wwPDB X-ray Structure Validation Summary Report (i)

Jan 27, 2024 – 07:47 PM EST

PDB ID	:	1DEW
Title	:	CRYSTAL STRUCTURE OF HUMAN APE1 BOUND TO ABASIC DNA
Authors	:	Mol, C.D.; Izumi, T.; Mitra, S.; Tainer, J.A.
Deposited on	:	1999-11-15
Resolution	:	2.65 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

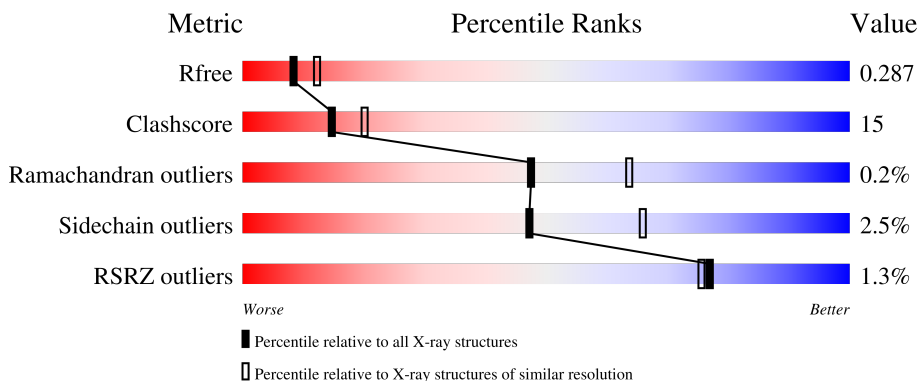|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance ⓘ

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

The reported resolution of this entry is 2.65 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.
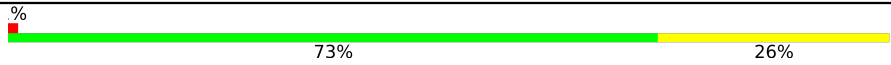
Metric	Percentile Ranks	Value
Rfree		0.287
Clashscore		15
Ramachandran outliers		0.2%
Sidechain outliers		2.5%
RSRZ outliers		1.3%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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	U	15	47% 53%
1	X	15	47% 53%
2	V	14	14% 86%
2	Y	14	29% 71%
3	A	279	2% 76% 22% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	B	279	% 73% 26%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 6034 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(*GP*CP*GP*TP*CP*CP*(3DR)P*CP*GP*AP*CP*GP*AP*CP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	X	15	Total	C	N	O	P	0	0	0
			294	139	55	86	14			
1	U	15	Total	C	N	O	P	0	0	0
			294	139	55	86	14			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	Y	14	Total	C	N	O	P	0	0	0
			288	136	56	83	13			
2	V	14	Total	C	N	O	P	0	0	0
			288	136	56	83	13			

- Molecule 3 is a protein called MAJOR APURINIC/APYRIMIDINIC ENDONUCLEASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	278	Total	C	N	O	S	0	0	0
			2206	1409	382	406	9			
3	B	279	Total	C	N	O	S	0	0	0
			2215	1414	383	409	9			

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total 5	O 4	S 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	X	18	Total 18	O 18	0	0
5	Y	27	Total 27	O 27	0	0
5	U	18	Total 18	O 18	0	0
5	V	16	Total 16	O 16	0	0
5	A	151	Total 151	O 151	0	0
5	B	214	Total 214	O 214	0	0

3 Residue-property plots ⓘ

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.
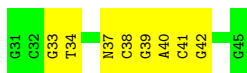
- Molecule 1: 5'-D(*GP*CP*GP*TP*CP*CP*(3DR)P*CP*GP*AP*CP*GP*AP*CP*G)-3'

Chain X:

- Molecule 1: 5'-D(*GP*CP*GP*TP*CP*CP*(3DR)P*CP*GP*AP*CP*GP*AP*CP*G)-3'

Chain U:
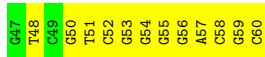

- Molecule 2: 5'-D(*GP*TP*CP*GP*TP*CP*GP*GP*GP*GP*AP*CP*GP*C)-3'

Chain Y:

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

Chain V:

- Molecule 3: MAJOR APURINIC/APYRIMIDINIC ENDONUCLEASE

Chain A:

L182 Q186 R193 L205 V206 L207 L211 E216 E217 I218 D219 L220 R221 K224 A230 L249 A250 D251 S252 F253 L256 Y257 P258 N259 Y264 T268 W280 R281 Y284 F285 L286 L287 L292 P293 A294 L295 K299 I300 R301 A304 P311 L318

• Molecule 3: MAJOR APURINIC/APYRIMIDINIC ENDONUCLEASE

Chain B:

%

73% 26%

E40 G41 P42 A43 D47 P48 P49 D50 T53 S54 K58 W67 L72 R73 A74 W75 I76 W83 E87 L92 C93 E101 L104 P105 A106 E107 E110 Q117 P122 K125 E126 L133 L134 E154 V159 S164 V168 T169 V180 R181

L182 R185 Q186 F192 L196 K197 G198 R202 K203 P204 N212 V213 E217 R221 N222 P223 K224 G225 N226 F240 L243 R254 H255 L256 N259 T260 P261 T265 T268 N272 S275 W280 S288 L291 L292 P293 K299 I300 R301 A304

S307 D308 H309 C310 P311 A317 L318

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	71.22Å 72.25Å 93.75Å 90.00° 94.29° 90.00°	Depositor
Resolution (Å)	20.00 – 2.65 19.93 – 2.63	Depositor EDS
% Data completeness (in resolution range)	90.9 (20.00-2.65) 90.7 (19.93-2.63)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.72 (at 2.63Å)	Xtriage
Refinement program	X-PLOR 3.851	Depositor
R, R_{free}	0.195 , 0.286 0.200 , 0.287	Depositor DCC
R_{free} test set	2591 reflections (10.05%)	wwPDB-VP
Wilson B-factor (Å^2)	41.8	Xtriage
Anisotropy	0.348	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 58.1	EDS
L-test for twinning[2]	$< |L| > = 0.52, < L^2 > = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6034	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 56.12 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.8661e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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: 3DR, SO4

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	U	0.43	0/316	0.69	0/483
1	X	0.44	0/316	0.72	0/483
2	V	0.40	0/323	0.69	0/498
2	Y	0.37	0/323	0.70	0/498
3	A	0.39	0/2266	0.69	1/3074 (0.0%)
3	B	0.40	0/2275	0.66	0/3086
All	All	0.40	0/5819	0.68	1/8122 (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	A	0	1

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	A	43	ALA	N-CA-C	5.18	124.98	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	171	TYR	Sidechain

WORLDWIDE
wwPDB
PROTEIN DATA BANK

5.2 Too-close contacts ⓘ

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	U	294	0	165	11	0
1	X	294	0	165	16	0
2	V	288	0	158	21	0
2	Y	288	0	158	15	0
3	A	2206	0	2172	51	0
3	B	2215	0	2178	58	0
4	B	5	0	0	0	0
5	A	151	0	0	4	0
5	B	214	0	0	10	0
5	U	18	0	0	0	0
5	V	16	0	0	1	0
5	X	18	0	0	1	0
5	Y	27	0	0	3	0
All	All	6034	0	4996	162	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 15.

The worst 5 of 162 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:Y:20:DG:H2”	2:Y:21:DT:H5’	1.21	1.18
2:V:50:DG:H2”	2:V:51:DT:H5’	1.11	1.09
2:V:50:DG:H2”	2:V:51:DT:C5’	1.92	0.99
1:U:40:DA:H2”	1:U:41:DC:H5”	1.46	0.95
3:B:40:GLU:HG2	3:B:41:GLY:H	1.36	0.91

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	276/279 (99%)	263 (95%)	13 (5%)	0	100	100
3	B	277/279 (99%)	261 (94%)	15 (5%)	1 (0%)	34	48
All	All	553/558 (99%)	524 (95%)	28 (5%)	1 (0%)	47	64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	B	125	LYS

5.3.2 Protein sidechains ⓘ

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	236/238 (99%)	230 (98%)	6 (2%)	47	66
3	B	237/238 (100%)	231 (98%)	6 (2%)	47	66
All	All	473/476 (99%)	461 (98%)	12 (2%)	47	66

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

Mol	Chain	Res	Type
3	B	126	GLU
3	B	164	SER
3	B	308	ASP
3	B	212	ASN
3	A	221	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	289	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	B	153	GLN
3	A	229	ASN
3	A	226	ASN
3	A	289	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues are 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	$\#|Z| > 2$
1	3DR	U	37	1	8,11,12	0.22	0	9,14,17	0.70	0
1	3DR	X	7	1	8,11,12	0.26	0	9,14,17	0.86	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
1	3DR	U	37	1	-	0/3/15/16	0/1/1/1
1	3DR	X	7	1	-	0/3/15/16	0/1/1/1

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	U	37	3DR	1	0

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

1 ligand 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	$\#|Z| > 2$
4	SO4	B	2055	-	4,4,4	0.45	0	6,6,6	0.22	0

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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

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, 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	U	14/15 (93%)	-0.22	0	100	100	30, 39, 64, 68	0
1	X	14/15 (93%)	-0.09	0	100	100	27, 36, 66, 72	0
2	V	14/14 (100%)	-0.19	0	100	100	22, 47, 67, 70	0
2	Y	14/14 (100%)	-0.20	0	100	100	18, 48, 65, 67	0
3	A	278/279 (99%)	-0.44	5 (1%)	68	65	5, 29, 47, 83	0
3	B	279/279 (100%)	-0.67	3 (1%)	80	79	9, 21, 43, 78	0
All	All	613/616 (99%)	-0.52	8 (1%)	77	75	5, 25, 56, 83	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	41	GLY	4.3
3	B	42	PRO	3.9
3	B	41	GLY	3.3
3	A	42	PRO	3.2
3	A	44	LEU	2.3

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

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
1	3DR	X	7	11/12	0.95	0.17	21,28,34,34	0
1	3DR	U	37	11/12	0.96	0.13	21,25,29,32	0

6.3 Carbohydrates ⓘ

There are no monosaccharides 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, 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
4	SO4	B	2055	5/5	0.94	0.24	64,66,68,68	0

6.5 Other polymers ⓘ

There are no such residues in this entry.