



wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 02:12 pm BST

PDB ID : 2ZFH
Title : Crystal structure of putative CutA1 from Homo sapiens at 2.05Å resolution
Authors : Bagautdinov, B.; Yutani, K.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on : 2008-01-07
Resolution : 2.05 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

<https://www.wwpdb.org/validation/2017/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
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

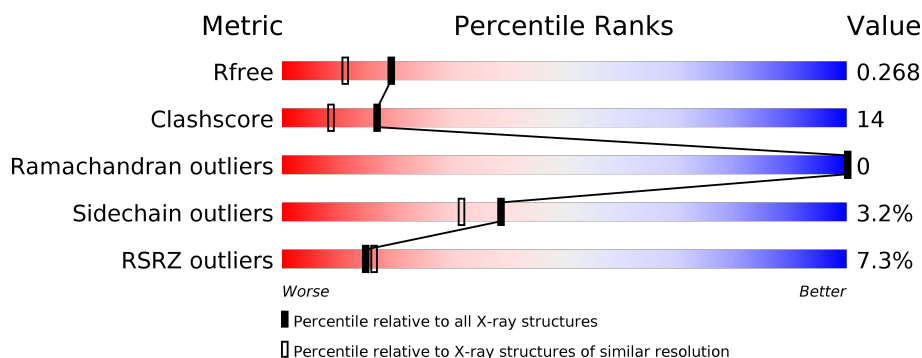
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.05 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}	130704	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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 respectively. 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	179	
1	B	179	
1	C	179	
1	D	179	
1	E	179	
1	F	179	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5279 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 CutA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	109	Total	C	N	O	S	0	0	0
			851	548	139	160	4			
1	B	107	Total	C	N	O	S	0	0	0
			836	540	137	155	4			
1	C	106	Total	C	N	O	S	0	0	0
			832	538	136	154	4			
1	D	107	Total	C	N	O	S	0	0	0
			836	540	137	155	4			
1	E	107	Total	C	N	O	S	0	0	0
			836	540	137	155	4			
1	F	107	Total	C	N	O	S	0	0	0
			836	540	137	155	4			

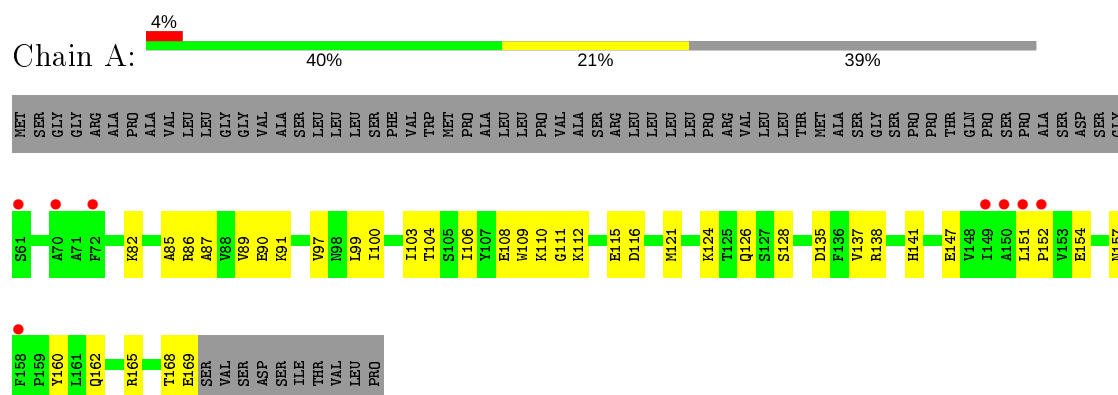
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	35	Total	O	0	0
			35	35		
2	B	37	Total	O	0	0
			37	37		
2	C	50	Total	O	0	0
			50	50		
2	D	48	Total	O	0	0
			48	48		
2	E	44	Total	O	0	0
			44	44		
2	F	38	Total	O	0	0
			38	38		

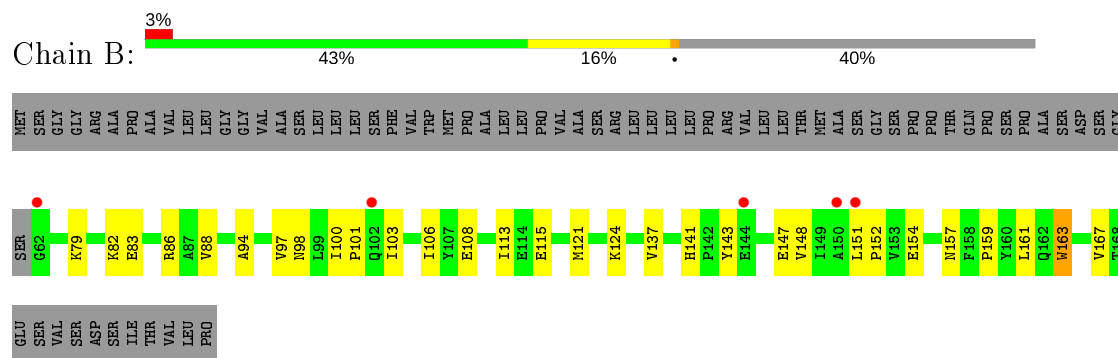
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

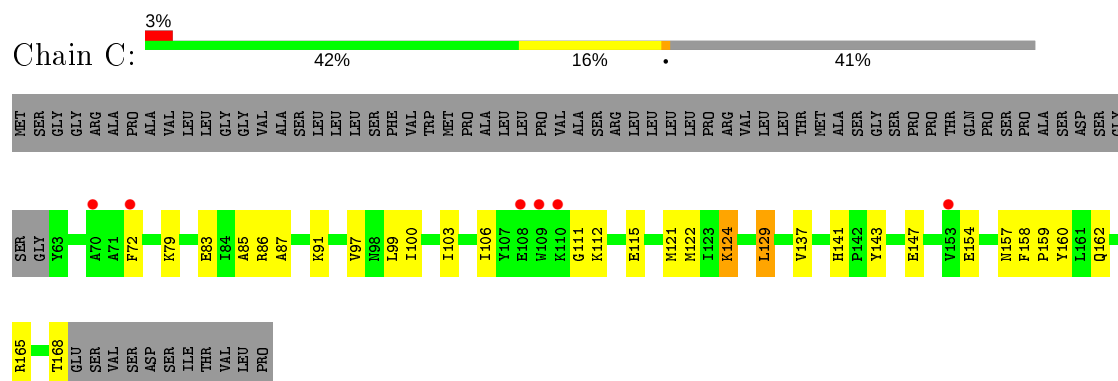
• Molecule 1: CutA



• Molecule 1: CutA



• Molecule 1: CutA



Chain D:

9% 39% 20% 40%

MET SER GLY ARG ALA PRO VAL LEU LEU GLY GLY VAL ALA SER LEU LEU LEU SER PHE VAL TRP MET PRO ALA LEU LEU PRO VAL VAL ALA SER ARG LEU LEU LEU LEU PRO ARG VAL LEU LEU THR MET ALA SER SER GLY SER PRO PRO GLN PRO SER PRO ALA ASP SER GLY

SER G62 G63 V64 A70 A71 T74 K79 A85 A86 A87 V88 V89 E90 K91 A94 I99 I100 I106 I107 E108 E109 W109 K110 G111 I112 I113 I114 E115 E118 V119 L120 M121 K124 L133 F136 V137 H141 E147 V148 I149 A150 L151 P152 V153 E154 V157

F168 P159 Q162 R165 Q166 V167 T168 GLU SER SER SER ASP ASP ILE THR VAL PRO

Chain E:

Position	Most Conserved Amino Acid	Conservation Percentage
1	LEU	4%
2	LEU	47%
3	LEU	13%
4	LEU	40%

Chain F:

Amino Acid Composition (Chain F):

Amino Acid	Percentage
ALA	~1%
ARG	~1%
ASP	~1%
ASN	~1%
CYS	~1%
GLN	~1%
GLY	~1%
HIS	~1%
ILE	~1%
LEU	~1%
LYS	~1%
PHE	~1%
PRO	~1%
SER	~1%
THR	~1%
TRP	~1%
Tyr	~1%
VAL	~1%

The remaining 98% of the composition is distributed among other amino acids, with LEU being the most abundant at approximately 42%, followed by ARG at 17%, and ASP and GLY each at 40%.

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	68.69Å 88.84Å 125.33Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.81 – 2.05 37.81 – 2.06	Depositor EDS
% Data completeness (in resolution range)	94.7 (37.81-2.05) 94.8 (37.81-2.06)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.86 (at 2.06Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.221 , 0.265 0.223 , 0.268	Depositor DCC
R_{free} test set	2291 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	35.9	Xtriage
Anisotropy	0.906	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 53.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	5279	wwPDB-VP
Average B, all atoms (Å ²)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.80% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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	0.35	0/869	0.57	0/1185
1	B	0.35	0/854	0.60	0/1165
1	C	0.38	0/850	0.59	0/1160
1	D	0.40	0/854	0.63	0/1165
1	E	0.35	0/854	0.57	0/1165
1	F	0.37	0/854	0.58	0/1165
All	All	0.37	0/5135	0.59	0/7005

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	A	851	0	863	30	0
1	B	836	0	852	24	0
1	C	832	0	849	26	0
1	D	836	0	852	39	0
1	E	836	0	852	24	0
1	F	836	0	852	28	0
2	A	35	0	0	0	0
2	B	37	0	0	0	0
2	C	50	0	0	0	0
2	D	48	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	44	0	0	2	0
2	F	38	0	0	1	0
All	All	5279	0	5120	143	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 14.

The worst 5 of 143 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:108:GLU:HG3	1:D:113:ILE:HD13	1.42	0.97
1:D:162:GLN:HG3	1:D:165:ARG:NH2	1.92	0.84
1:C:122:MET:HG2	1:C:124:LYS:HZ3	1.46	0.80
1:B:100:ILE:HG22	1:B:103:ILE:HD11	1.62	0.80
1:B:106:ILE:HG12	1:B:115:GLU:HG2	1.69	0.73

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	A	107/179 (60%)	104 (97%)	3 (3%)	0	100	100
1	B	105/179 (59%)	105 (100%)	0	0	100	100
1	C	104/179 (58%)	102 (98%)	2 (2%)	0	100	100
1	D	105/179 (59%)	103 (98%)	2 (2%)	0	100	100
1	E	105/179 (59%)	103 (98%)	2 (2%)	0	100	100
1	F	105/179 (59%)	104 (99%)	1 (1%)	0	100	100
All	All	631/1074 (59%)	621 (98%)	10 (2%)	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	95/152 (62%)	93 (98%)	2 (2%)	53	48
1	B	93/152 (61%)	89 (96%)	4 (4%)	29	22
1	C	93/152 (61%)	91 (98%)	2 (2%)	52	46
1	D	93/152 (61%)	89 (96%)	4 (4%)	29	22
1	E	93/152 (61%)	91 (98%)	2 (2%)	52	46
1	F	93/152 (61%)	89 (96%)	4 (4%)	29	22
All	All	560/912 (61%)	542 (97%)	18 (3%)	39	32

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	79	LYS
1	D	109	TRP
1	F	114	GLU
1	C	124	LYS
1	C	129	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	166	GLN
1	D	98	ASN
1	F	98	ASN
1	C	98	ASN
1	E	162	GLN

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](#)

There are no ligands in this entry.

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 ⓘ

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, 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	A	109/179 (60%)	0.28	8 (7%) 15 16	35, 48, 62, 73	0
1	B	107/179 (59%)	0.16	5 (4%) 31 33	34, 47, 64, 77	0
1	C	106/179 (59%)	0.23	6 (5%) 23 25	33, 45, 68, 78	0
1	D	107/179 (59%)	0.68	16 (14%) 2 1	36, 49, 79, 92	0
1	E	107/179 (59%)	0.36	8 (7%) 14 15	33, 49, 78, 99	0
1	F	107/179 (59%)	0.27	4 (3%) 41 45	35, 48, 62, 67	0
All	All	643/1074 (59%)	0.33	47 (7%) 15 16	33, 48, 67, 99	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	109	TRP	7.8
1	D	111	GLY	5.6
1	D	110	LYS	5.4
1	C	108	GLU	4.3
1	E	110	LYS	4.3

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 ⓘ

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.