wwPDB X-ray Structure Validation Summary Report (i)

Sep 18, 2023 – 06:46 PM EDT

PDB ID	:	5DII
Title	:	Structure of an engineered bacterial microcompartment shell protein binding a [4Fe-4S] cluster
Authors	:	Sutter, M.; Aussignargues, C.; Turmo, A.; Kerfeld, C.A.
Deposited on	:	2015-09-01
Resolution	:	1.80 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

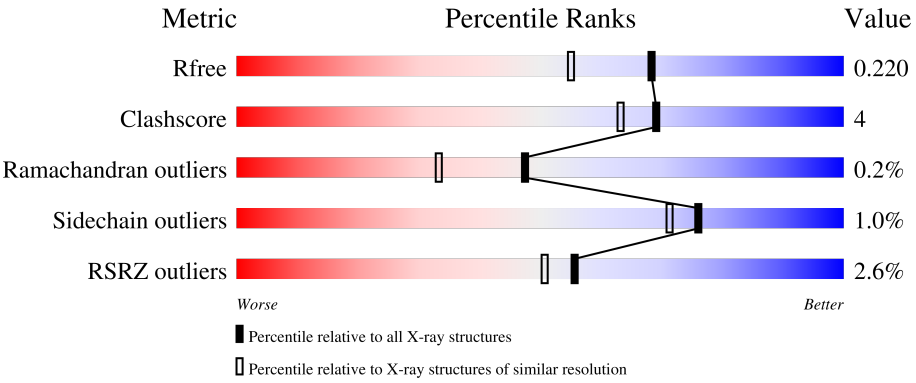
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.80 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.



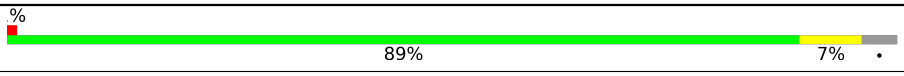
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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 of 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 <=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	205	3% 87% 7% 5%
1	B	205	90% 5% .
1	C	205	6% 89% 7% . .
1	D	205	3% 88% 9% .
1	E	205	% 88% 6% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	205	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 18044 atoms, of which 8809 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.

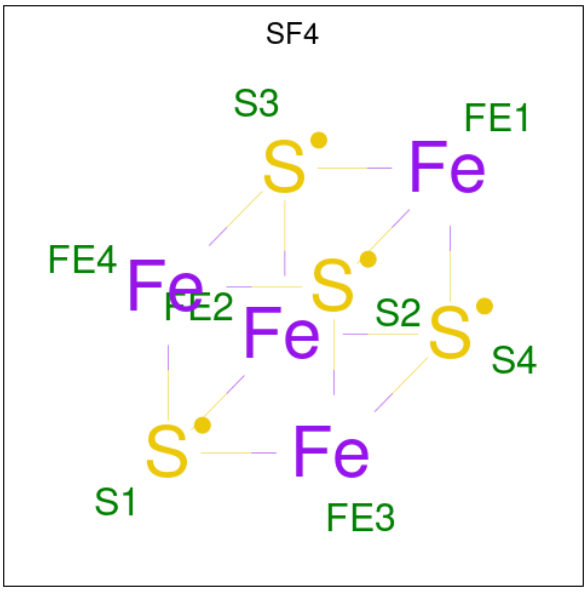
- Molecule 1 is a protein called Microcompartments protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	194	Total	C	H	N	O	S	0	0	0
			2925	914	1477	253	272	9			
1	B	196	Total	C	H	N	O	S	0	0	0
			2933	925	1469	256	274	9			
1	C	199	Total	C	H	N	O	S	0	0	0
			2957	943	1464	261	279	10			
1	D	199	Total	C	H	N	O	S	0	0	0
			2946	934	1465	256	282	9			
1	E	195	Total	C	H	N	O	S	0	0	0
			2921	920	1466	255	271	9			
1	F	197	Total	C	H	N	O	S	0	0	0
			2939	930	1468	257	275	9			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	55	CYS	SER	engineered mutation	UNP D0LHE3
B	55	CYS	SER	engineered mutation	UNP D0LHE3
C	55	CYS	SER	engineered mutation	UNP D0LHE3
D	55	CYS	SER	engineered mutation	UNP D0LHE3
E	55	CYS	SER	engineered mutation	UNP D0LHE3
F	55	CYS	SER	engineered mutation	UNP D0LHE3

- Molecule 2 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Fe S 8 4 4	0	0
2	D	1	Total Fe S 8 4 4	0	0

- Molecule 3 is water.

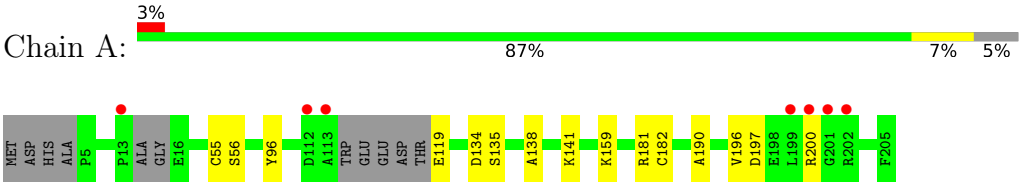
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	53	Total O 53 53	0	0
3	B	81	Total O 81 81	0	0
3	C	59	Total O 59 59	0	0
3	D	65	Total O 65 65	0	0
3	E	80	Total O 80 80	0	0
3	F	69	Total O 69 69	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

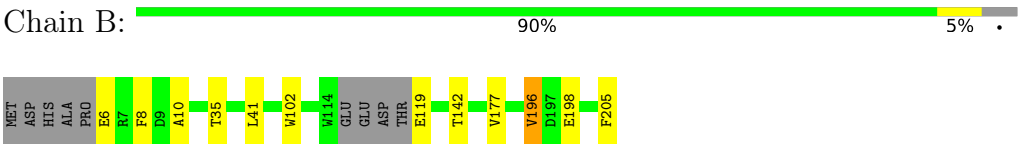
• Molecule 1: Microcompartments protein

Chain A:



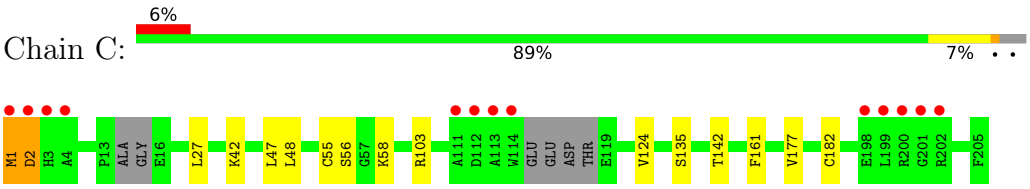
• Molecule 1: Microcompartments protein

Chain B:



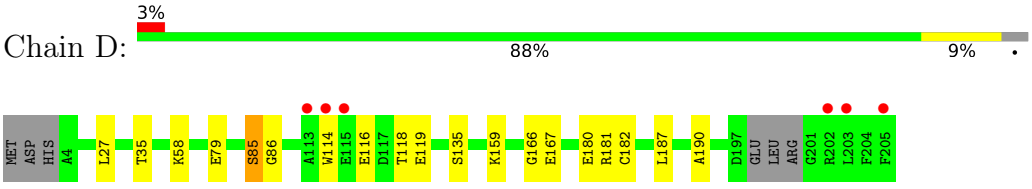
• Molecule 1: Microcompartments protein

Chain C:



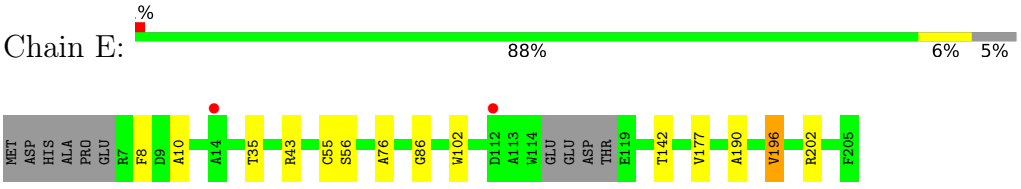
• Molecule 1: Microcompartments protein

Chain D:



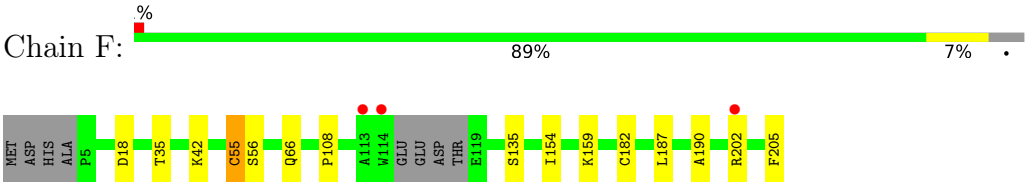
• Molecule 1: Microcompartments protein

Chain E:



• Molecule 1: Microcompartments protein

Chain F:


%
89%
7%
MET
ASP
HIS
ALA
P5
D18
T35
K42
C55
S56
Q66
P108
A113
W114
GLU
GLU
ASP
THR
E119
S135
I154
K159
C182
L187
A190
R202
F205

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	42.90Å 55.56Å 117.93Å 83.45° 81.22° 86.95°	Depositor
Resolution (Å)	38.62 – 1.80 38.62 – 1.81	Depositor EDS
% Data completeness (in resolution range)	89.5 (38.62-1.80) 86.7 (38.62-1.81)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.23 (at 1.81Å)	Xtriage
Refinement program	PHENIX 1.9_1690	Depositor
R, R_{free}	0.188 , 0.218 0.193 , 0.220	Depositor DCC
R_{free} test set	1787 reflections (2.11%)	wwPDB-VP
Wilson B-factor (Å^2)	20.5	Xtriage
Anisotropy	0.563	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.43 , 52.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	18044	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 18.47% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SF4

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	0.47	0/1467	0.68	0/1987
1	B	0.46	0/1485	0.65	0/2014
1	C	0.45	0/1515	0.66	0/2054
1	D	0.47	0/1503	0.62	0/2041
1	E	0.46	0/1476	0.63	0/2002
1	F	0.50	1/1493 (0.1%)	0.64	0/2025
All	All	0.47	1/8939 (0.0%)	0.65	0/12123

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	55	CYS	CB-SG	-5.31	1.73	1.81

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-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 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	1448	1477	1485	10	0
1	B	1464	1469	1496	11	0
1	C	1493	1464	1522	18	0
1	D	1481	1465	1501	19	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1455	1466	1490	10	0
1	F	1471	1468	1504	10	0
2	A	8	0	0	0	0
2	D	8	0	0	0	0
3	A	53	0	0	1	0
3	B	81	0	0	1	0
3	C	59	0	0	2	0
3	D	65	0	0	5	0
3	E	80	0	0	0	0
3	F	69	0	0	0	0
All	All	9235	8809	8998	66	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 66 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:1:MET:H1	1:C:1:MET:HE2	1.34	0.92
1:F:135:SER:OG	1:F:182:CYS:SG	2.37	0.82
1:D:116:GLU:CB	1:D:119:GLU:HG3	2.10	0.81
1:A:119:GLU:N	1:A:119:GLU:OE1	2.15	0.79
1:B:198:GLU:HG2	1:C:1:MET:HG2	1.68	0.76

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	188/205 (92%)	182 (97%)	6 (3%)	0	100	100
1	B	192/205 (94%)	186 (97%)	6 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	193/205 (94%)	188 (97%)	5 (3%)	0	100	100
1	D	195/205 (95%)	187 (96%)	7 (4%)	1 (0%)	29	15
1	E	191/205 (93%)	184 (96%)	6 (3%)	1 (0%)	29	15
1	F	193/205 (94%)	187 (97%)	6 (3%)	0	100	100
All	All	1152/1230 (94%)	1114 (97%)	36 (3%)	2 (0%)	47	33

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	86	GLY
1	E	86	GLY

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	148/156 (95%)	146 (99%)	2 (1%)	67	59
1	B	148/156 (95%)	146 (99%)	2 (1%)	67	59
1	C	152/156 (97%)	150 (99%)	2 (1%)	69	62
1	D	150/156 (96%)	148 (99%)	2 (1%)	69	62
1	E	147/156 (94%)	146 (99%)	1 (1%)	84	81
1	F	149/156 (96%)	149 (100%)	0	100	100
All	All	894/936 (96%)	885 (99%)	9 (1%)	76	71

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	180	GLU
1	E	196	VAL
1	B	196	VAL
1	C	1	MET
1	C	2	ASP

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

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

2 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
2	SF4	D	301	1,3	0,12,12	-	-	-		
2	SF4	A	301	1,3	0,12,12	-	-	-		

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
2	SF4	D	301	1,3	-	-	0/6/5/5
2	SF4	A	301	1,3	-	-	0/6/5/5

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.

No monomer is involved in short contacts.

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, 95^{th} 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	194/205 (94%)	-0.01	7 (3%) 42 37	20, 29, 70, 91	0
1	B	196/205 (95%)	-0.20	0 100 100	19, 26, 50, 76	0
1	C	199/205 (97%)	0.04	13 (6%) 18 15	21, 29, 79, 101	0
1	D	199/205 (97%)	-0.11	6 (3%) 50 44	19, 28, 65, 96	0
1	E	195/205 (95%)	-0.16	2 (1%) 82 80	20, 27, 56, 86	0
1	F	197/205 (96%)	-0.17	3 (1%) 73 70	18, 26, 64, 91	0
All	All	1180/1230 (95%)	-0.10	31 (2%) 56 51	18, 27, 64, 101	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1	MET	6.1
1	C	114	TRP	5.9
1	A	200	ARG	4.2
1	D	203	LEU	3.9
1	D	113	ALA	3.6

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 monosaccharides 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. The B-factors column lists the minimum, median, 95^{th} 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	B-factors(Å^2)	Q<0.9
2	SF4	A	301	8/8	0.97	0.08	23,29,32,36	8
2	SF4	D	301	8/8	0.97	0.09	22,26,30,34	8

6.5 Other polymers

There are no such residues in this entry.