Full wwPDB X-ray Structure Validation Report (i)

Aug 16, 2023 – 06:04 PM EDT

PDB ID	:	2ETW
Title	:	Principles of protein-DNA recognition revealed in the structural analysis of Ndt80-MSE DNA complexes
Authors	:	Lamoureux, J.S.; Glover, J.N.
Deposited on	:	2005-10-27
Resolution	:	1.67 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
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 (1)) were used in the production of this report:

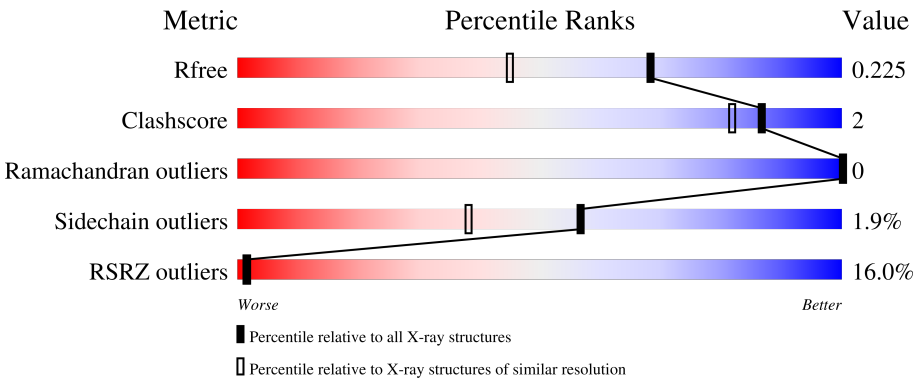
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 1.67 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	6780 (1.70-1.66)
Clashscore	141614	7310 (1.70-1.66)
Ramachandran outliers	138981	7173 (1.70-1.66)
Sidechain outliers	138945	7172 (1.70-1.66)
RSRZ outliers	127900	6661 (1.70-1.66)

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% 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	B	14	29%; 64%; 29%; 7%
2	C	14	36%; 57%; 43%
3	A	345	12%; 79%; 6% • 14%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3376 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*GP*CP*CP*AP*CP*AP*CP*AP*AP*AP*AP*AP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	13	264	125	55	71	13	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	14	287	139	47	88	13	0	0	0

- Molecule 3 is a protein called NDT80 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	297	2494	1589	422	472	11	0	27	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	cloning artifact	UNP P38830
A	-3	PRO	-	cloning artifact	UNP P38830
A	-2	LEU	-	cloning artifact	UNP P38830
A	-1	GLY	-	cloning artifact	UNP P38830
A	0	SER	-	cloning artifact	UNP P38830
A	146	GLY	SER	engineered mutation	UNP P38830
A	200	THR	ILE	engineered mutation	UNP P38830

- Molecule 4 is water.

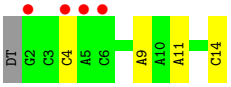
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	B	24	24	24	0	0
			Total	O		
4	C	25	25	25	0	0
			Total	O		
4	A	282	282	282	0	0

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 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.

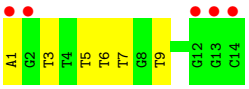
• Molecule 1: 5'-D(*TP*GP*CP*CP*AP*CP*AP*CP*AP*AP*AP*AP*AP*C)-3'

Chain B:



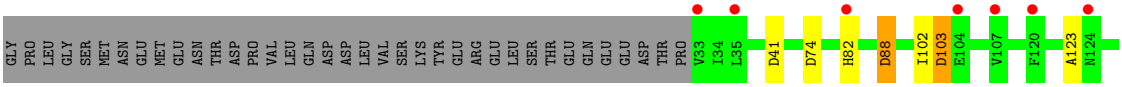
• Molecule 2: 5'-D(*AP*GP*TP*TP*TP*TP*TP*GP*TP*GP*TP*GP*GP*C)-3'

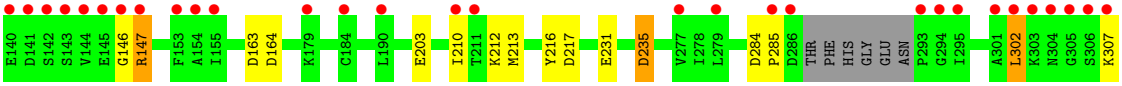
Chain C:

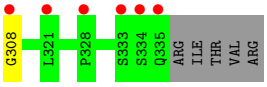


• Molecule 3: NDT80 protein

Chain A:







4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	70.30Å 78.92Å 161.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	23.57 – 1.67 23.57 – 1.67	Depositor EDS
% Data completeness (in resolution range)	98.7 (23.57-1.67) 98.7 (23.57-1.67)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	0.04	Depositor
$< I/\sigma(I) >$ [1]	3.71 (at 1.67Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.180 , 0.199 0.214 , 0.225	Depositor DCC
R_{free} test set	2636 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	22.2	Xtriage
Anisotropy	0.299	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 53.1	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	3376	wwPDB-VP
Average B, all atoms (Å^2)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.71% 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 (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	B	0.68	0/297	1.38	4/454 (0.9%)
2	C	0.98	1/320 (0.3%)	1.57	8/494 (1.6%)
3	A	0.47	0/2664	0.81	8/3609 (0.2%)
All	All	0.56	1/3281 (0.0%)	0.99	20/4557 (0.4%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	3	DT	C3'-O3'	6.22	1.52	1.44

All (20) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	1	DA	O4'-C1'-N9	8.76	114.13	108.00
3	A	163	ASP	CB-CG-OD2	6.81	124.43	118.30
2	C	1	DA	O4'-C1'-C2'	-6.36	100.81	105.90
3	A	164	ASP	CB-CG-OD2	6.19	123.87	118.30
1	B	11	DA	P-O3'-C3'	6.13	127.06	119.70
2	C	5	DT	O4'-C1'-N1	-6.02	103.79	108.00
2	C	9	DT	O4'-C1'-N1	6.01	112.21	108.00
1	B	14	DC	O4'-C1'-N1	5.95	112.16	108.00
3	A	217	ASP	CB-CG-OD2	5.61	123.34	118.30
3	A	41	ASP	CB-CG-OD2	5.46	123.21	118.30
1	B	14	DC	C1'-O4'-C4'	-5.43	104.67	110.10
3	A	103[A]	ASP	CB-CG-OD2	5.38	123.14	118.30
3	A	103[B]	ASP	CB-CG-OD2	5.38	123.14	118.30
2	C	6	DT	O4'-C1'-N1	-5.37	104.24	108.00
3	A	74	ASP	CB-CG-OD2	5.31	123.08	118.30
1	B	9	DA	O4'-C1'-N9	5.22	111.66	108.00
3	A	284	ASP	CB-CG-OD2	5.13	122.92	118.30
2	C	1	DA	C1'-O4'-C4'	-5.08	105.02	110.10
2	C	3	DT	O4'-C4'-C3'	5.00	109.00	106.00
2	C	7	DT	C6-C5-C7	-5.00	119.90	122.90

There are no chirality outliers.

There are no planarity outliers.

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	B	264	0	144	1	0
2	C	287	0	163	0	0
3	A	2494	0	2439	10	0
4	A	282	0	0	3	1
4	B	24	0	0	1	0
4	C	25	0	0	0	0
All	All	3376	0	2746	11	1

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 2.

All (11) 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)
3:A:235[B]:ASP:OD1	4:A:618:HOH:O	1.76	1.02
3:A:102[A]:ILE:HG22	3:A:103[A]:ASP:OD2	1.66	0.95
3:A:231:GLU:OE2	4:A:587:HOH:O	1.97	0.81
3:A:203:GLU:HB3	3:A:216:TYR:OH	2.01	0.60
3:A:285:PRO:HG3	3:A:307:LYS:O	2.07	0.55
3:A:302:LEU:HD13	3:A:308:GLY:HA3	1.92	0.51
3:A:88[A]:ASP:OD1	3:A:123:ALA:HB3	2.12	0.49
3:A:210:ILE:HD13	3:A:213[A]:MET:CE	2.45	0.46
3:A:212:LYS:NZ	4:A:455:HOH:O	2.49	0.45
1:B:4:DC:P	4:B:32:HOH:O	2.77	0.42
3:A:146:GLY:H	3:A:147:ARG:HG2	1.85	0.42

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:A:604:HOH:O	4:A:604:HOH:O[4_566]	1.90	0.30

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	321/345 (93%)	313 (98%)	8 (2%)	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	294/318 (92%)	286 (97%)	8 (3%)	44 24

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	82[A]	HIS
3	A	82[B]	HIS
3	A	88[A]	ASP
3	A	88[B]	ASP
3	A	147	ARG
3	A	235[A]	ASP
3	A	235[B]	ASP
3	A	302	LEU

Sometimes 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 monosaccharides 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	B	13/14 (92%)	1.00	4 (30%)	0	0	19, 25, 42, 42	0
2	C	14/14 (100%)	1.67	5 (35%)	0	0	16, 23, 48, 51	0
3	A	297/345 (86%)	0.84	43 (14%)	2	2	10, 16, 34, 44	0
All	All	324/373 (86%)	0.88	52 (16%)	1	1	10, 16, 35, 51	0

All (52) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	302	LEU	7.6
3	A	304	ASN	7.3
3	A	142	SER	7.3
3	A	293	PRO	6.9
3	A	144	VAL	6.5
3	A	305	GLY	5.9
3	A	303	LYS	5.7
2	C	14	DC	5.6
3	A	141	ASP	5.5
3	A	146	GLY	5.2
3	A	328	PRO	4.8
2	C	13	DG	4.7
3	A	286	ASP	4.7
3	A	306	SER	4.0
3	A	307	LYS	3.9
3	A	82[A]	HIS	3.8
3	A	145	GLU	3.8
3	A	179	LYS	3.8
2	C	1	DA	3.8
3	A	210	ILE	3.7
3	A	294	GLY	3.7
3	A	140	GLU	3.7
2	C	12	DG	3.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
3	A	295	ILE	3.6
1	B	5	DA	3.5
3	A	190[A]	LEU	3.5
3	A	301	ALA	3.4
3	A	285	PRO	3.4
3	A	143	SER	3.1
3	A	155	ILE	2.9
3	A	211	THR	2.8
3	A	33	VAL	2.7
3	A	333	SER	2.6
3	A	124	ASN	2.5
3	A	154	ALA	2.5
3	A	35	LEU	2.5
3	A	308	GLY	2.5
3	A	334	SER	2.5
3	A	321	LEU	2.4
1	B	4	DC	2.4
1	B	2	DG	2.3
3	A	184[A]	CYS	2.3
2	C	2	DG	2.2
3	A	279	LEU	2.2
3	A	104	GLU	2.2
3	A	335	GLN	2.2
3	A	120	PHE	2.2
3	A	153	PHE	2.2
1	B	6	DC	2.1
3	A	277	VAL	2.1
3	A	107	VAL	2.1
3	A	147	ARG	2.0

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 monosaccharides 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.