wwPDB X-ray Structure Validation Summary Report (i)

May 28, 2020 – 07:35 pm BST

PDB ID	:	1H88
Title	:	CRYSTAL STRUCTURE OF TERNARY PROTEIN-DNA COMPLEX1
Authors	:	Tahirov, T.H.; Ogata, K.
Deposited on	:	2001-01-29
Resolution	:	2.80 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 following versions of software and data (see references (1)) were used in the production of this report:

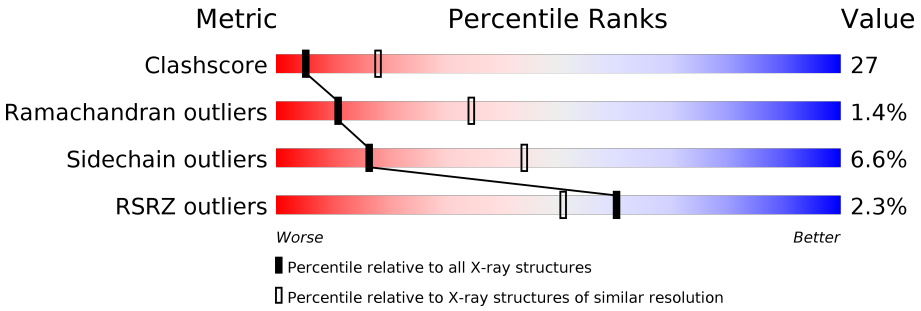
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.80 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))
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 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 <=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	58% 27% 5% 10%
1	B	78	6% 59% 32% 9%
2	C	159	48% 41% 6%
3	D	26	27% 69%
4	E	26	92%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3576 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 CCAAT/ENHANCER BINDING PROTEIN BETA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	70	Total	C	N	O	S	0	0	0
			596	364	119	112	1			
1	B	71	Total	C	N	O	S	0	0	0
			601	367	120	113	1			

- Molecule 2 is a protein called MYB PROTO-ONCOGENE PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	152	Total	C	N	O	S	0	0	0
			1293	813	252	225	3			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	35	MET	LYS	engineered mutation	UNP P17676
C	36	GLY	ARG	engineered mutation	UNP P17676

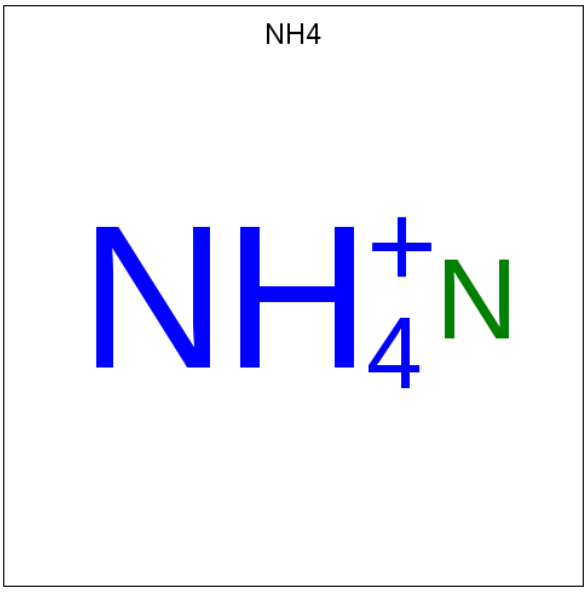
- Molecule 3 is a DNA chain called DNA(5'-(*GP*AP*TP*GP*TP*GP*GP*CP*GP*CP*AP* AP*TP*CP*CP*TP*TP*AP*AP*CP*GP*GP*AP*CP*TP*G)-3').

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

- 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
4	E	26	Total	C	N	O	P	0	0	0
			527	252	96	154	25			

- Molecule 5 is AMMONIUM ION (three-letter code: NH4) (formula: H_4N).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	1	Total N 1 1	0	0

- Molecule 6 is water.

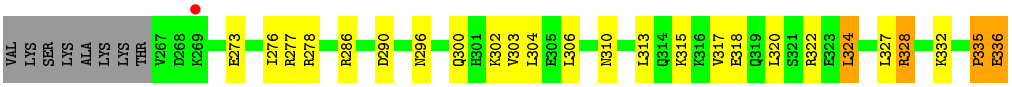
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total O 1 1	0	0
6	B	5	Total O 5 5	0	0
6	C	7	Total O 7 7	0	0
6	D	6	Total O 6 6	0	0
6	E	6	Total O 6 6	0	0

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 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.

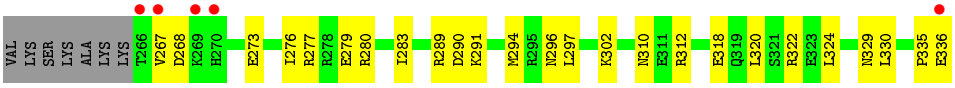
• Molecule 1: CCAAT/ENHANCER BINDING PROTEIN BETA

Chain A:



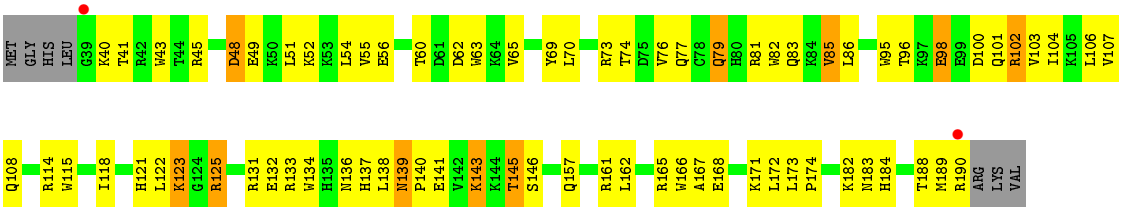
• Molecule 1: CCAAT/ENHANCER BINDING PROTEIN BETA

Chain B:



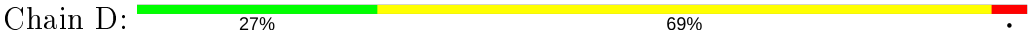
• Molecule 2: MYB PROTO-ONCOGENE PROTEIN

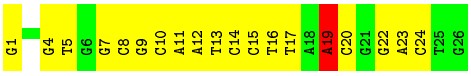
Chain C:



• 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')

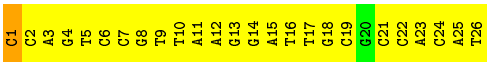
Chain D:





• 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
C2
A3
G4
T5
C6
C7
G8
T9
T10
A11
A12
G13
G14
A15
T16
T17
G18
C19
G20
C21
C22
A23
C24
A25
T26

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	41.73Å 156.97Å 55.87Å 90.00° 100.16° 90.00°	Depositor
Resolution (Å)	19.87 – 2.80 19.87 – 2.70	Depositor EDS
% Data completeness (in resolution range)	96.1 (19.87-2.80) 94.5 (19.87-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	2.66 (at 2.71Å)	Xtriage
Refinement program	CNS 0.9	Depositor
R, R_{free}	0.222 , 0.277 0.220 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	63.2	Xtriage
Anisotropy	0.960	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 69.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3576	wwPDB-VP
Average B, all atoms (Å^2)	80.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.12% 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: NH4

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	0.38	0/600	0.54	0/796
1	B	0.36	0/605	0.57	0/803
2	C	0.36	0/1325	0.57	0/1786
3	D	0.55	0/598	0.90	1/922 (0.1%)
4	E	0.56	0/590	0.77	0/908
All	All	0.43	0/3718	0.67	1/5215 (0.0%)

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.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	19	DA	N9-C1'-C2'	5.02	122.14	112.60

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	1	DG	Sidechain
3	D	19	DA	Sidechain

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
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	596	0	622	23	0
1	B	601	0	624	20	0
2	C	1293	0	1302	79	0
3	D	533	0	294	22	0
4	E	527	0	294	42	0
5	C	1	0	0	0	0
6	A	1	0	0	0	0
6	B	5	0	0	3	0
6	C	7	0	0	0	0
6	D	6	0	0	0	0
6	E	6	0	0	0	0
All	All	3576	0	3136	178	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 27.

The worst 5 of 178 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)
4:E:8:DG:H1'	4:E:9:DT:H5'	1.39	1.05
4:E:6:DC:H2"	4:E:7:DC:H5'	1.44	0.99
3:D:12:DA:H1'	3:D:13:DT:H5'	1.54	0.90
4:E:16:DT:H2"	4:E:17:DT:H5'	1.51	0.90
1:A:335:PRO:O	1:A:336:GLU:HB3	1.72	0.89

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	68/78 (87%)	59 (87%)	8 (12%)	1 (2%)	10 33
1	B	69/78 (88%)	65 (94%)	4 (6%)	0	100 100
2	C	150/159 (94%)	126 (84%)	21 (14%)	3 (2%)	7 24
All	All	287/315 (91%)	250 (87%)	33 (12%)	4 (1%)	11 34

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	145	THR
1	A	335	PRO
2	C	85	VAL
2	C	174	PRO

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	67/74 (90%)	63 (94%)	4 (6%)	19 48
1	B	67/74 (90%)	63 (94%)	4 (6%)	19 48
2	C	138/144 (96%)	128 (93%)	10 (7%)	14 38
All	All	272/292 (93%)	254 (93%)	18 (7%)	16 44

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

Mol	Chain	Res	Type
2	C	48	ASP
2	C	56	GLU
2	C	125	ARG
1	B	297	LEU
1	B	302	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	57	GLN
2	C	79	GLN
2	C	139	ASN
1	B	333	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)

Of 1 ligands modelled in this entry, 1 is modelled with single atom - leaving 0 for Mogul analysis.

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	70/78 (89%)	-0.15	1 (1%) 75 70	46, 77, 116, 132	0
1	B	71/78 (91%)	-0.05	5 (7%) 16 9	59, 82, 130, 140	0
2	C	152/159 (95%)	-0.05	2 (1%) 77 72	45, 87, 119, 141	0
3	D	26/26 (100%)	-0.61	0 100 100	49, 65, 83, 91	0
4	E	26/26 (100%)	-0.64	0 100 100	54, 64, 76, 97	0
All	All	345/367 (94%)	-0.16	8 (2%) 60 51	45, 80, 124, 141	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	39	GLY	4.2
1	B	266	THR	3.9
1	B	269	LYS	3.8
1	B	336	GLU	3.7
1	B	270	HIS	3.0

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

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
5	NH4	C	1191	1/1	0.91	0.29	49,49,49,49	0

6.5 Other polymers (i)

There are no such residues in this entry.