wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 07:48 PM GMT

PDB ID	:	1H8A
Title	:	CRYSTAL STRUCTURE OF TERNARY PROTEIN-DNA COMPLEX3
Authors	:	Tahirov, T.H.; Ogata, K.
Deposited on	:	2001-01-31
Resolution	:	2.23 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

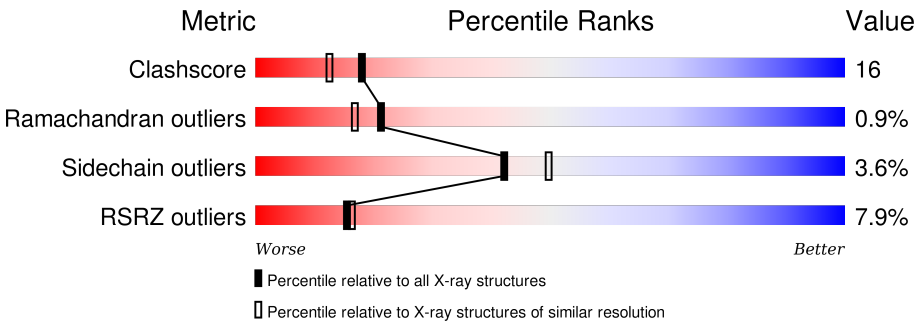
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.23 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))
Clashscore	102246	1764 (2.26-2.22)
Ramachandran outliers	100387	1724 (2.26-2.22)
Sidechain outliers	100360	1724 (2.26-2.22)
RSRZ outliers	91569	1616 (2.26-2.22)

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 <=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	78	5%, 73%, 14%, 13%
1	B	78	%, 72%, 12%, •, 14%
2	C	128	14%, 55%, 25%, •, 18%
3	D	26	46%, 46%, 8%
4	E	26	31%, 69%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3276 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 CAAT/ENHANCER BINDING PROTEIN BETA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	68	579	354	117	107	1	0	0	0
1	B	67	572	349	116	106	1	0	0	0

- Molecule 2 is a protein called MYB TRANSFORMING PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	105	894	557	180	155	2	0	0	0

- Molecule 3 is a DNA chain called DNA(5'-(*GP*AP*TP*GP*TP*GP*GP*CP*GP*CP*AP* AP*TP*CP*CP*TP*TP*AP*AP*CP*GP*GP*AP*CP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	D	26	533	254	100	154	25	0	0	0

- Molecule 4 is a DNA chain called DNA(5'-(*CP*CP*AP*GP*TP*CP*CP*GP*TP*TP*AP* AP*GP*GP*AP*TP*TP*GP*CP*GP*CP*CP*AP*CP*AP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	E	26	527	252	96	154	25	0	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	42	42	42	0	0
5	B	46	46	46	0	0

Continued on next page...

Continued from previous page...

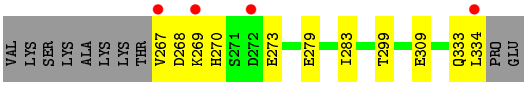
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	C	7	7	7	0	0
			Total	O		
5	D	42	42	42	0	0
			Total	O		
5	E	34	34	34	0	0

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.

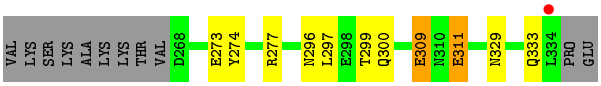
• Molecule 1: CAAT/ENHANCER BINDING PROTEIN BETA

Chain A:



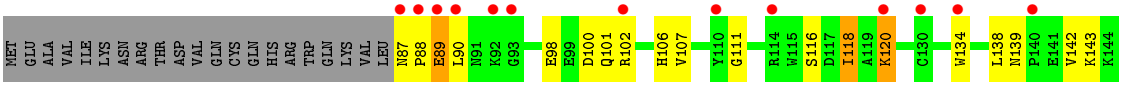
• Molecule 1: CAAT/ENHANCER BINDING PROTEIN BETA

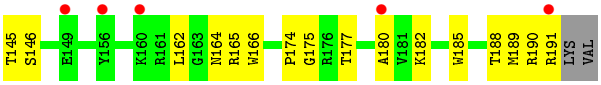
Chain B:



• Molecule 2: MYB TRANSFORMING PROTEIN

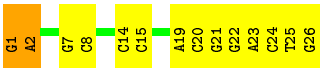
Chain C:





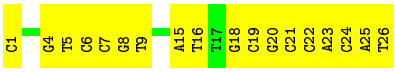
• Molecule 3: DNA(5'-(*GP*AP*TP*GP*TP*GP*GP*CP*GP*CP*AP* AP*TP*CP*CP*TP* TP*AP*AP*CP*GP*GP*AP*CP*TP*G)-3')

Chain D:



• Molecule 4: DNA(5'-(*CP*CP*AP*GP*TP*CP*CP*GP*TP*TP*AP* AP*GP*GP*AP*TP* TP*GP*CP*GP*CP*CP*AP*CP*AP*T)-3')

Chain E:


C1
G4
T5
C6
C7
G8
T9
A15
T16
T17
G18
C19
G20
C21
C22
A23
C24
A25
T26

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	109.86Å 166.70Å 39.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.85 – 2.23 19.85 – 2.10	Depositor EDS
% Data completeness (in resolution range)	97.9 (19.85-2.23) 97.2 (19.85-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	2.76 (at 2.11Å)	Xtriage
Refinement program	CNS 0.9	Depositor
R, R_{free}	0.245 , 0.288 0.248 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	49.2	Xtriage
Anisotropy	0.184	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 51.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	18 of 42089 reflections (0.043%)	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3276	wwPDB-VP
Average B, all atoms (Å^2)	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.05% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.37	0/582	0.52	0/772
1	B	0.37	0/575	0.52	0/762
2	C	0.27	0/918	0.49	0/1236
3	D	0.48	0/598	0.87	0/922
4	E	0.46	0/590	0.81	0/908
All	All	0.39	0/3263	0.66	0/4600

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
3	D	0	2
4	E	0	1
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	1	DG	Sidechain
3	D	2	DA	Sidechain
4	E	1	DC	Sidechain

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	579	0	609	14	0
1	B	572	0	600	10	0
2	C	894	0	884	40	0
3	D	533	0	294	14	0
4	E	527	0	294	16	0
5	A	42	0	0	1	0
5	B	46	0	0	1	0
5	C	7	0	0	1	0
5	D	42	0	0	0	0
5	E	34	0	0	0	0
All	All	3276	0	2681	91	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 16.

The worst 5 of 91 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)
2:C:87:ASN:HB3	2:C:88:PRO:HD2	1.47	0.95
2:C:98:GLU:HG3	2:C:102:ARG:HH12	1.36	0.90
4:E:8:DG:H1'	4:E:9:DT:H5'	1.60	0.83
2:C:87:ASN:HB3	2:C:88:PRO:CD	2.09	0.83
3:D:14:DC:H1'	3:D:15:DC:H5'	1.65	0.78

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	66/78 (85%)	65 (98%)	1 (2%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	65/78 (83%)	64 (98%)	1 (2%)	0	100	100
2	C	103/128 (80%)	91 (88%)	10 (10%)	2 (2%)	10	4
All	All	234/284 (82%)	220 (94%)	12 (5%)	2 (1%)	21	17

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	164	ASN
2	C	188	THR

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	65/74 (88%)	63 (97%)	2 (3%)	47	57
1	B	64/74 (86%)	61 (95%)	3 (5%)	32	35
2	C	94/116 (81%)	91 (97%)	3 (3%)	46	56
All	All	223/264 (84%)	215 (96%)	8 (4%)	42	50

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	309	GLU
2	C	120	LYS
2	C	89	GLU
1	B	273	GLU
1	B	311	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	333	GLN
1	B	300	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	C	101	GLN
2	C	108	GLN
2	C	136	ASN

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 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 (i)

6.1 Protein, DNA and RNA chains (i)

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	68/78 (87%)	0.23	4 (5%) 26 27	30, 52, 102, 113	0
1	B	67/78 (85%)	0.23	1 (1%) 76 78	27, 47, 77, 96	0
2	C	105/128 (82%)	1.18	18 (17%) 2 2	54, 87, 111, 137	0
3	D	26/26 (100%)	-0.22	0 100 100	36, 55, 71, 72	0
4	E	26/26 (100%)	-0.28	0 100 100	37, 53, 71, 80	0
All	All	292/336 (86%)	0.49	23 (7%) 15 16	27, 62, 107, 137	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	87	ASN	8.7
2	C	90	LEU	7.4
2	C	88	PRO	7.2
2	C	93	GLY	6.9
2	C	110	TYR	6.5

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

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

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.