wwPDB X-ray Structure Validation Summary Report (i)

Jun 12, 2024 – 09:07 PM EDT

PDB ID	:	3HIF
Title	:	The crystal structure of apo wild type CAP at 3.6 A resolution.
Authors	:	Steitz, T.A.; Sharma, H.; Wang, J.; Kong, J.; Yu, S.
Deposited on	:	2009-05-19
Resolution	:	3.59 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:

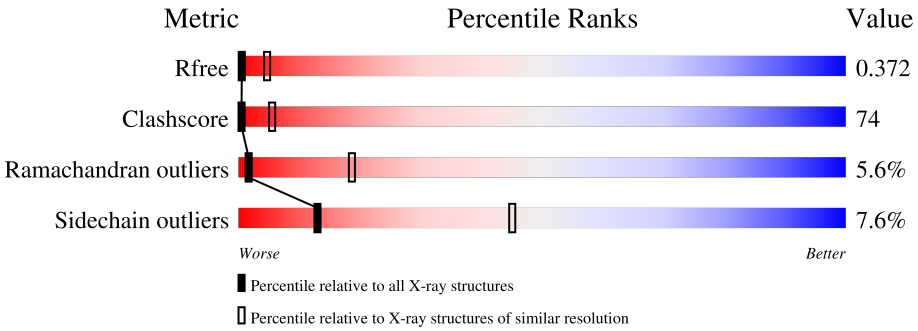
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 3.59 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}	130704	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)

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%

Mol	Chain	Length	Quality of chain
1	A	210	43% 39% 9% • 5%
1	B	210	45% 35% 12% • 5%
1	C	210	40% 40% 11% • 5%
1	D	210	53% 33% 8% • 5%
1	E	210	38% 44% 10% • 5%
1	F	210	36% 41% 12% 5% 5%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 9456 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 Catabolite gene activator.

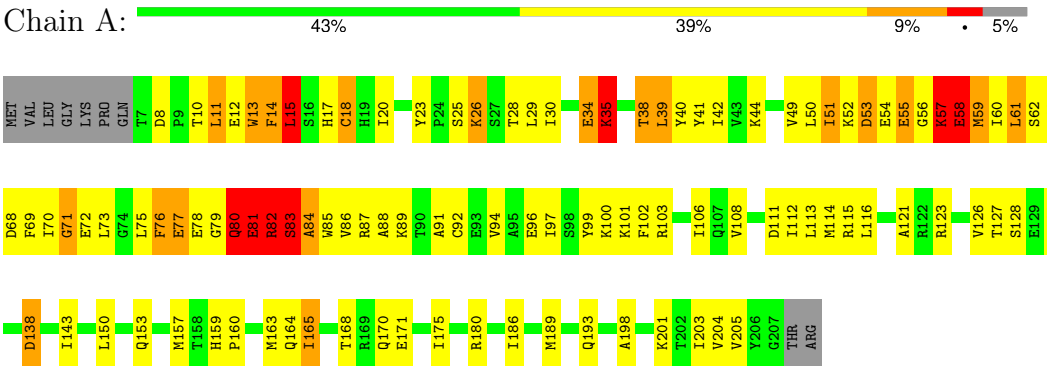
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	199	Total	C	N	O	S	0	0	0
			1576	998	275	294	9			
1	B	199	Total	C	N	O	S	0	0	0
			1576	998	275	294	9			
1	C	199	Total	C	N	O	S	0	0	0
			1576	998	275	294	9			
1	D	199	Total	C	N	O	S	0	0	0
			1576	998	275	294	9			
1	E	199	Total	C	N	O	S	0	0	0
			1576	998	275	294	9			
1	F	199	Total	C	N	O	S	0	0	0
			1576	998	275	294	9			

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. 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. 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.

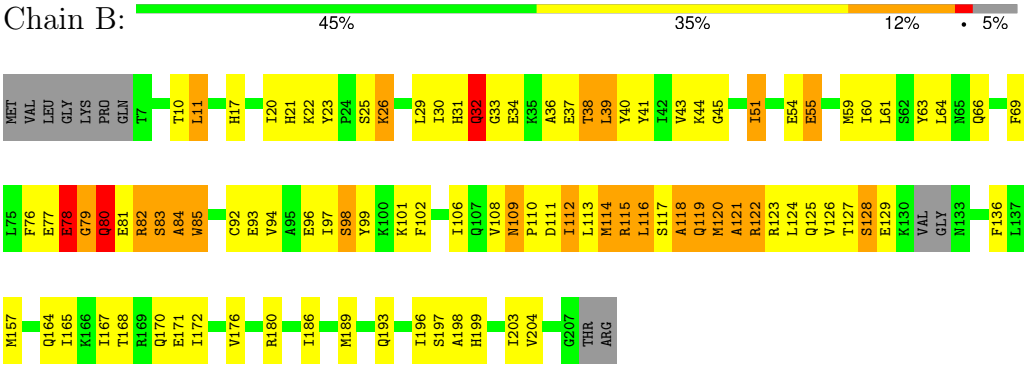
• Molecule 1: Catabolite gene activator

Chain A:



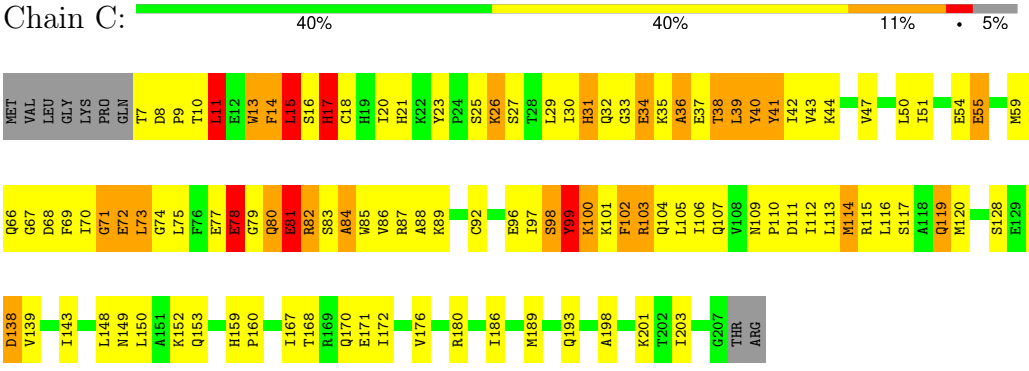
• Molecule 1: Catabolite gene activator

Chain B:



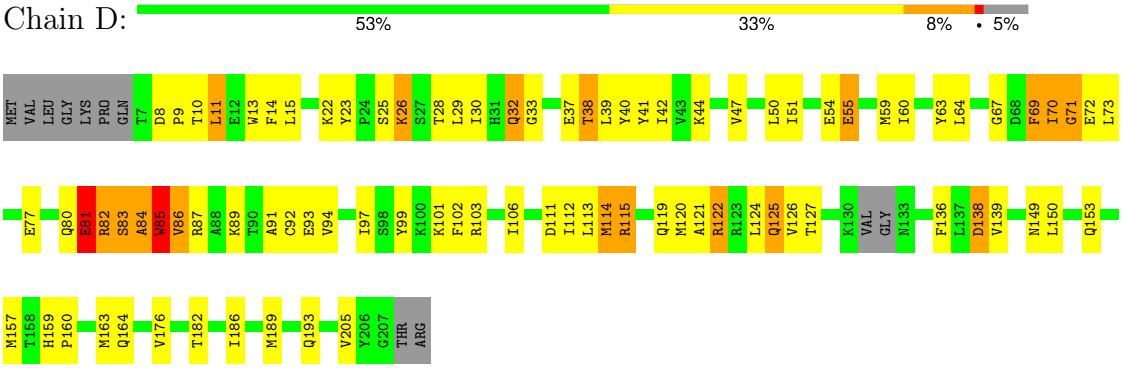
• Molecule 1: Catabolite gene activator

Chain C:



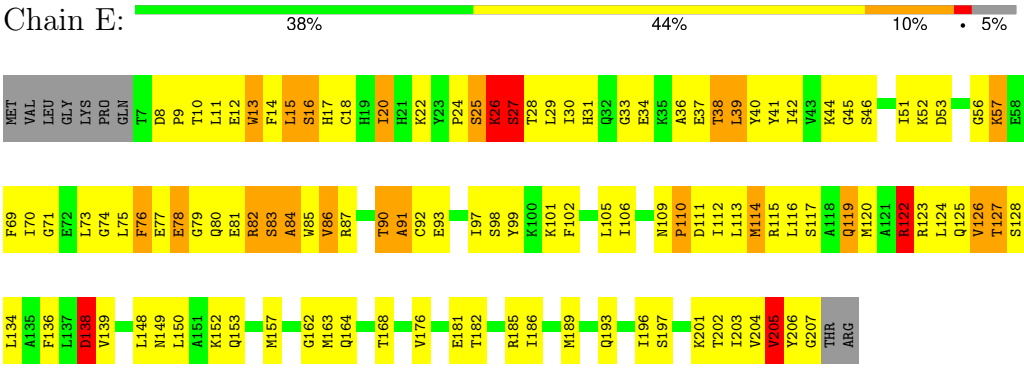
• Molecule 1: Catabolite gene activator

Chain D:



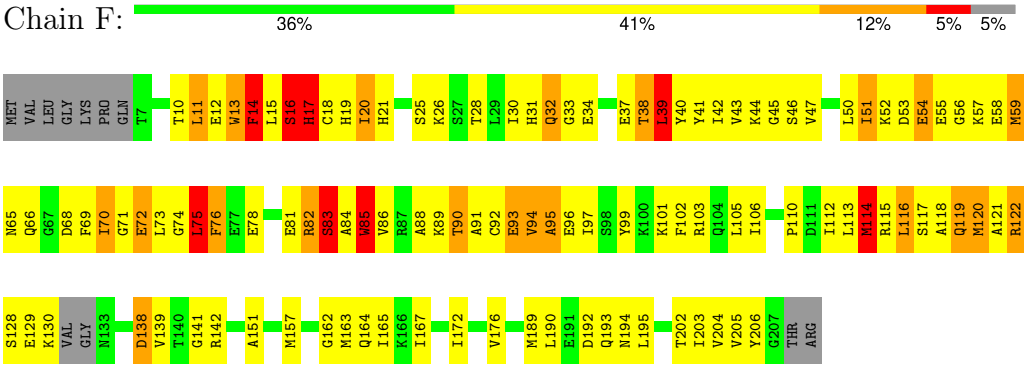
• Molecule 1: Catabolite gene activator

Chain E:



• Molecule 1: Catabolite gene activator

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	125.29Å 125.29Å 224.68Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.06 – 3.59 48.05 – 3.59	Depositor EDS
% Data completeness (in resolution range)	98.9 (48.06-3.59) 98.9 (48.05-3.59)	Depositor EDS
R_{merge}	0.46	Depositor
R_{sym}	0.53	Depositor
$< I/\sigma(I) >$ [1]	1.27 (at 3.57Å)	Xtriage
Refinement program	REFMAC 5.4.0077	Depositor
R, R_{free}	0.295 , 0.318 0.362 , 0.372	Depositor DCC
R_{free} test set	1232 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	100.5	Xtriage
Anisotropy	0.139	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , -10.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.043 for -h,-k,l	Xtriage
F_o,F_c correlation	0.83	EDS
Total number of atoms	9456	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.02% 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

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	0.54	3/1600 (0.2%)	1.07	19/2154 (0.9%)
1	B	0.71	3/1600 (0.2%)	0.82	11/2154 (0.5%)
1	C	0.65	3/1600 (0.2%)	1.72	23/2154 (1.1%)
1	D	0.54	3/1600 (0.2%)	0.81	13/2154 (0.6%)
1	E	0.51	2/1600 (0.1%)	1.05	16/2154 (0.7%)
1	F	0.54	3/1600 (0.2%)	1.12	28/2154 (1.3%)
All	All	0.59	17/9600 (0.2%)	1.14	110/12924 (0.9%)

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	A	0	2
1	B	0	1
1	C	1	3
1	D	0	2
1	E	0	1
1	F	0	2
All	All	1	11

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	112	ILE	CG1-CD1	20.25	2.90	1.50
1	C	78	GLU	C-N	-14.81	1.06	1.33
1	C	138	ASP	CG-OD1	11.91	1.52	1.25
1	A	138	ASP	CG-OD1	11.88	1.52	1.25
1	F	138	ASP	CG-OD1	11.88	1.52	1.25

The worst 5 of 110 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	82	ARG	N-CA-CB	-40.78	37.20	110.60
1	C	81	GLU	CB-CA-C	-29.48	51.45	110.40
1	C	78	GLU	C-N-CA	26.64	178.25	122.30
1	C	102	PHE	CB-CA-C	-19.82	70.76	110.40
1	E	82	ARG	CB-CA-C	-19.36	71.67	110.40

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	C	81	GLU	CA

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	138	ASP	Sidechain
1	A	58	GLU	Mainchain
1	B	138	ASP	Sidechain
1	C	78	GLU	Mainchain
1	C	99	TYR	Peptide

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	1576	0	1608	196	0
1	B	1576	0	1609	221	0
1	C	1576	0	1608	229	0
1	D	1576	0	1608	131	0
1	E	1576	0	1608	315	0
1	F	1576	0	1608	366	0
All	All	9456	0	9649	1412	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 74.

The worst 5 of 1412 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:41:TYR:HD1	1:C:69:PHE:CE1	0.96	1.65
1:C:41:TYR:CD1	1:C:69:PHE:CE1	1.86	1.63
1:F:51:ILE:HD12	1:F:61:LEU:CG	1.29	1.62
1:E:134:LEU:HG	1:F:195:LEU:CD1	1.21	1.57
1:F:68:ASP:CA	1:F:119:GLN:HE22	1.16	1.56

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	195/210 (93%)	163 (84%)	21 (11%)	11 (6%)	2	19
1	B	195/210 (93%)	164 (84%)	18 (9%)	13 (7%)	1	15
1	C	195/210 (93%)	154 (79%)	29 (15%)	12 (6%)	1	17
1	D	195/210 (93%)	173 (89%)	18 (9%)	4 (2%)	7	40
1	E	195/210 (93%)	159 (82%)	25 (13%)	11 (6%)	2	19
1	F	195/210 (93%)	149 (76%)	32 (16%)	14 (7%)	1	13
All	All	1170/1260 (93%)	962 (82%)	143 (12%)	65 (6%)	2	19

5 of 65 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	14	PHE
1	A	15	LEU
1	A	35	LYS
1	A	57	LYS
1	A	82	ARG

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	172/181 (95%)	157 (91%)	15 (9%)	10 41
1	B	172/181 (95%)	156 (91%)	16 (9%)	9 38
1	C	172/181 (95%)	159 (92%)	13 (8%)	13 45
1	D	172/181 (95%)	164 (95%)	8 (5%)	26 61
1	E	172/181 (95%)	163 (95%)	9 (5%)	23 58
1	F	172/181 (95%)	155 (90%)	17 (10%)	8 35
All	All	1032/1086 (95%)	954 (92%)	78 (8%)	13 45

5 of 78 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	86	VAL
1	F	90	THR
1	E	114	MET
1	F	38	THR
1	F	122	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 49 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	164	GLN
1	E	119	GLN
1	D	174	GLN
1	E	66	GLN
1	E	153	GLN

5.3.3 RNA (i)

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 monosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1
1	D	1
1	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	58:GLU	C	59:MET	N	1.13
1	D	85:TRP	C	86:VAL	N	1.11
1	C	78:GLU	C	79:GLY	N	1.06

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.