wwPDB X-ray Structure Validation Summary Report (i)

Mar 9, 2018 – 09:29 am GMT

PDB ID	:	2Q10
Title	:	RESTRICTION ENDONUCLEASE BcnI (WILD TYPE)-COGNATE DNA SUBSTRATE COMPLEX
Authors	:	Sokolowska, M.; Kaus-Drobek, M.; Czapinska, H.; Tamulaitis, G.; Szczepanowski, R.H.; Urbanke, C.; Siksnys, V.; Bochtler, M.
Deposited on	:	2007-05-23
Resolution	:	1.75 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 following versions of software and data (see references (1)) were used in the production of this report:

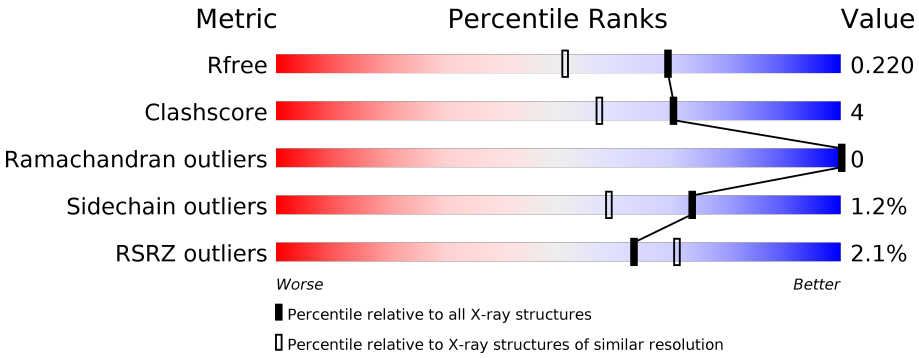
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	trunk30967
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30967

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 1.75 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}	111664	1952 (1.76-1.76)
Clashscore	122126	2072 (1.76-1.76)
Ramachandran outliers	120053	2050 (1.76-1.76)
Sidechain outliers	120020	2050 (1.76-1.76)
RSRZ outliers	108989	1913 (1.76-1.76)

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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	C	11	82% 18%
1	E	11	82% 18%
2	D	11	82% 18%
2	F	11	82% 18%
3	A	238	3% 94% 5%
3	B	238	2% 89% 11%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 5638 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 (5'-D(*AP*AP*CP*CP*CP*GP*GP*AP*GP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	11	Total	C	N	O	P	0	0	0
			223	106	47	60	10			
1	E	11	Total	C	N	O	P	0	0	0
			223	106	47	60	10			

- Molecule 2 is a DNA chain called DNA (5'-D(*CP*TP*CP*CP*GP*GP*GP*TP*TP*GP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	11	Total	C	N	O	P	0	0	0
			222	107	37	68	10			
2	F	11	Total	C	N	O	P	0	0	0
			222	107	37	68	10			

- Molecule 3 is a protein called R.BcnI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	238	Total	C	N	O	S	3	17	0
			2056	1302	362	382	10			
3	B	238	Total	C	N	O	S	0	15	0
			2042	1299	358	373	12			

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

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

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total Cl 1 1	0	0
5	A	1	Total Cl 1 1	0	0

- Molecule 6 is water.

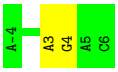
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	C	36	Total O 36 36	0	2
6	D	40	Total O 40 40	0	1
6	E	37	Total O 37 37	0	4
6	F	43	Total O 43 43	0	1
6	A	251	Total O 251 251	0	6
6	B	238	Total O 238 238	0	8

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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 geometryand 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.

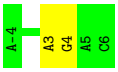
• Molecule 1: DNA (5'-D(*AP*AP*CP*CP*CP*GP*GP*AP*GP*AP*C)-3')

Chain C:



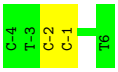
• Molecule 1: DNA (5'-D(*AP*AP*CP*CP*CP*GP*GP*AP*GP*AP*C)-3')

Chain E:



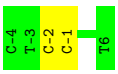
• Molecule 2: DNA (5'-D(*CP*TP*CP*CP*GP*GP*GP*TP*TP*GP*T)-3')

Chain D:



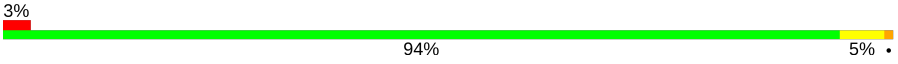
• Molecule 2: DNA (5'-D(*CP*TP*CP*CP*GP*GP*GP*TP*TP*GP*T)-3')

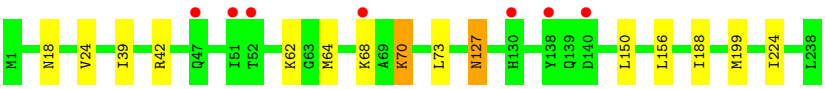
Chain F:



• Molecule 3: R.BcnI

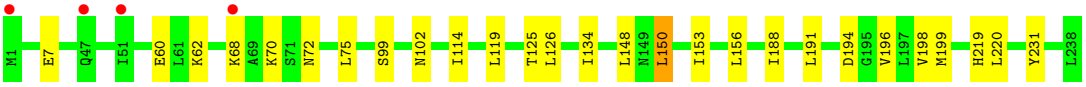
Chain A:





• Molecule 3: R.BcnI

Chain B:


M1
E7
Q47
I51
E60
L61
K62
K68
A69
K70
S71
N72
L75
S99
N102
I114
L119
T125
L126
I134
L148
N149
L150
I153
L156
I188
L191
D194
G195
V196
L197
V198
M199
H219
L220
Y231
L238

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	81.35Å 87.55Å 79.16Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.72 – 1.75 29.36 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.8 (29.72-1.75) 99.8 (29.36-1.75)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	2.83 (at 1.75Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.174 , 0.213 0.184 , 0.220	Depositor DCC
R_{free} test set	2882 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	13.1	Xtriage
Anisotropy	0.207	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 43.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.000 for l,-k,h	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5638	wwPDB-VP
Average B, all atoms (Å^2)	11.0	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 60.76 % 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 1.4506e-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: CA, CL

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	0.38	0/251	0.71	0/385
1	E	0.39	0/251	0.73	0/385
2	D	0.38	0/247	0.84	2/380 (0.5%)
2	F	0.38	0/247	0.81	2/380 (0.5%)
3	A	0.53	0/2090	0.66	0/2808
3	B	0.55	0/2076	0.68	1/2787 (0.0%)
All	All	0.52	0/5162	0.69	5/7125 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	-2	DC	C1'-O4'-C4'	-6.55	103.55	110.10
2	F	-2	DC	C1'-O4'-C4'	-6.22	103.88	110.10
2	D	-1	DC	C1'-O4'-C4'	-5.94	104.16	110.10
3	B	194	ASP	CB-CG-OD1	5.55	123.30	118.30
2	F	-1	DC	C1'-O4'-C4'	-5.48	104.62	110.10

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	C	223	0	123	1	0
1	E	223	0	123	2	0
2	D	222	0	127	0	0
2	F	222	0	127	0	0
3	A	2056	0	2098	13	0
3	B	2042	0	2105	22	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
6	A	251	0	0	3	0
6	B	238	0	0	3	0
6	C	36	0	0	0	0
6	D	40	0	0	0	0
6	E	37	0	0	0	0
6	F	43	0	0	0	0
All	All	5638	0	4703	38	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 4.

The worst 5 of 38 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:126:LEU:HD21	3:B:199:MET:HE1	1.57	0.87
3:A:62:LYS:HG3	3:A:156[B]:LEU:HD11	1.63	0.79
3:A:199:MET:HG3	6:A:762:HOH:O	1.88	0.73
3:B:68[B]:LYS:HD3	3:B:70[B]:LYS:HD3	1.75	0.66
3:B:62:LYS:HG3	3:B:156[A]:LEU:HD11	1.77	0.65

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	253/238 (106%)	242 (96%)	11 (4%)	0	100	100
3	B	250/238 (105%)	238 (95%)	12 (5%)	0	100	100
All	All	503/476 (106%)	480 (95%)	23 (5%)	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	
3	A	231/215 (107%)	227 (98%)	4 (2%)	63	45
3	B	229/215 (106%)	227 (99%)	2 (1%)	81	69
All	All	460/430 (107%)	454 (99%)	6 (1%)	74	57

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	70[B]	LYS
3	B	219	HIS
3	A	127	ASN
3	A	70[A]	LYS
3	B	150	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	127	ASN
3	B	84	GLN
3	B	102	ASN
3	B	135	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry

Of 5 ligands modelled in this entry, 5 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	C	11/11 (100%)	-0.26	0 100 100	5, 7, 13, 13	0
1	E	11/11 (100%)	-0.18	0 100 100	4, 7, 14, 15	0
2	D	11/11 (100%)	-0.20	0 100 100	6, 7, 14, 20	0
2	F	11/11 (100%)	-0.27	0 100 100	6, 7, 11, 19	0
3	A	238/238 (100%)	0.11	7 (2%) 51 57	6, 10, 20, 30	14 (5%)
3	B	238/238 (100%)	-0.01	4 (1%) 70 78	6, 10, 16, 26	13 (5%)
All	All	520/520 (100%)	0.03	11 (2%) 63 71	4, 10, 18, 30	27 (5%)

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	51	ILE	5.7
3	A	52	THR	4.6
3	A	47	GLN	3.8
3	B	51	ILE	3.4
3	B	68[A]	LYS	2.8

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 carbohydrates 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, 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
4	CA	C	703	1/1	0.91	0.19	22,22,22,22	1
5	CL	A	705	1/1	0.99	0.08	17,17,17,17	0
5	CL	B	704	1/1	0.99	0.04	12,12,12,12	0
4	CA	B	702	1/1	1.00	0.03	4,4,4,4	0
4	CA	A	701	1/1	1.00	0.03	6,6,6,6	0

6.5 Other polymers (i)

There are no such residues in this entry.