wwPDB X-ray Structure Validation Summary Report (i)

Feb 28, 2024 – 06:43 PM EST

PDB ID	:	5THG
Title	:	Engineered variant of I-OnuI meganuclease targeting the HIV CCR5 gene; harbors 43 point mutations relative to wild-type I-OnuI
Authors	:	Hallinan, J.P.; Stoddard, B.L.
Deposited on	:	2016-09-29
Resolution	:	3.11 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
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:

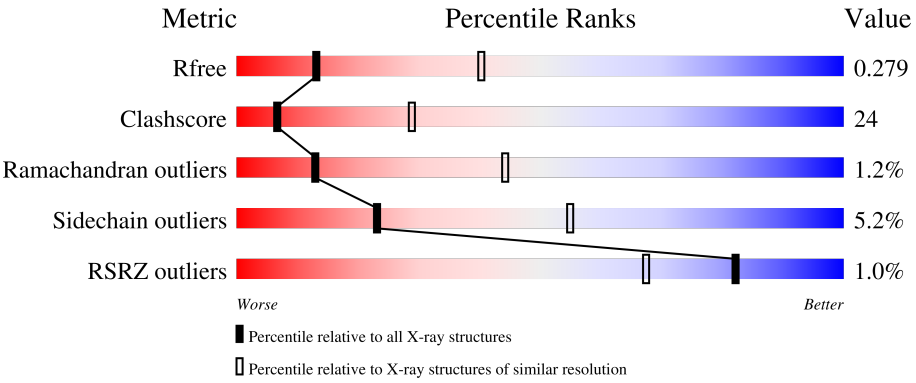
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.11 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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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	304	64% 29%
1	D	304	64% 29%
2	B	29	24% 72%
2	E	29	41% 55%
3	C	29	21% 76%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	29	14% 86%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 6772 atoms, of which 16 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 I-OnuI_e-hCCR5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	294	Total	C	N	O	S	0	0	0
			2261	1455	390	408	8			
1	D	291	Total	C	N	O	S	0	0	0
			2149	1382	369	390	8			

- Molecule 2 is a DNA chain called DNA (29-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	28	Total	C	N	O	P	0	0	0
			565	270	96	171	28			
2	E	28	Total	C	N	O	P	0	0	0
			565	270	96	171	28			

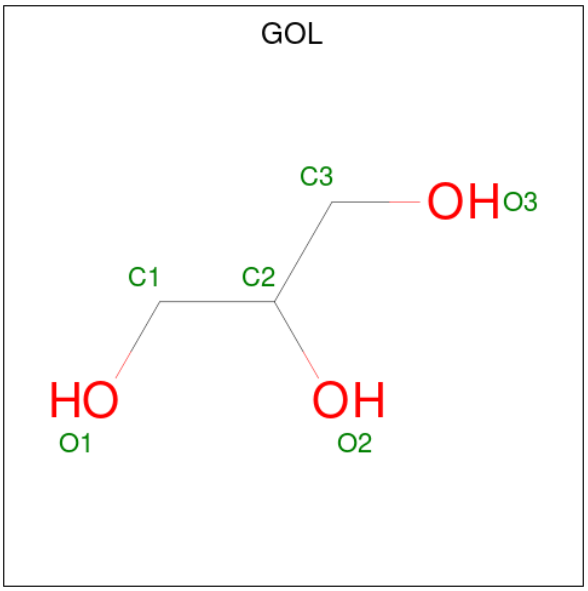
- Molecule 3 is a DNA chain called DNA (29-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	28	Total	C	N	O	P	0	0	0
			583	275	115	165	28			
3	F	29	Total	C	N	O	P	0	0	0
			605	285	120	171	29			

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Ca	0	0
			2	2		
4	D	2	Total	Ca	0	0
			2	2		

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total C H O 14 3 8 3	0	0
5	E	1	Total C H O 14 3 8 3	0	0

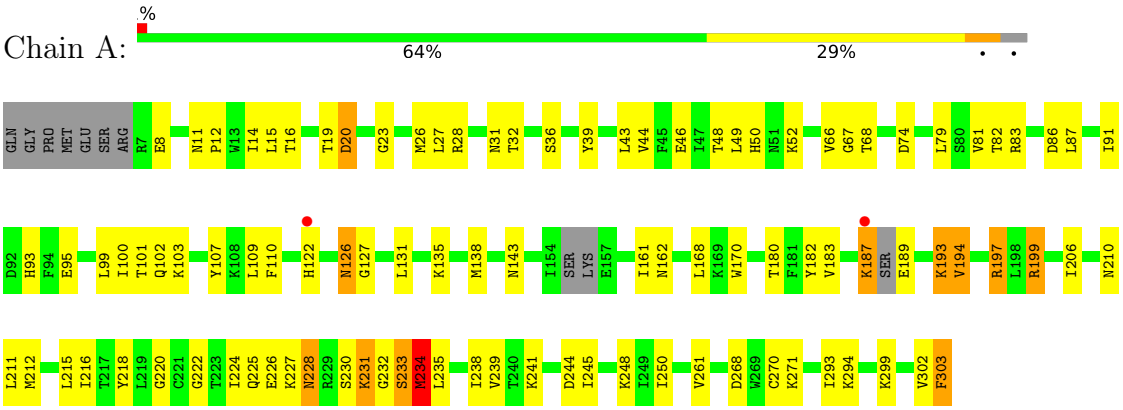
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	2	Total O 2 2	0	0
6	B	2	Total O 2 2	0	0
6	C	1	Total O 1 1	0	0
6	D	1	Total O 1 1	0	0
6	E	4	Total O 4 4	0	0
6	F	2	Total O 2 2	0	0

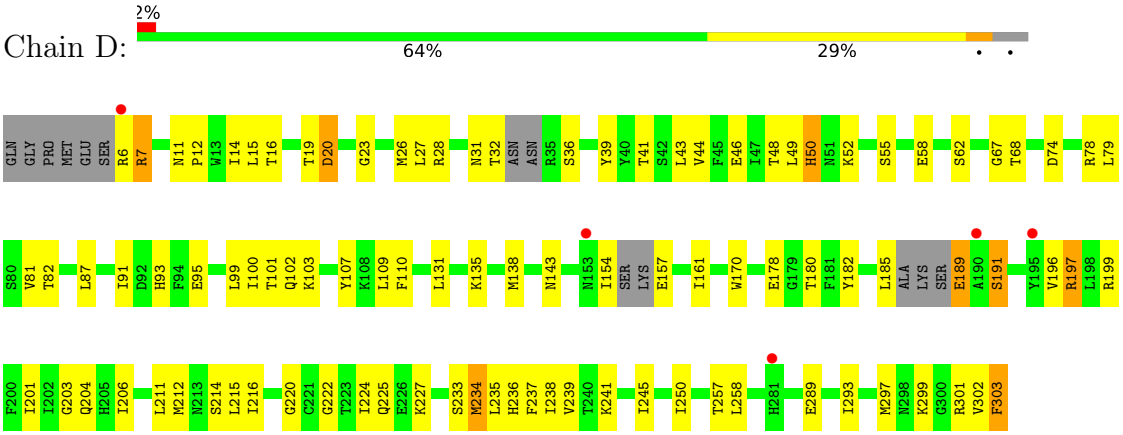
3 Residue-property plots (i)

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.

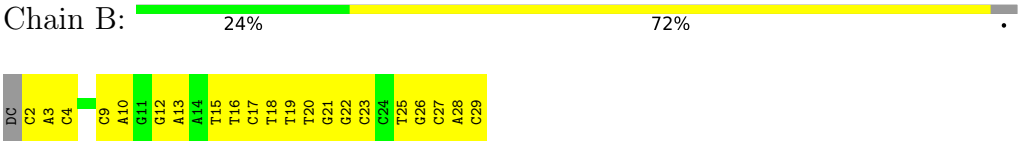
• Molecule 1: I-OnuI_e-hCCR5



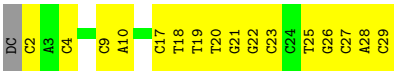
• Molecule 1: I-OnuI_e-hCCR5



• Molecule 2: DNA (29-MER)

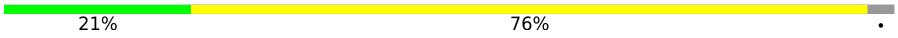


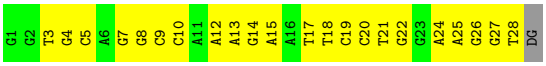
• Molecule 2: DNA (29-MER)



• Molecule 3: DNA (29-MER)

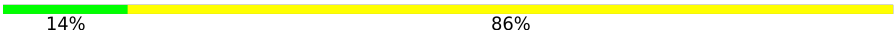
Chain C:

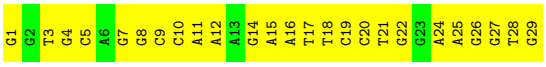




• Molecule 3: DNA (29-MER)

Chain F:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	75.96Å 112.35Å 122.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.77 – 3.11 47.77 – 3.11	Depositor EDS
% Data completeness (in resolution range)	98.7 (47.77-3.11) 90.8 (47.77-3.11)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.11 (at 3.12Å)	Xtriage
Refinement program	PHENIX (1.10_2155)	Depositor
R, R_{free}	0.221 , 0.279 0.221 , 0.279	Depositor DCC
R_{free} test set	1948 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å^2)	55.5	Xtriage
Anisotropy	0.832	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 51.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6772	wwPDB-VP
Average B, all atoms (Å^2)	55.0	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 79.62 % 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 5.3637e-07. 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: CA, GOL

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	0.30	0/2304	0.45	0/3115
1	D	0.30	0/2191	0.45	0/2976
2	B	0.87	0/630	1.02	0/968
2	E	0.78	0/630	0.99	0/968
3	C	0.80	0/656	0.91	0/1012
3	F	0.74	0/681	0.91	0/1051
All	All	0.54	0/7092	0.70	0/10090

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	A	2261	0	2178	96	0
1	D	2149	0	1952	91	1
2	B	565	0	317	37	0
2	E	565	0	317	31	0
3	C	583	0	314	30	0
3	F	605	0	325	37	1
4	A	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	2	0	0	0	0
5	B	6	8	8	0	0
5	E	6	8	8	1	0
6	A	2	0	0	0	0
6	B	2	0	0	1	0
6	C	1	0	0	0	0
6	D	1	0	0	0	0
6	E	4	0	0	2	0
6	F	2	0	0	0	0
All	All	6756	16	5419	286	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 24.

The worst 5 of 286 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:E:21:DG:OP1	6:E:201:HOH:O	1.83	0.96
5:E:101:GOL:HO1	5:E:101:GOL:HO3	1.13	0.93
2:E:22:DG:N1	3:F:9:DC:N3	2.16	0.93
2:E:22:DG:N2	3:F:9:DC:O2	2.04	0.90
1:A:232:GLY:H	1:A:233:SER:HA	1.38	0.88

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)
1:D:257:THR:O	3:F:1:DG:N2[2_485]	2.16	0.04

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	288/304 (95%)	256 (89%)	27 (9%)	5 (2%)	9 36
1	D	283/304 (93%)	259 (92%)	22 (8%)	2 (1%)	22 57
All	All	571/608 (94%)	515 (90%)	49 (9%)	7 (1%)	13 44

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	8	GLU
1	A	231	LYS
1	A	194	VAL
1	D	7	ARG
1	D	302	VAL

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	227/274 (83%)	215 (95%)	12 (5%)	22 54
1	D	199/274 (73%)	189 (95%)	10 (5%)	24 57
All	All	426/548 (78%)	404 (95%)	22 (5%)	23 55

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	50	HIS
1	D	191	SER
1	D	189	GLU
1	D	197	ARG
1	A	197	ARG

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

5.3.3 RNA

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

Of 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 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	GOL	B	101	-	5,5,5	0.35	0	5,5,5	0.31	0
5	GOL	E	101	-	5,5,5	0.40	0	5,5,5	0.32	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	GOL	B	101	-	-	0/4/4/4	-
5	GOL	E	101	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	E	101	GOL	O1-C1-C2-C3
5	E	101	GOL	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	E	101	GOL	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

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, 95^{th} 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	294/304 (96%)	0.03	2 (0%) 87 75	26, 44, 91, 110	0
1	D	291/304 (95%)	0.04	5 (1%) 70 49	32, 54, 98, 127	0
2	B	28/29 (96%)	-0.28	0 100 100	33, 47, 75, 94	0
2	E	28/29 (96%)	-0.22	0 100 100	48, 60, 90, 113	0
3	C	28/29 (96%)	-0.16	0 100 100	37, 49, 73, 80	0
3	F	29/29 (100%)	-0.17	0 100 100	47, 60, 94, 95	0
All	All	698/724 (96%)	-0.00	7 (1%) 82 67	26, 50, 94, 127	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	153	ASN	4.0
1	A	122	HIS	3.4
1	A	187	LYS	3.0
1	D	281	HIS	2.4
1	D	195	TYR	2.2

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

There are no non-standard protein/DNA/RNA residues in this entry.

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, 95^{th} 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	GOL	B	101	6/6	0.88	0.21	46,59,76,79	0
4	CA	D	401	1/1	0.91	0.11	56,56,56,56	0
5	GOL	E	101	6/6	0.92	0.28	58,69,80,80	0
4	CA	A	401	1/1	0.93	0.13	50,50,50,50	0
4	CA	D	402	1/1	0.95	0.27	65,65,65,65	0
4	CA	A	402	1/1	0.96	0.15	36,36,36,36	0

6.5 Other polymers

There are no such residues in this entry.