	B	1638	1/1	0.18	-0.64	34,34,34,34	0
2	ZN	A	1639	1/1	0.18	-0.67	52,52,52,52	0
2	ZN	B	1639	1/1	0.16	-0.67	35,35,35,35	0

## 6.5 Other polymers

There are no such residues in this entry.