wwPDB X-ray Structure Validation Summary Report (i)

Oct 12, 2024 – 08:40 am BST

PDB ID	:	5CO8
Title	:	Crystal structure of the Holliday junction-resolving enzyme GEN1 (WT) in complex with product DNA and Mg2+ ion
Authors	:	Liu, Y.J.; Freeman, A.D.J.; Declais, A.C.; Wilson, T.J.; Gartner, A.; Lilley, D.M.J.
Deposited on	:	2015-07-20
Resolution	:	2.40 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:

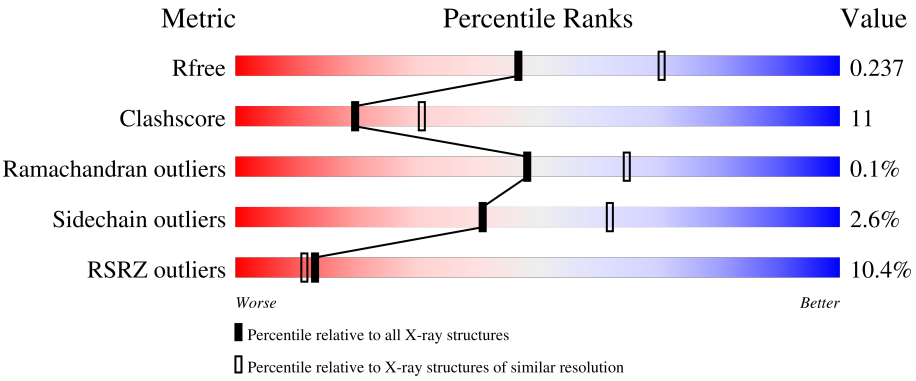
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 2.40 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}	164625	4642 (2.40-2.40)
Clashscore	180529	5218 (2.40-2.40)
Ramachandran outliers	177936	5158 (2.40-2.40)
Sidechain outliers	177891	5159 (2.40-2.40)
RSRZ outliers	164620	4642 (2.40-2.40)

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	R	31	39% 61%
2	C	464	10% 71% 13% · 14%
3	X	16	6% 62% 38%
4	A	464	8% 68% 15% · 15%
5	H	15	7% 27% 73%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 7414 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 (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	R	31	602	287	106	179	30	0	0	1

- Molecule 2 is a protein called Nuclease-like protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
2	C	397	3109	1984	542	570	6	7	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	X	16	296	138	54	89	15	0	0	1

- Molecule 4 is a protein called Nuclease-like protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
4	A	393	3078	1962	538	565	8	5	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	H	15	298	139	59	86	14	0	0	1

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	C	1	Total Mg 1 1	0	0
6	A	1	Total Mg 1 1	0	0

- Molecule 7 is water.

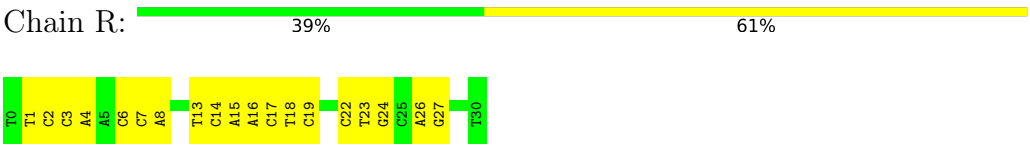
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	C	14	Total O 14 14	0	0
7	X	1	Total O 1 1	0	0
7	A	11	Total O 11 11	0	0
7	H	3	Total O 3 3	0	0

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.

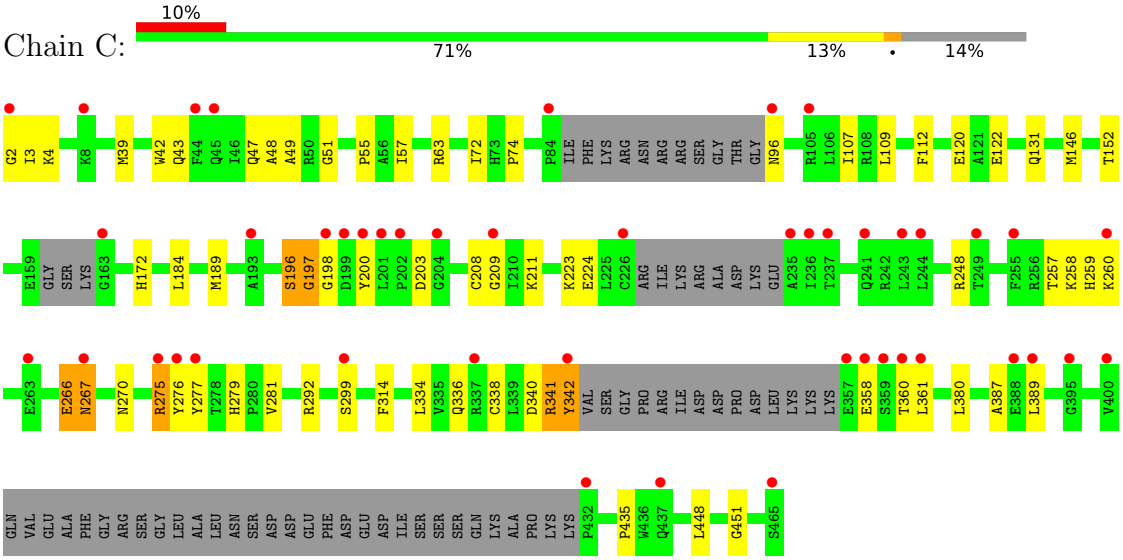
• Molecule 1: DNA (31-MER)

Chain R:



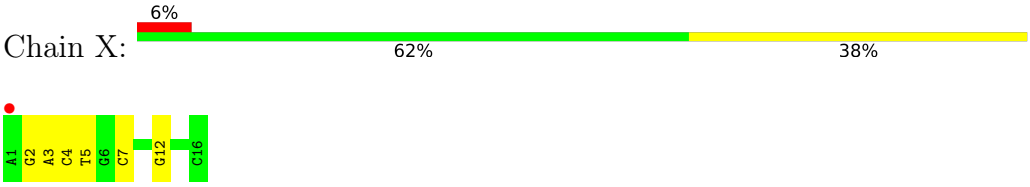
• Molecule 2: Nuclease-like protein

Chain C:



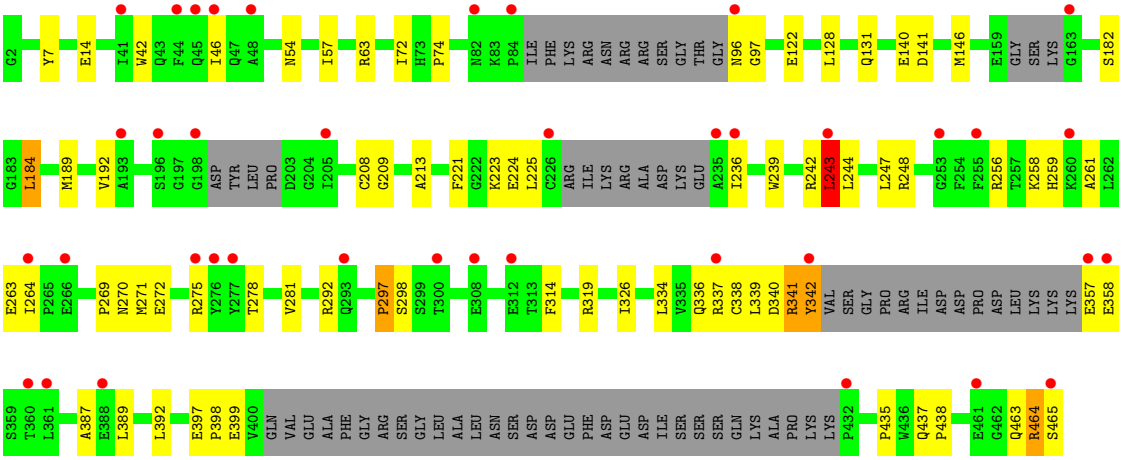
• Molecule 3: DNA (5'-D(*AP*GP*AP*CP*TP*GP*CP*AP*GP*TP*TP*GP*AP*GP*TP*C)-3')

Chain X:

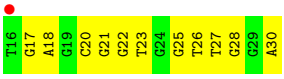


• Molecule 4: Nuclease-like protein

Chain A:



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



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	98.07Å 98.07Å 119.57Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.04 – 2.40 49.04 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.2 (49.04-2.40) 99.5 (49.04-2.40)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.02 (at 2.10Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.239 , 0.255 0.249 , 0.237	Depositor DCC
R_{free} test set	1354 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	53.6	Xtriage
Anisotropy	0.165	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 53.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.036 for -h,-k,l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	7414	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.92% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG

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	R	0.78	0/672	0.92	1/1032 (0.1%)
2	C	0.24	0/3174	0.46	3/4286 (0.1%)
3	X	0.48	0/331	0.93	0/512
4	A	0.38	1/3142 (0.0%)	0.49	3/4243 (0.1%)
5	H	0.73	1/335 (0.3%)	0.88	0/519
All	All	0.42	2/7654 (0.0%)	0.59	7/10592 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	297	PRO	C-O	14.64	1.52	1.23
5	H	17	DG	O3'-P	-7.71	1.51	1.61

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	297	PRO	O-C-N	-10.05	106.61	122.70
2	C	196	SER	N-CA-C	7.69	131.77	111.00
4	A	297	PRO	CA-C-O	7.08	137.20	120.20
2	C	196	SER	CB-CA-C	-6.67	97.42	110.10
2	C	197	GLY	N-CA-C	-6.21	97.57	113.10

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	R	602	0	337	32	0
2	C	3109	0	3095	65	31
3	X	296	0	158	16	0
4	A	3078	0	3066	60	31
5	H	298	0	158	18	0
6	A	1	0	0	0	0
6	C	1	0	0	0	0
7	A	11	0	0	4	0
7	C	14	0	0	2	0
7	H	3	0	0	0	0
7	X	1	0	0	4	0
All	All	7414	0	6814	155	31

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 11.

The worst 5 of 155 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)
4:A:261:ALA:HB2	5:H:27:DT:P	1.91	1.10
2:C:341:ARG:NH2	2:C:358:GLU:OE2	1.88	1.07
2:C:203:ASP:CG	2:C:267:ASN:HB3	1.78	1.04
1:R:6:DC:H42	5:H:25:DG:H1	1.06	0.97
2:C:122:GLU:OE1	7:C:602:HOH:O	1.82	0.97

The worst 5 of 31 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)
2:C:342:TYR:OH	4:A:342:TYR:OH[5_444]	0.19	2.01
2:C:342:TYR:CZ	4:A:342:TYR:CZ[5_444]	0.23	1.97
2:C:342:TYR:CE2	4:A:342:TYR:CE2[5_444]	0.23	1.97
2:C:342:TYR:CD2	4:A:342:TYR:CD2[5_444]	0.30	1.90
2:C:342:TYR:CE1	4:A:342:TYR:CE1[5_444]	0.31	1.89

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
2	C	385/464 (83%)	359 (93%)	26 (7%)	0	100 100
4	A	379/464 (82%)	358 (94%)	20 (5%)	1 (0%)	37 51
All	All	764/928 (82%)	717 (94%)	46 (6%)	1 (0%)	48 65

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	464	ARG

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
2	C	332/383 (87%)	324 (98%)	8 (2%)	44 64
4	A	329/385 (86%)	320 (97%)	9 (3%)	40 60
All	All	661/768 (86%)	644 (97%)	17 (3%)	41 62

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	341	ARG
4	A	464	ARG
2	C	342	TYR
4	A	128	LEU
4	A	184	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	43	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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.

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	R	31/31 (100%)	0.22	0 100 100	28, 35, 49, 55	31 (100%)
2	C	390/464 (84%)	0.65	46 (11%) 10 9	14, 24, 41, 122	390 (100%)
3	X	16/16 (100%)	0.24	1 (6%) 27 25	23, 34, 38, 46	16 (100%)
4	A	388/464 (83%)	0.57	39 (10%) 14 12	14, 24, 40, 122	388 (100%)
5	H	15/15 (100%)	0.13	1 (6%) 25 23	27, 32, 38, 43	15 (100%)
All	All	840/990 (84%)	0.58	87 (10%) 13 11	14, 24, 42, 122	840 (100%)

The worst 5 of 87 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	200	TYR	11.1
2	C	201	LEU	9.2
4	A	235	ALA	8.5
2	C	235	ALA	7.9
2	C	202	PRO	6.8

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
6	MG	C	501	1/1	0.97	0.09	52,52,52,52	1
6	MG	A	501	1/1	0.97	0.07	62,62,62,62	1

6.5 Other polymers (i)

There are no such residues in this entry.