Full wwPDB X-ray Structure Validation Report (i)

Jun 25, 2024 – 10:23 AM EDT

PDB ID	:	6OE7
Title	:	Crystal structure of HMCES cross-linked to DNA abasic site
Authors	:	Halabelian, L.; Li, Y.; Zeng, H.; Bountra, C.; Edwards, A.M.; Arrowsmith, C.H.; Structural Genomics Consortium (SGC)
Deposited on	:	2019-03-27
Resolution	:	2.20 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 (i)) were used in the production of this report:

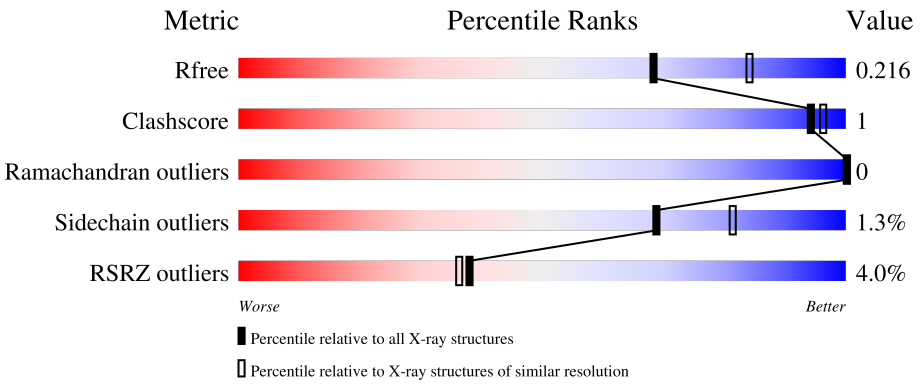
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	A	276	4%, 90%, •, 7%
2	C	6	67%, 33%
3	B	12	67%, 8%, 25%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2402 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 Embryonic stem cell-specific 5-hydroxymethylcytosine-bindin g protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	258	2052	1292	365	381	14	8	0	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	271	ALA	-	expression tag	UNP Q96FZ2
A	272	GLU	-	expression tag	UNP Q96FZ2
A	273	ASN	-	expression tag	UNP Q96FZ2
A	274	LEU	-	expression tag	UNP Q96FZ2
A	275	TYR	-	expression tag	UNP Q96FZ2
A	276	PHE	-	expression tag	UNP Q96FZ2
A	277	GLN	-	expression tag	UNP Q96FZ2

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	6	122	59	22	36	5	0	0	0

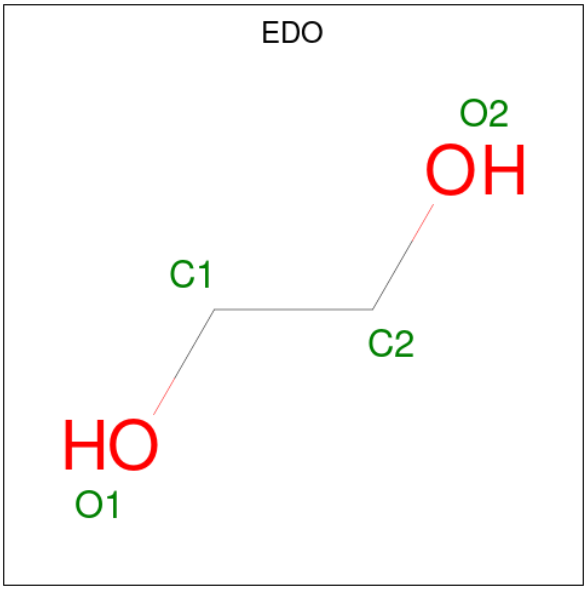
- Molecule 3 is a DNA chain called DNA (5'-D(*CP*CP*AP*GP*AP*CP*GP*TP*(DRZ)P* GP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	B	9	171	82	31	50	8	0	0	0

- Molecule 4 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	X		
4	A	3	3	3	0	0
			Total	X		
4	B	1	1	1	0	0

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
5	A	1	4	2	2	0	0
			Total	C	O		
5	A	1	4	2	2	0	0
			Total	C	O		
5	A	1	4	2	2	0	0
			Total	C	O		
5	A	1	4	2	2	0	0
			Total	C	O		
5	A	1	4	2	2	0	0

- Molecule 6 is water.

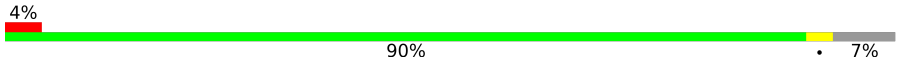
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	31	31	31	0	0
			Total	O		
6	B	2	2	2	0	0

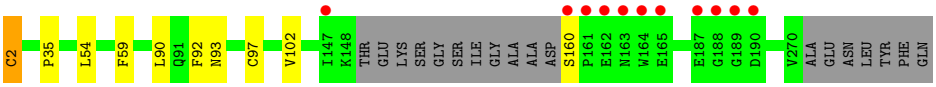
3 Residue-property plots (i)

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: Embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein

Chain A:



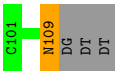


• Molecule 2: DNA (5'-D(*GP*TP*CP*TP*GP*G)-3')

Chain C:

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

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	55.51Å 51.53Å 149.72Å 90.00° 92.72° 90.00°	Depositor
Resolution (Å)	48.72 – 2.20 48.72 – 2.20	Depositor EDS
% Data completeness (in resolution range)	96.4 (48.72-2.20) 96.5 (48.72-2.20)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.86 (at 2.20Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.194 , 0.216 0.201 , 0.216	Depositor DCC
R_{free} test set	1036 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	53.0	Xtriage
Anisotropy	0.578	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 45.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.029 for h,-k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	2402	wwPDB-VP
Average B, all atoms (Å^2)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.19% 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)

Bond lengths and bond angles in the following residue types are not validated in this section: EDO, UNX, DRZ

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	0.67	0/2109	0.82	1/2871 (0.0%)
2	C	0.50	0/136	0.81	0/209
3	B	0.52	0/179	0.92	0/274
All	All	0.65	0/2424	0.83	1/3354 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	2	CYS	CA-CB-SG	-7.23	100.99	114.00

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	A	2052	0	1944	4	0
2	C	122	0	70	1	0
3	B	171	0	98	1	0
4	A	3	0	0	0	0
4	B	1	0	0	0	0
5	A	20	0	30	1	0
6	A	31	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	2	0	0	0	0
All	All	2402	0	2142	6	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 1.

All (6) 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:A:35:PRO:HD2	5:A:304:EDO:H21	1.91	0.50
2:C:1:DG:H1'	2:C:2:DT:H5'	1.94	0.48
1:A:97:CYS:SG	1:A:102:VAL:HG23	2.54	0.47
1:A:90:LEU:HD11	1:A:92:PHE:CZ	2.50	0.46
3:B:109:DRZ:H5"	3:B:109:DRZ:H2'	1.63	0.44
1:A:54:LEU:HD11	1:A:59:PHE:CE2	2.54	0.43

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	254/276 (92%)	246 (97%)	8 (3%)	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
1	A	224/247 (91%)	221 (99%)	3 (1%)	69 81

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	CYS
1	A	93	ASN
1	A	160	SER

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)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	DRZ	B	109	1,3	7,10,12	0.84	0	6,12,16	2.50	3 (50%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	DRZ	B	109	1,3	-	0/9/11/13	-

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	109	DRZ	O3'-C3'-C2'	4.05	116.72	109.35
3	B	109	DRZ	C1'-C2'-C3'	3.41	119.99	113.61
3	B	109	DRZ	C2'-C3'-C4'	-2.73	106.88	113.35

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	109	DRZ	1	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 4 are unknown - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
5	EDO	A	307	-	3,3,3	0.23	0	2,2,2	0.33	0
5	EDO	A	304	-	3,3,3	0.20	0	2,2,2	0.17	0
5	EDO	A	303	-	3,3,3	0.15	0	2,2,2	0.18	0
5	EDO	A	302	-	3,3,3	0.12	0	2,2,2	0.07	0
5	EDO	A	306	-	3,3,3	0.13	0	2,2,2	0.15	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the

Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	EDO	A	307	-	-	0/1/1/1	-
5	EDO	A	304	-	-	1/1/1/1	-
5	EDO	A	303	-	-	1/1/1/1	-
5	EDO	A	302	-	-	1/1/1/1	-
5	EDO	A	306	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	304	EDO	O1-C1-C2-O2
5	A	306	EDO	O1-C1-C2-O2
5	A	302	EDO	O1-C1-C2-O2
5	A	303	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	304	EDO	1	0

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	258/276 (93%)	0.01	11 (4%) 35 33	41, 58, 94, 137	3 (1%)
2	C	6/6 (100%)	-0.53	0 100 100	58, 63, 71, 78	0
3	B	8/12 (66%)	-0.61	0 100 100	52, 57, 65, 74	0
All	All	272/294 (92%)	-0.02	11 (4%) 38 36	41, 59, 93, 137	3 (1%)

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	147	ILE	3.9
1	A	164	TRP	3.7
1	A	162	GLU	3.6
1	A	188	GLY	3.3
1	A	189	GLY	3.2
1	A	163	ASN	3.2
1	A	161	PRO	3.1
1	A	187	GLU	3.1
1	A	165	GLU	2.5
1	A	160	SER	2.4
1	A	190	ASP	2.1

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

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, 95th 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
3	DRZ	B	109	11/13	0.97	0.10	59,68,82,85	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

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, 95th 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	EDO	A	307	4/4	0.59	0.18	88,91,94,94	0
4	UNX	A	305	1/1	0.83	0.19	45,45,45,45	0
4	UNX	A	308	1/1	0.85	0.31	50,50,50,50	0
4	UNX	A	301	1/1	0.86	0.70	72,72,72,72	0
5	EDO	A	302	4/4	0.90	0.21	59,61,63,65	0
5	EDO	A	303	4/4	0.91	0.14	63,64,65,67	0
5	EDO	A	306	4/4	0.94	0.07	75,83,84,86	0
5	EDO	A	304	4/4	0.94	0.17	63,66,71,75	0
4	UNX	B	201	1/1	0.98	0.33	52,52,52,52	0

6.5 Other polymers (i)

There are no such residues in this entry.