wwPDB X-ray Structure Validation Summary Report (i)

Nov 4, 2024 – 03:01 AM EST

PDB ID	:	2F03
Title	:	Crystal structure of tetrameric restriction endonuclease SfiI in complex with cognate DNA (partial bound form)
Authors	:	Aggarwal, A.K.; Vanamee, E.S.; Viadiu, H.
Deposited on	:	2005-11-11
Resolution	:	3.05 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:

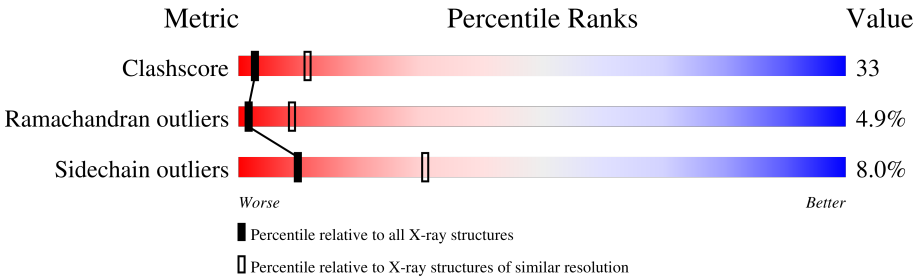
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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 3.05 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	180529	2399 (3.10-3.02)
Ramachandran outliers	177936	2269 (3.10-3.02)
Sidechain outliers	177891	2268 (3.10-3.02)

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	1-E	21	24% 14% 62%
1	1-G	21	24% 76%
1	2-E	21	5% 14% 14% 67%
1	2-G	21	24% 76%
2	1-F	21	5% 10% 19% 67%
2	1-H	21	5% 19% 76%
2	2-F	21	14% 24% 62%
2	2-H	21	5% 19% 76%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	1-A	269	44% 48% 7%
3	1-C	269	45% 46% 9%
3	2-A	269	46% 46% 7%
3	2-C	269	46% 45% 9%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 9827 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*TP*G*TP*GP*GP*CP*CP*AP*AP*CP*AP*AP*GP*GP*CP*CP*TP*AP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	1-E	8	Total	C	N	O	P	0	8	0
			160	77	31	45	7			
1	2-E	7	Total	C	N	O	P	0	7	0
			144	68	28	41	7			
1	1-G	5	Total	C	N	O	P	0	5	0
			99	48	18	29	4			
1	2-G	5	Total	C	N	O	P	0	5	0
			99	48	18	29	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	1-F	7	Total	C	N	O	P	0	7	0
			143	68	25	43	7			
2	2-F	8	Total	C	N	O	P	0	8	0
			162	78	30	47	7			
2	1-H	5	Total	C	N	O	P	0	5	0
			100	48	21	27	4			
2	2-H	5	Total	C	N	O	P	0	5	0
			100	48	21	27	4			

- Molecule 3 is a protein called Type II restriction enzyme SfiI.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	1-A	269	Total	C	N	O	S	Se	0	269	0
			2155	1366	380	404	1	4			
3	2-A	269	Total	C	N	O	S	Se	0	269	0
			2155	1366	380	404	1	4			
3	1-C	269	Total	C	N	O	S	Se	0	269	0
			2142	1359	377	401	1	4			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	2-C	269	Total	C	N	O	S	Se	0	269	0
			2142	1359	377	401	1	4			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	modified residue	UNP O52512
A	119	MSE	MET	modified residue	UNP O52512
A	134	MSE	MET	modified residue	UNP O52512
A	143	MSE	MET	modified residue	UNP O52512
C	1	MSE	MET	modified residue	UNP O52512
C	119	MSE	MET	modified residue	UNP O52512
C	134	MSE	MET	modified residue	UNP O52512
C	143	MSE	MET	modified residue	UNP O52512

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	1-A	1	Total	Ca	0	1
			1	1		
4	2-A	1	Total	Ca	0	1
			1	1		
4	1-C	1	Total	Ca	0	1
			1	1		
4	2-C	1	Total	Ca	0	1
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	1-E	5	Total	O	0	5
			5	5		
5	2-E	93	Total	O	0	93
			93	93		
5	1-F	3	Total	O	0	3
			3	3		
5	2-F	18	Total	O	0	18
			18	18		
5	1-G	3	Total	O	0	3
			3	3		
5	1-A	55	Total	O	0	55
			55	55		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	1-C	45	45	45	0	45

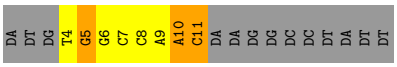
3 Residue-property plots (i)

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: DNA (5'-D(*AP*TP*G*TP*GP*GP*CP*CP*AP*AP*CP*AP*AP*GP*GP*CP*CP*TP*AP*TP*T)-3')

Chain 1-E:



• Molecule 1: DNA (5'-D(*AP*TP*G*TP*GP*GP*CP*CP*AP*AP*CP*AP*AP*GP*GP*CP*CP*TP*AP*TP*T)-3')

Chain 1-G:

• Molecule 1: DNA (5'-D(*AP*TP*G*TP*GP*GP*CP*CP*AP*AP*CP*AP*AP*GP*GP*CP*CP*TP*AP*TP*T)-3')

Chain 2-E:

• Molecule 1: DNA (5'-D(*AP*TP*G*TP*GP*GP*CP*CP*AP*AP*CP*AP*AP*GP*GP*CP*CP*TP*AP*TP*T)-3')

Chain 2-G:

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

Chain 1-F:

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

Chain 1-H:

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

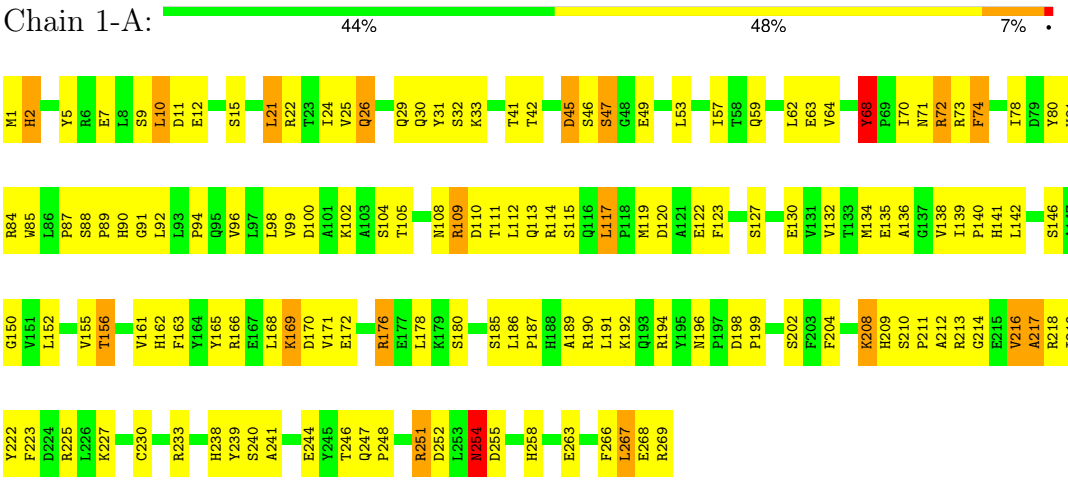
Chain 2-F:

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

Chain 2-H:

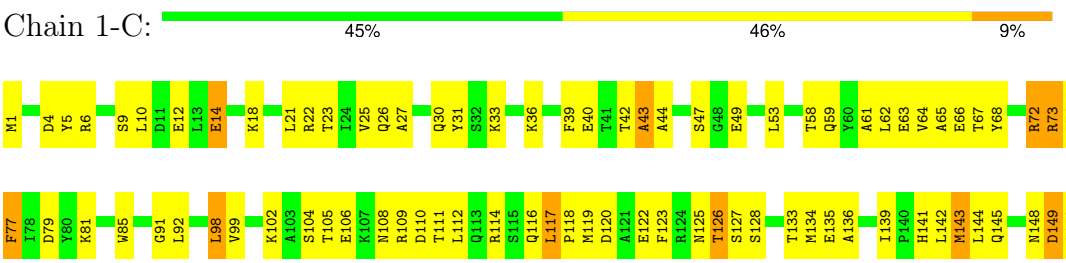
• Molecule 3: Type II restriction enzyme SfiI

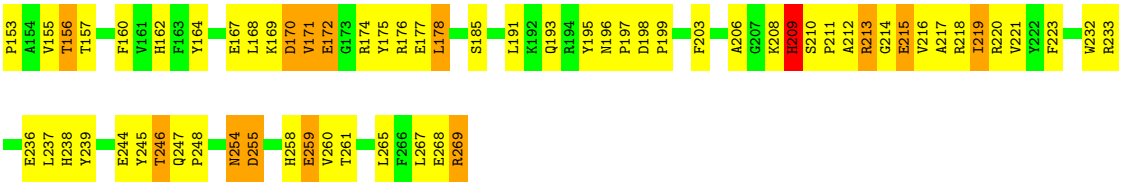
Chain 1-A:



• Molecule 3: Type II restriction enzyme SfiI

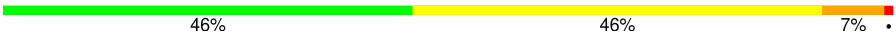
Chain 1-C:

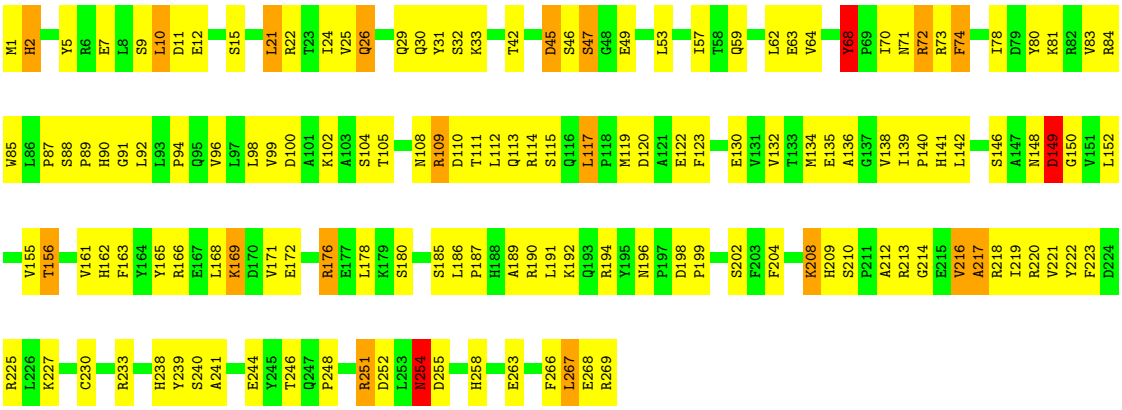




• Molecule 3: Type II restriction enzyme SfiI

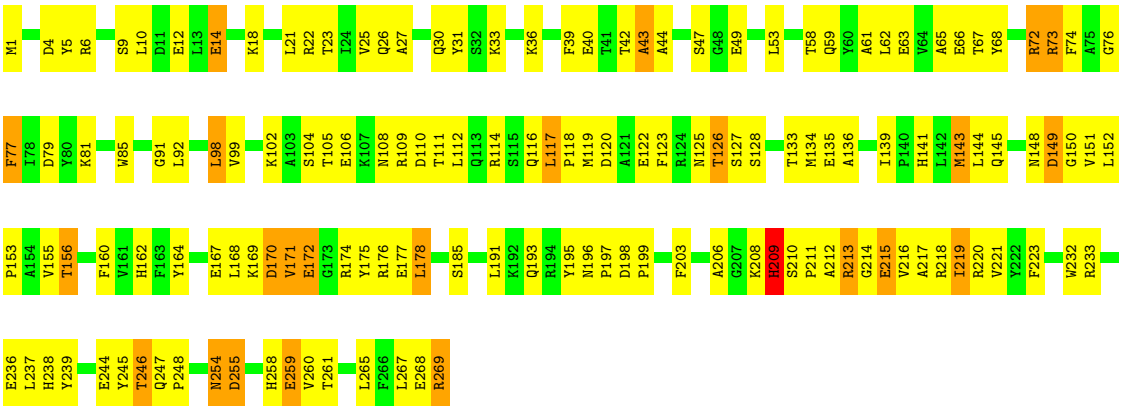
Chain 2-A:





• Molecule 3: Type II restriction enzyme SfiI

Chain 2-C:



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α, β, γ	85.60Å 85.60Å 418.70Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 3.05	Depositor
% Data completeness (in resolution range)	95.8 (50.00-3.05)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.241 , 0.293	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	9827	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

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	1-E	0.63	0/179	1.68	7/274 (2.6%)
1	1-G	0.69	0/110	1.31	0/168
1	2-E	0.47	0/161	1.66	10/246 (4.1%)
1	2-G	0.69	0/110	1.31	0/168
2	1-F	0.55	0/159	1.66	8/243 (3.3%)
2	1-H	0.75	0/112	1.04	0/171
2	2-F	0.78	0/181	1.83	13/278 (4.7%)
2	2-H	0.75	0/112	1.04	0/171
3	1-A	0.45	0/2202	0.69	0/2984
3	1-C	0.50	2/2189 (0.1%)	0.68	0/2968
3	2-A	0.45	0/2202	0.69	0/2984
3	2-C	0.50	2/2189 (0.1%)	0.68	0/2968
All	All	0.50	4/9906 (0.0%)	0.84	38/13623 (0.3%)

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	1-C	0	1
3	2-C	0	1
All	All	0	2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	1-C	143[A]	MSE	CG-SE	-5.50	1.76	1.95
3	2-C	143[B]	MSE	CG-SE	-5.50	1.76	1.95
3	1-C	255[A]	ASP	C-N	5.36	1.46	1.34
3	2-C	255[B]	ASP	C-N	5.36	1.46	1.34

The worst 5 of 38 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	2-F	11[A]	DG	O4'-C1'-C2'	8.45	112.66	105.90
2	2-F	11[A]	DG	O4'-C1'-N9	7.69	113.38	108.00
2	1-F	14[A]	DG	O4'-C1'-C2'	6.38	111.00	105.90
2	1-F	13[A]	DT	O4'-C1'-N1	6.31	112.42	108.00
2	2-F	10[A]	DT	O4'-C1'-N1	6.23	112.36	108.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	1-C	255[A]	ASP	Mainchain
3	2-C	255[B]	ASP	Mainchain

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	1-E	160	0	90	13	0
1	1-G	99	0	58	11	0
1	2-E	144	0	79	4	0
1	2-G	99	0	58	12	0
2	1-F	143	0	80	9	0
2	1-H	100	0	57	12	0
2	2-F	162	0	91	6	0
2	2-H	100	0	57	13	0
3	1-A	2155	0	2096	167	0
3	1-C	2142	0	2071	134	0
3	2-A	2155	0	2096	146	0
3	2-C	2142	0	2071	127	0
4	1-A	1	0	0	0	0
4	1-C	1	0	0	0	0
4	2-A	1	0	0	0	0
4	2-C	1	0	0	0	0
5	1-A	55	0	0	10	0
5	1-C	45	0	0	6	0
5	1-E	5	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	1-F	3	0	0	0	0
5	1-G	3	0	0	0	0
5	2-E	93	0	0	0	0
5	2-F	18	0	0	1	0
All	All	9827	0	8904	607	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 33.

The worst 5 of 607 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:F:12[A]:DT:H4'	2:F:13[A]:DT:H5'	1.23	1.13
3:A:22[A]:ARG:HH12	3:A:26[A]:GLN:HG2	1.17	1.06
3:A:22[B]:ARG:HH12	3:A:26[B]:GLN:HG2	1.17	1.06
3:A:45[A]:ASP:H	3:A:49[A]:GLU:HG3	1.19	1.04
3:A:45[B]:ASP:H	3:A:49[B]:GLU:HG3	1.19	1.04

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	1-A	267/269 (99%)	232 (87%)	23 (9%)	12 (4%)	2	9
3	1-C	267/269 (99%)	226 (85%)	27 (10%)	14 (5%)	1	8
3	2-A	267/269 (99%)	232 (87%)	23 (9%)	12 (4%)	2	9
3	2-C	267/269 (99%)	226 (85%)	27 (10%)	14 (5%)	1	8
All	All	1068/1076 (99%)	916 (86%)	100 (9%)	52 (5%)	2	9

5 of 52 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	1-A	47[A]	SER
3	1-A	130[A]	GLU
3	1-A	149[A]	ASP
3	1-A	254[A]	ASN
3	1-C	43[A]	ALA

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	1-A	228/233 (98%)	210 (92%)	18 (8%)	10	31
3	1-C	224/233 (96%)	206 (92%)	18 (8%)	10	31
3	2-A	228/233 (98%)	210 (92%)	18 (8%)	10	31
3	2-C	224/233 (96%)	206 (92%)	18 (8%)	10	31
All	All	904/932 (97%)	832 (92%)	72 (8%)	10	31

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	2-C	77[B]	PHE
3	2-C	269[B]	ARG
3	2-C	99[B]	VAL
3	2-C	209[B]	HIS
3	1-C	105[A]	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
3	2-A	26[B]	GLN
3	2-A	59[B]	GLN
3	2-C	141[B]	HIS
3	2-A	30[B]	GLN
3	2-A	141[B]	HIS

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 4 ligands modelled in this entry, 4 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

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.