Full wwPDB X-ray Structure Validation Report (i)

Jan 31, 2016 – 07:03 PM GMT

PDB ID	:	1DU0
Title	:	ENGRAILED HOMEODOMAIN Q50A VARIANT DNA COMPLEX
Authors	:	Grant, R.A.; Rould, M.A.; Klemm, J.D.; Pabo, C.O.
Deposited on	:	2000-01-13
Resolution	:	2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i)) were used in the production of this report:

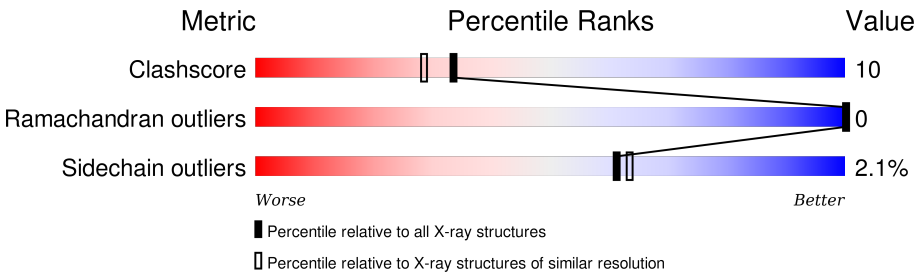
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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 2.00 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	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	21	67% 29% 5%
2	D	21	52% 29% 19%
3	A	57	75% 18% 7%
3	B	57	75% 23% .

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1888 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 DNA (5'-D(*TP*TP*TP*TP*GP*CP*CP*AP*TP*GP*TP*AP*AP*TP*TP*AP*CP*CP*TP*AP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	21	423	206	70	127	20	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	21	432	208	83	121	20	0	0	0

- Molecule 3 is a protein called ENGRAILED HOMEODOMAIN.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	A	53	450	281	89	80	0	0	0
3	B	56	480	299	98	83	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	50	ALA	GLN	engineered	UNP P02836
B	150	ALA	GLN	engineered	UNP P02836

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	21	21	21	0	0
4	B	30	30	30	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	C	29	29	29	0	0
			Total	O		
4	D	23	23	23	0	0

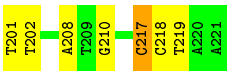
3 Residue-property plots (i)

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.

Note EDS was not executed.

• Molecule 1: DNA (5'-D(*TP*TP*TP*TP*GP*CP*CP*AP*TP*GP*TP*AP*AP*TP*TP*AP*CP*CP*TP*AP*A)-3')

Chain C:

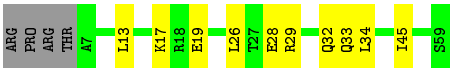


• Molecule 2: DNA (5'-D(*AP*TP*TP*AP*GP*GP*TP*AP*AP*TP*TP*AP*CP*AP*TP*GP*GP*CP*AP*AP*A)-3')

Chain D:

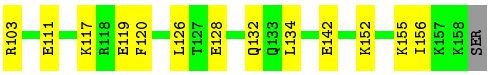
• Molecule 3: ENGRAILED HOMEODOMAIN

Chain A:



• Molecule 3: ENGRAILED HOMEODOMAIN

Chain B:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	126.90Å 45.55Å 71.62Å 90.00° 119.90° 90.00°	Depositor
Resolution (Å)	20.00 – 2.00	Depositor
% Data completeness (in resolution range)	89.0 (20.00-2.00)	Depositor
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	X-PLOR 3.851	Depositor
R, R_{free}	0.234 , 0.270	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1888	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

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	C	0.48	0/472	0.89	0/726
2	D	0.48	0/486	0.83	0/749
3	A	0.39	0/455	0.57	0/603
3	B	0.42	0/486	0.61	0/645
All	All	0.44	0/1899	0.75	0/2723

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	C	0	1
2	D	0	5
All	All	0	6

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (6) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	217	DC	Sidechain
2	D	301	DA	Sidechain
2	D	305	DG	Sidechain
2	D	306	DG	Sidechain
2	D	311	DT	Sidechain
2	D	314	DA	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	C	423	0	242	8	0
2	D	432	0	239	14	0
3	A	450	0	465	5	0
3	B	480	0	500	10	0
4	A	21	0	0	0	0
4	B	30	0	0	3	0
4	C	29	0	0	0	0
4	D	23	0	0	0	0
All	All	1888	0	1446	33	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 10.

All (33) 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:201:DT:H2'	1:C:202:DT:H72	1.59	0.85
2:D:316:DG:H2"	2:D:317:DG:H5"	1.58	0.84
2:D:316:DG:H2"	2:D:317:DG:C5'	2.11	0.81
2:D:316:DG:C2'	2:D:317:DG:H5"	2.13	0.77
3:A:19:GLU:HG3	3:A:34:LEU:HD21	1.65	0.77
1:C:218:DC:H2'	1:C:219:DT:H71	1.67	0.76
1:C:201:DT:H2'	1:C:202:DT:C7	2.23	0.69
3:B:119:GLU:HG2	3:B:134:LEU:HD11	1.79	0.65
1:C:217:DC:OP1	3:B:128:GLU:HG2	1.96	0.65
3:A:28:GLU:O	3:A:32:GLN:HG3	1.97	0.64
1:C:218:DC:H2'	1:C:219:DT:C7	2.28	0.64
2:D:306:DG:H2'	2:D:307:DT:H72	1.81	0.61
3:B:152:LYS:HE2	4:B:473:HOH:O	2.02	0.59
2:D:317:DG:H8	2:D:317:DG:H5'	1.69	0.58
2:D:306:DG:H2'	2:D:307:DT:C7	2.34	0.58
3:B:132:GLN:HG2	3:B:142:GLU:OE2	2.06	0.56
2:D:301:DA:H2'	2:D:302:DT:H72	1.87	0.56
3:A:26:LEU:HD11	3:A:45:ILE:HG22	1.87	0.55
2:D:317:DG:C8	2:D:317:DG:H5'	2.42	0.53

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:302:DT:OP1	3:B:103:ARG:HB3	2.08	0.52
2:D:305:DG:H2"	2:D:306:DG:OP2	2.12	0.50
3:B:155:LYS:HD3	4:B:470:HOH:O	2.12	0.50
3:A:29:ARG:O	3:A:33:GLN:HG3	2.12	0.49
3:A:13:LEU:O	3:A:17:LYS:HG3	2.13	0.49
2:D:302:DT:H2'	2:D:303:DT:H71	1.95	0.48
1:C:210:DG:C2	2:D:314:DA:C2	3.00	0.48
3:B:128:GLU:O	3:B:132:GLN:HG3	2.13	0.47
2:D:316:DG:H2"	2:D:317:DG:H5'	1.95	0.45
1:C:208:DA:C2	2:D:316:DG:C2	3.05	0.45
1:C:201:DT:H2"	1:C:202:DT:C6	2.53	0.44
3:B:152:LYS:O	3:B:156:ILE:HG12	2.19	0.42
3:B:120:PHE:CE2	3:B:152:LYS:HB3	2.56	0.41
3:B:117:LYS:HE2	4:B:473:HOH:O	2.20	0.41

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	A	51/57 (90%)	51 (100%)	0	0	100	100
3	B	54/57 (95%)	54 (100%)	0	0	100	100
All	All	105/114 (92%)	105 (100%)	0	0	100	100

There are no Ramachandran outliers to report.

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	A	47/51 (92%)	47 (100%)	0	100	100
3	B	50/51 (98%)	48 (96%)	2 (4%)	38	33
All	All	97/102 (95%)	95 (98%)	2 (2%)	61	63

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	111	GLU
3	B	126	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.