



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 1, 2016 – 03:21 AM EDT

PDB ID : 5EIH
Title : Structural determination of an protein complex of a Fab with increased solubility
Authors : Bailey, L.J.; Kossiakoff, A.A.; Chaperone-Enabled Studies of Epigenetic Regulation Enzymes (CEBS)
Deposited on : 2015-10-29
Resolution : 2.44 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

<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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028320
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)	:	rb-20028320

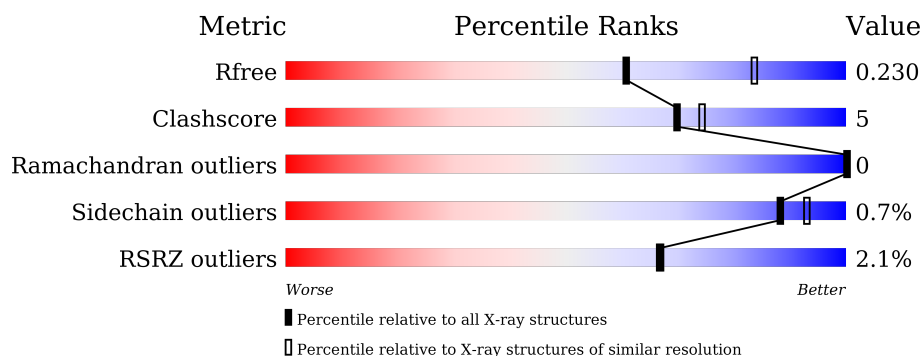
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.44 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	1003 (2.46-2.42)
Clashscore	102246	1071 (2.46-2.42)
Ramachandran outliers	100387	1065 (2.46-2.42)
Sidechain outliers	100360	1065 (2.46-2.42)
RSRZ outliers	91569	1005 (2.46-2.42)

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 $\leq 5\%$. 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	235	<div> <div>85%</div> <div>9%</div> <div>6%</div> </div>
1	H	235	<div> <div>%</div> <div>92%</div> <div>.</div> <div>.</div> </div>
2	B	215	<div> <div>89%</div> <div>10%</div> <div>.</div> </div>
2	L	215	<div> <div>89%</div> <div>9%</div> <div>.</div> </div>
3	G	156	<div> <div>5%</div> <div>72%</div> <div>19%</div> <div>8%</div> </div>
3	I	156	<div> <div>8%</div> <div>63%</div> <div>11%</div> <div>26%</div> </div>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 9056 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 Fab Heavy Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	220	Total	C	N	O	S	0	0	0
			1655	1049	276	324	6			
1	H	225	Total	C	N	O	S	0	0	0
			1681	1063	281	331	6			

- Molecule 2 is a protein called Fab Light Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	212	Total	C	N	O	S	1	0	0
			1606	1000	271	330	5			
2	L	212	Total	C	N	O	S	0	0	0
			1614	1005	272	332	5			

- Molecule 3 is a protein called Histone chaperone ASF1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	144	Total	C	N	O	S	0	0	0
			1155	748	184	221	2			
3	I	116	Total	C	N	O	S	0	0	0
			916	598	148	168	2			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	103	Total	O	0	0
			103	103		
4	B	33	Total	O	0	0
			33	33		
4	G	20	Total	O	0	0
			20	20		
4	H	120	Total	O	0	0
			120	120		

Continued on next page...

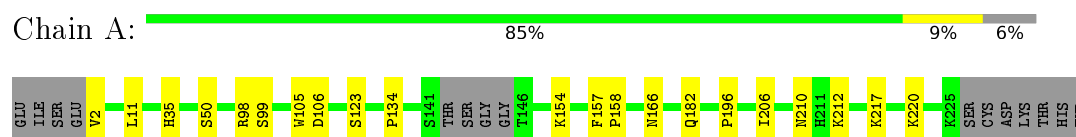
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	I	14	Total 14	O 14	0	0
4	L	139	Total 139	O 139	0	0

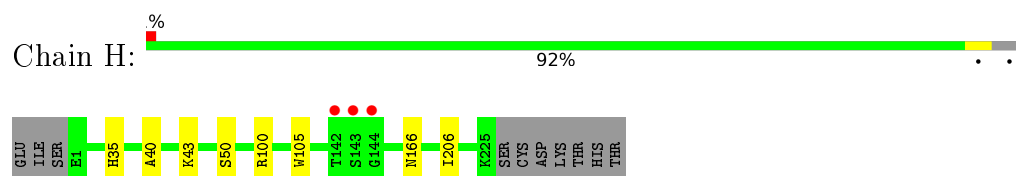
3 Residue-property plots [i](#)

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.

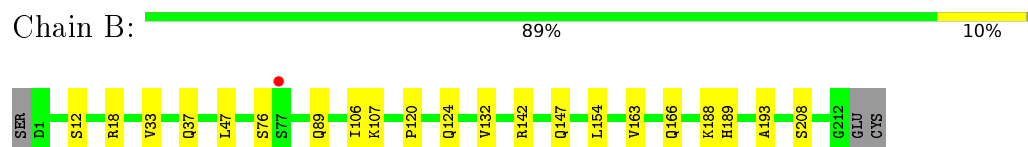
• Molecule 1: Fab Heavy Chain



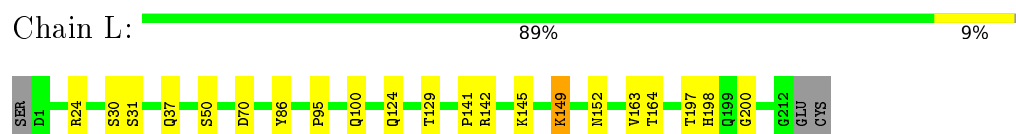
• Molecule 1: Fab Heavy Chain



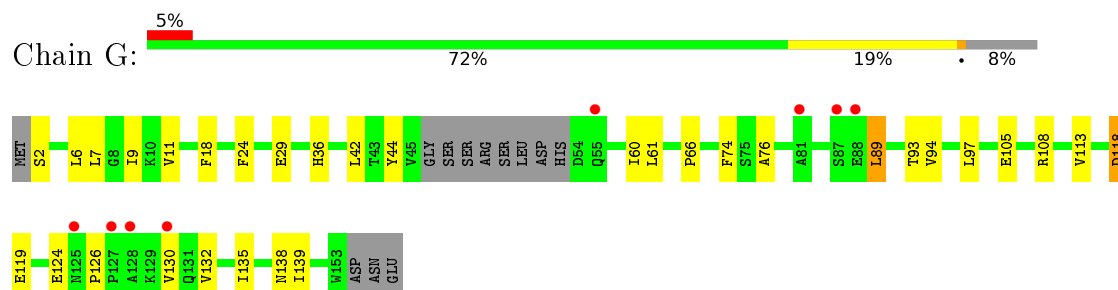
• Molecule 2: Fab Light Chain



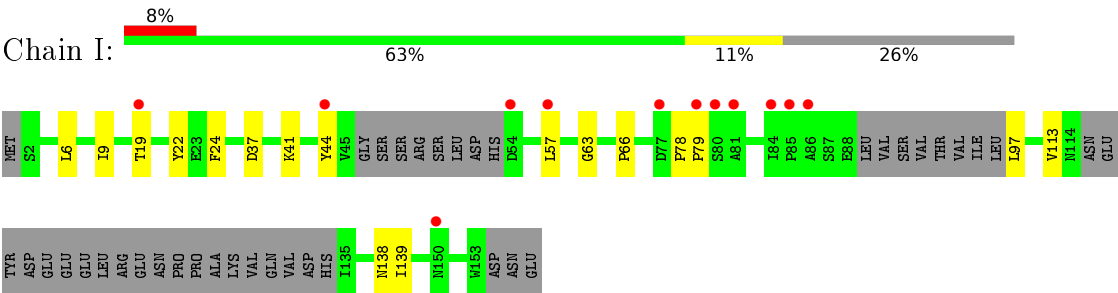
• Molecule 2: Fab Light Chain



• Molecule 3: Histone chaperone ASF1



● Molecule 3: Histone chaperone ASF1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	81.49Å 62.95Å 161.83Å 90.00° 94.00° 90.00°	Depositor
Resolution (Å)	46.38 – 2.44 46.38 – 2.44	Depositor EDS
% Data completeness (in resolution range)	99.4 (46.38-2.44) 95.8 (46.38-2.44)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.24 (at 2.45Å)	Xtriage
Refinement program	PHENIX (1.10_2152: ???)	Depositor
R, R_{free}	0.185 , 0.233 0.183 , 0.230	Depositor DCC
R_{free} test set	2966 reflections (5.03%)	DCC
Wilson B-factor (Å ²)	37.0	Xtriage
Anisotropy	0.284	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 41.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	9056	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.60% 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	0.30	0/1699	0.51	0/2319
1	H	0.32	0/1726	0.51	0/2357
2	B	0.28	0/1639	0.52	0/2227
2	L	0.32	0/1647	0.53	0/2236
3	G	0.28	0/1182	0.47	0/1614
3	I	0.28	0/937	0.50	0/1274
All	All	0.30	0/8830	0.51	0/12027

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	1655	0	1595	13	0
1	H	1681	0	1619	11	0
2	B	1606	0	1559	11	0
2	L	1614	0	1574	15	0
3	G	1155	0	1138	25	0
3	I	916	0	897	10	0
4	A	103	0	0	1	0
4	B	33	0	0	1	0
4	G	20	0	0	2	0
4	H	120	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	I	14	0	0	0	0
4	L	139	0	0	2	0
All	All	9056	0	8382	77	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 5.

The worst 5 of 77 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)
3:G:60:ILE:CG2	1:H:100:ARG:HH21	1.79	0.95
3:G:60:ILE:CG2	1:H:100:ARG:NH2	2.43	0.82
2:L:198:HIS:CD2	2:L:200:GLY:H	2.05	0.75
3:G:60:ILE:HG23	1:H:100:ARG:HH21	1.55	0.72
1:A:134:PRO:HD3	1:A:220:LYS:HE2	1.72	0.71

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	216/235 (92%)	214 (99%)	2 (1%)	0	100	100
1	H	223/235 (95%)	220 (99%)	3 (1%)	0	100	100
2	B	210/215 (98%)	204 (97%)	6 (3%)	0	100	100
2	L	210/215 (98%)	203 (97%)	7 (3%)	0	100	100
3	G	140/156 (90%)	131 (94%)	9 (6%)	0	100	100
3	I	108/156 (69%)	106 (98%)	2 (2%)	0	100	100
All	All	1107/1212 (91%)	1078 (97%)	29 (3%)	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	182/196 (93%)	180 (99%)	2 (1%)	80	87
1	H	184/196 (94%)	184 (100%)	0	100	100
2	B	185/190 (97%)	185 (100%)	0	100	100
2	L	187/190 (98%)	186 (100%)	1 (0%)	92	95
3	G	132/144 (92%)	129 (98%)	3 (2%)	58	73
3	I	101/144 (70%)	100 (99%)	1 (1%)	82	88
All	All	971/1060 (92%)	964 (99%)	7 (1%)	88	93

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	118	ASP
2	L	149	LYS
3	G	138	ASN
1	A	196	PRO
3	I	138	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	124	GLN
2	L	198	HIS
1	H	166	ASN
1	A	166	ASN
2	B	138	ASN

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	220/235 (93%)	-0.42	0 100 100	22, 37, 60, 89	0
1	H	225/235 (95%)	-0.38	3 (1%) 79 80	20, 33, 60, 136	0
2	B	212/215 (98%)	-0.10	1 (0%) 91 92	28, 53, 79, 89	1 (0%)
2	L	212/215 (98%)	-0.51	0 100 100	19, 30, 51, 68	1 (0%)
3	G	144/156 (92%)	0.39	8 (5%) 28 26	38, 59, 105, 119	0
3	I	116/156 (74%)	0.34	12 (10%) 9 7	40, 63, 89, 102	0
All	All	1129/1212 (93%)	-0.19	24 (2%) 67 67	19, 43, 83, 136	2 (0%)

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	143	SER	6.6
3	G	128	ALA	5.0
3	I	81	ALA	4.2
1	H	144	GLY	3.7
3	I	80	SER	3.7

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.