wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 07:41 PM GMT

PDB ID	:	1GT0
Title	:	Crystal structure of a POU/HMG/DNA ternary complex
Authors	:	Remenyi, A.; Wilmanns, M.
Deposited on	:	2002-01-09
Resolution	:	2.60 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:

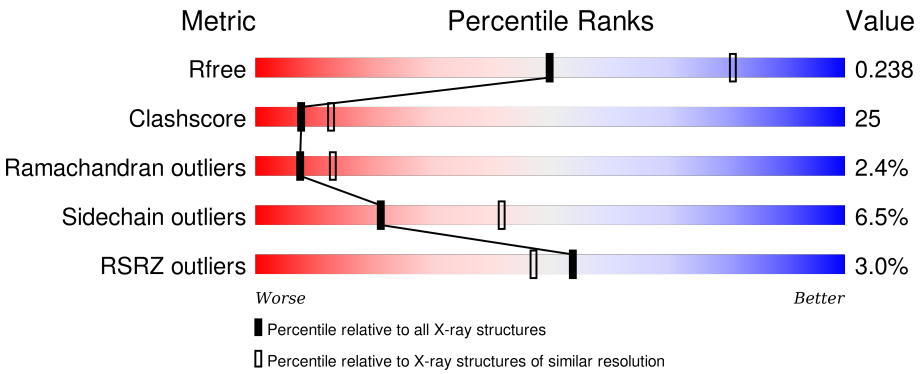
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	24	13% 21% 38% 29%
2	B	24	25% 71% •
3	C	159	% 52% 31% • 13%
4	D	80	8% 64% 31% 5%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2833 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 DNA chain called 5'-D(*TP*TP*CP*TP*TP*TP*GP*TP*TP*TP* GP*GP*AP* TP*GP*CP*TP*AP*AP*TP*GP*GP*GP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	24	493	238	83	149	23	0	0	0

- Molecule 2 is a DNA chain called 5'-D(*AP*TP*CP*CP*CP*AP*TP*TP*AP*GP* CP*A P*TP*CP*CP*AP*AP*AP*CP*AP*AP*AP*GP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	24	485	233	94	135	23	0	0	0

- Molecule 3 is a protein called OCTAMER-BINDING TRANSCRIPTION FACTOR 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	138	1118	702	202	210	4	0	0	1

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	61	SER	CYS	ENGINEERED MUTATION	UNP P14859
C	150	SER	CYS	ENGINEERED MUTATION	UNP P14859

- Molecule 4 is a protein called TRANSCRIPTION FACTOR SOX-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	80	647	402	132	108	5	0	0	1

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	25	25	25	0	0
			Total	O		
5	B	18	18	18	0	0
			Total	O		
5	C	30	30	30	0	0
			Total	O		
5	D	17	17	17	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: 5'-D(*TP*TP*CP*TP*TP*TP*GP*TP*TP*TP* GP*GP*AP* TP*GP*CP*TP* AP*AP*TP*GP*GP*GP*A)-3'

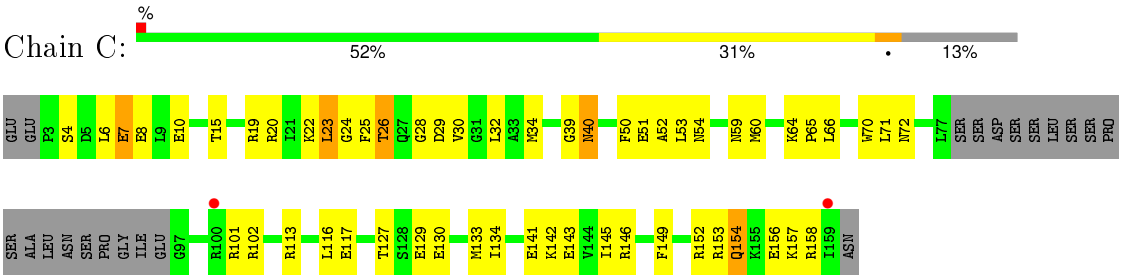
Chain A:

• Molecule 2: 5'-D(*AP*TP*CP*CP*CP*AP*TP*TP*AP*GP* CP*AP*TP*CP*CP*AP*AP* AP*CP*AP*AP*AP*GP*A)-3'

Chain B:

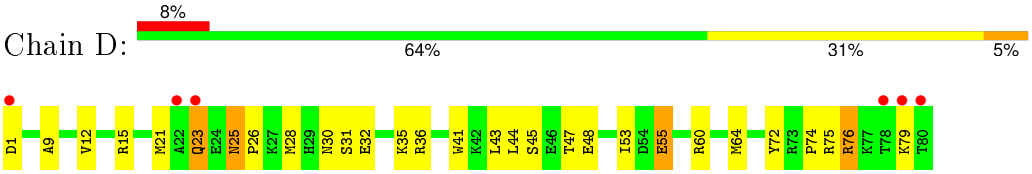
• Molecule 3: OCTAMER-BINDING TRANSCRIPTION FACTOR 1

Chain C:



• Molecule 4: TRANSCRIPTION FACTOR SOX-2

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	72.80Å 72.80Å 172.40Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 2.60 29.61 – 2.60	Depositor EDS
% Data completeness (in resolution range)	98.2 (30.00-2.60) 98.5 (29.61-2.60)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	1.76 (at 2.61Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.232 , 0.285 0.234 , 0.238	Depositor DCC
R_{free} test set	845 reflections (5.36%)	DCC
Wilson B-factor (Å^2)	66.7	Xtriage
Anisotropy	0.119	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 38.0	EDS
Estimated twinning fraction	0.033 for -h,-k,l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 16611 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2833	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.21% 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	0.84	0/551	1.59	17/851 (2.0%)
2	B	0.70	0/545	1.04	2/837 (0.2%)
3	C	0.58	0/1132	0.75	0/1512
4	D	0.48	0/662	0.62	0/888
All	All	0.64	0/2890	1.02	19/4088 (0.5%)

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	10

There are no bond length outliers.

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	17	DT	C2'-C3'-O3'	-8.55	84.38	112.60
1	A	14	DT	N1-C1'-C2'	7.35	126.56	112.60
1	A	8	DT	N1-C1'-C2'	6.39	124.75	112.60
1	A	7	DG	N9-C1'-C2'	6.31	124.58	112.60
1	A	19	DA	N9-C1'-C2'	6.26	124.49	112.60

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	13	DA	Sidechain
1	A	14	DT	Sidechain
1	A	15	DG	Sidechain
1	A	6	DT	Sidechain

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	7	DG	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	493	0	277	35	0
2	B	485	0	270	26	0
3	C	1118	0	1132	50	0
4	D	647	0	632	26	0
5	A	25	0	0	0	0
5	B	18	0	0	0	0
5	C	30	0	0	3	0
5	D	17	0	0	2	0
All	All	2833	0	2311	127	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 25.

The worst 5 of 127 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:B:25:DA:H3'	2:B:26:DT:H5"	1.30	1.09
1:A:13:DA:H2"	1:A:14:DT:O5'	1.53	1.07
2:B:25:DA:H3'	2:B:26:DT:C5'	1.90	1.02
3:C:26:THR:HG22	3:C:29:ASP:H	1.26	0.96
1:A:17:DT:H2"	1:A:18:DA:C8	2.03	0.93

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
3	C	134/159 (84%)	126 (94%)	6 (4%)	2 (2%)	13 26
4	D	78/80 (98%)	68 (87%)	7 (9%)	3 (4%)	4 5
All	All	212/239 (89%)	194 (92%)	13 (6%)	5 (2%)	7 13

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	23	LEU
4	D	79	LYS
4	D	21	MET
3	C	157	LYS
4	D	76	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
3	C	122/145 (84%)	115 (94%)	7 (6%)	25 49
4	D	64/73 (88%)	59 (92%)	5 (8%)	16 30
All	All	186/218 (85%)	174 (94%)	12 (6%)	21 42

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	59	ASN
3	C	154	GLN
4	D	31	SER
3	C	40	ASN
4	D	25	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such

sidechains are listed below:

Mol	Chain	Res	Type
3	C	72	ASN
4	D	67	HIS
3	C	124	GLN
3	C	59	ASN
4	D	25	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	24/24 (100%)	-0.54	0	100	100	42, 49, 64, 66	0
2	B	24/24 (100%)	-0.55	0	100	100	39, 54, 75, 88	0
3	C	138/159 (86%)	-0.15	2 (1%)	78	74	36, 55, 85, 90	0
4	D	80/80 (100%)	0.10	6 (7%)	17	12	41, 67, 98, 99	0
All	All	266/287 (92%)	-0.15	8 (3%)	54	47	36, 57, 88, 99	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	78	THR	3.9
4	D	23	GLN	3.7
4	D	79	LYS	3.2
4	D	22	ALA	2.9
4	D	1	ASP	2.6

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.