wwPDB X-ray Structure Validation Summary Report (i)

Aug 2, 2023 – 10:30 PM EDT

PDB ID	:	1G9Y
Title	:	HOMING ENDONUCLEASE I-CREI / DNA SUBSTRATE COMPLEX WITH CALCIUM
Authors	:	Chevalier, B.; Monnat, R.J.; Stoddard, B.L.
Deposited on	:	2000-11-28
Resolution	:	2.05 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:

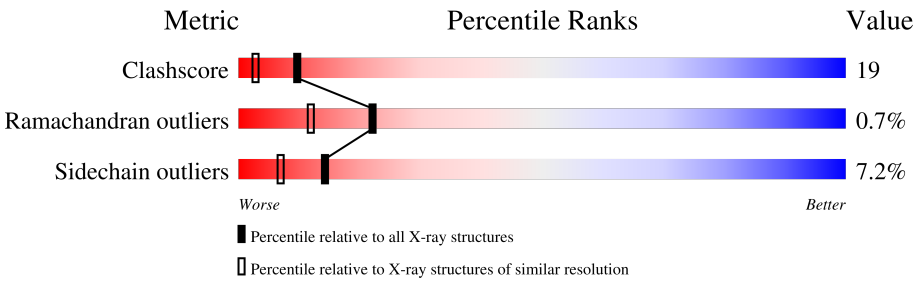
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.34

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 2.05 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))
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	24	54% 42% .
2	D	24	54% 42% .
3	A	152	73% 25% .
3	B	152	61% 34% 5%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3890 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(*CP*GP*AP*AP*AP*CP*TP*GP*TP*CP*TP*CP*AP*CP*GP*AP*CP*GP*TP*TP*TP*TP*GP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	24	485	233	85	144	23	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	24	493	235	95	140	23	0	0	0

- Molecule 3 is a protein called DNA ENDONUCLEASE I-CREI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	152	1237	796	210	230	1	0	0	0
3	B	152	1237	796	210	230	1	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	42	THR	ALA	SEE REMARK 999	UNP P05725
A	110	GLU	TRP	SEE REMARK 999	UNP P05725
A	111	GLN	ARG	SEE REMARK 999	UNP P05725
B	242	THR	ALA	SEE REMARK 999	UNP P05725
B	310	GLU	TRP	SEE REMARK 999	UNP P05725
B	311	GLN	ARG	SEE REMARK 999	UNP P05725

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
4	C	1	1	1	0	0
			Total	Ca		
4	A	1	1	1	0	0
			Total	Ca		
4	B	1	1	1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	C	83	83	83	0	0
			Total	O		
5	D	70	70	70	0	0
			Total	O		
5	A	152	152	152	0	0
			Total	O		
5	B	130	130	130	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. 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. 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.

Note EDS was not executed.

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

Chain C: 54% 42% •

C401 G402 A403 A404 A405 A413 C414 G415 A416 C417 G418 T419 T420 T421 T422 G423 C424

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

Chain D: 54% 42% •

G501 C502 A503 A504 A505 A506 G511 T512 G513 A514 G515 A516 C517 A518 G519 T520 G524

• Molecule 3: DNA ENDONUCLEASE I-CREI

Chain A: 73% 25% •

N2 N6 K7 E8 Y12 I27 Q31 S32 Y33 K36 H37 Q38 Q50 K57 V64 G65 R68 E80 I81 K82 P83 Q92 Q101 A102 N103 L106 P113 S114 A115 K116 E117 S118 K121 F122 L123 E124 V125 C126 T127 W128 R141 K142 E146

T147 V148 R149 A150 V151 L152 D153

• Molecule 3: DNA ENDONUCLEASE I-CREI

Chain B: 61% 34% 5%

N202 T203 I227 K228 P229 Y233 Q238 L239 S240 L241 T246 Q247 K248 T249 Q250 K257 V264 G265 R268 D269 D275 Y276 I277 L278 S279 E280 I281 K282 P283 Q292 L295 K296 Q301 A302 N303 L304 V305 L306 I309 E310 Q311 L312 P313 S314 A315 K316 P319

D320 E324 V325 C326 T327 W328 V329 D330 S338 K339 K342 T343 T344 S345 E346 T347 V348 R349 A350 V351 L352 D353

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	43.20Å 67.80Å 88.30Å 90.00° 91.60° 90.00°	Depositor
Resolution (Å)	26.99 – 2.05	Depositor
% Data completeness (in resolution range)	93.5 (26.99-2.05)	Depositor
R_{merge}	0.06	Depositor
R_{sym}	5.50	Depositor
Refinement program		Depositor
R, R_{free}	0.207 , 0.260	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3890	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

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: CA

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	C	0.31	0/542	0.74	0/834
2	D	0.32	0/554	0.76	0/854
3	A	0.36	0/1260	0.64	0/1700
3	B	0.36	0/1260	0.64	0/1700
All	All	0.35	0/3616	0.68	0/5088

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
1	C	0	1
2	D	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	418	DG	Sidechain
2	D	514	DA	Sidechain

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	C	485	0	273	17	0
2	D	493	0	271	15	0
3	A	1237	0	1270	39	0
3	B	1237	0	1270	65	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
5	A	152	0	0	9	0
5	B	130	0	0	9	0
5	C	83	0	0	2	0
5	D	70	0	0	2	0
All	All	3890	0	3084	127	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 19.

The worst 5 of 127 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:B:282:LYS:HA	3:B:282:LYS:HE3	1.50	0.94
3:B:241:LEU:HD23	3:B:305:VAL:HG13	1.50	0.93
1:C:421:DT:H2"	1:C:422:DT:H5"	1.55	0.87
3:A:118:SER:HB3	3:A:121:LYS:HB2	1.57	0.85
3:A:115:ALA:HB1	3:A:122:PHE:HA	1.64	0.80

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	
3	A	150/152 (99%)	140 (93%)	9 (6%)	1 (1%)	22	12
3	B	150/152 (99%)	139 (93%)	10 (7%)	1 (1%)	22	12
All	All	300/304 (99%)	279 (93%)	19 (6%)	2 (1%)	22	12

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	B	349	ARG
3	A	149	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	
3	A	139/139 (100%)	132 (95%)	7 (5%)	24	16
3	B	139/139 (100%)	126 (91%)	13 (9%)	8	3
All	All	278/278 (100%)	258 (93%)	20 (7%)	14	7

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	301	GLN
3	B	320	ASP
3	B	347	THR
3	B	328	TRP
3	A	152	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	226	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	B	250	GLN
3	A	99	GLN
3	A	103	ASN
3	B	202	ASN

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)

Of 3 ligands modelled in this entry, 3 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 (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)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.