wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 10:56 AM GMT

PDB ID	:	3NIC
Title	:	DNA binding and cleavage by the GIY-YIG endonuclease R.Eco29kI inactive variant Y49F
Authors	:	Mak, A.N.S.; Lambert, A.R.; Stoddard, B.L.
Deposited on	:	2010-06-15
Resolution	:	2.80 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
http://wwpdb.org/validation/2016/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:

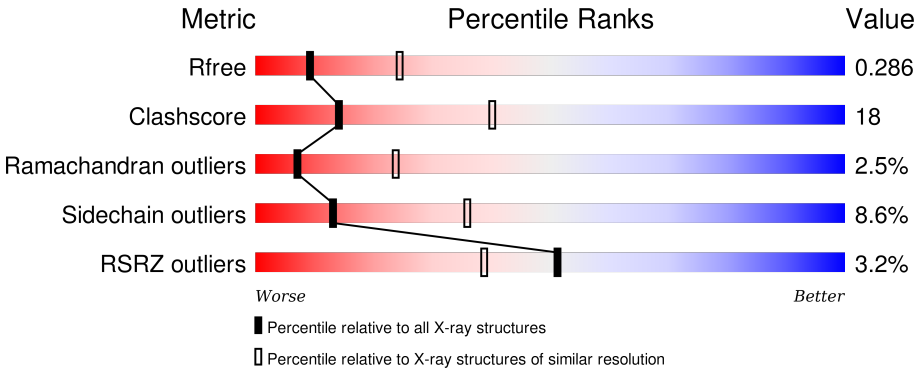
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.80 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}	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	A	235	4% 61% 25% • 11%
1	B	235	3% 55% 30% • 11%
1	C	235	% 59% 26% • • 11%
1	D	235	4% 59% 25% 5% 11%
1	E	235	4% 62% 23% • 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	235	3% 63% 24% 11%
1	G	235	3% 59% 26% 11%
1	H	235	3% 60% 23% 5% 11%
2	I	22	64% 36%
2	K	22	14% 36% 50%
2	M	22	18% 45% 36%
2	O	22	9% 9% 55% 36%
3	J	22	5% 50% 45%
3	L	22	5% 64% 32%
3	N	22	14% 41% 45%
3	P	22	5% 9% 32% 59%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	PO4	F	216	-	-	-	X
4	PO4	H	216	-	-	X	-

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 17279 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 protein called Eco29kIR.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	210	Total	C	N	O	S	0	0	0
			1671	1066	297	304	4			
1	B	210	Total	C	N	O	S	0	0	0
			1683	1072	299	308	4			
1	C	209	Total	C	N	O	S	0	0	0
			1635	1042	288	301	4			
1	D	209	Total	C	N	O	S	0	0	0
			1649	1053	290	302	4			
1	E	209	Total	C	N	O	S	0	0	0
			1631	1043	285	299	4			
1	F	209	Total	C	N	O	S	0	0	0
			1627	1038	285	300	4			
1	G	209	Total	C	N	O	S	0	0	0
			1612	1029	281	298	4			
1	H	209	Total	C	N	O	S	0	0	0
			1642	1050	286	302	4			

There are 192 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	EXPRESSION TAG	UNP Q46944
A	-19	ALA	-	EXPRESSION TAG	UNP Q46944
A	-18	SER	-	EXPRESSION TAG	UNP Q46944
A	-17	SER	-	EXPRESSION TAG	UNP Q46944
A	-16	HIS	-	EXPRESSION TAG	UNP Q46944
A	-15	HIS	-	EXPRESSION TAG	UNP Q46944
A	-14	HIS	-	EXPRESSION TAG	UNP Q46944
A	-13	HIS	-	EXPRESSION TAG	UNP Q46944
A	-12	HIS	-	EXPRESSION TAG	UNP Q46944
A	-11	HIS	-	EXPRESSION TAG	UNP Q46944
A	-10	SER	-	EXPRESSION TAG	UNP Q46944
A	-9	SER	-	EXPRESSION TAG	UNP Q46944
A	-8	GLY	-	EXPRESSION TAG	UNP Q46944

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	LEU	-	EXPRESSION TAG	UNP Q46944
A	-6	VAL	-	EXPRESSION TAG	UNP Q46944
A	-5	PRO	-	EXPRESSION TAG	UNP Q46944
A	-4	ARG	-	EXPRESSION TAG	UNP Q46944
A	-3	GLY	-	EXPRESSION TAG	UNP Q46944
A	-2	SER	-	EXPRESSION TAG	UNP Q46944
A	-1	SER	-	EXPRESSION TAG	UNP Q46944
A	0	MET	-	EXPRESSION TAG	UNP Q46944
A	1	GLY	-	EXPRESSION TAG	UNP Q46944
A	49	PHE	TYR	ENGINEERED MUTATION	UNP Q46944
A	69	LYS	LEU	ENGINEERED MUTATION	UNP Q46944
B	-20	MET	-	EXPRESSION TAG	UNP Q46944
B	-19	ALA	-	EXPRESSION TAG	UNP Q46944
B	-18	SER	-	EXPRESSION TAG	UNP Q46944
B	-17	SER	-	EXPRESSION TAG	UNP Q46944
B	-16	HIS	-	EXPRESSION TAG	UNP Q46944
B	-15	HIS	-	EXPRESSION TAG	UNP Q46944
B	-14	HIS	-	EXPRESSION TAG	UNP Q46944
B	-13	HIS	-	EXPRESSION TAG	UNP Q46944
B	-12	HIS	-	EXPRESSION TAG	UNP Q46944
B	-11	HIS	-	EXPRESSION TAG	UNP Q46944
B	-10	SER	-	EXPRESSION TAG	UNP Q46944
B	-9	SER	-	EXPRESSION TAG	UNP Q46944
B	-8	GLY	-	EXPRESSION TAG	UNP Q46944
B	-7	LEU	-	EXPRESSION TAG	UNP Q46944
B	-6	VAL	-	EXPRESSION TAG	UNP Q46944
B	-5	PRO	-	EXPRESSION TAG	UNP Q46944
B	-4	ARG	-	EXPRESSION TAG	UNP Q46944
B	-3	GLY	-	EXPRESSION TAG	UNP Q46944
B	-2	SER	-	EXPRESSION TAG	UNP Q46944
B	-1	SER	-	EXPRESSION TAG	UNP Q46944
B	0	MET	-	EXPRESSION TAG	UNP Q46944
B	1	GLY	-	EXPRESSION TAG	UNP Q46944
B	49	PHE	TYR	ENGINEERED MUTATION	UNP Q46944
B	69	LYS	LEU	ENGINEERED MUTATION	UNP Q46944
C	-20	MET	-	EXPRESSION TAG	UNP Q46944
C	-19	ALA	-	EXPRESSION TAG	UNP Q46944
C	-18	SER	-	EXPRESSION TAG	UNP Q46944
C	-17	SER	-	EXPRESSION TAG	UNP Q46944
C	-16	HIS	-	EXPRESSION TAG	UNP Q46944
C	-15	HIS	-	EXPRESSION TAG	UNP Q46944
C	-14	HIS	-	EXPRESSION TAG	UNP Q46944

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-13	HIS	-	EXPRESSION TAG	UNP Q46944
C	-12	HIS	-	EXPRESSION TAG	UNP Q46944
C	-11	HIS	-	EXPRESSION TAG	UNP Q46944
C	-10	SER	-	EXPRESSION TAG	UNP Q46944
C	-9	SER	-	EXPRESSION TAG	UNP Q46944
C	-8	GLY	-	EXPRESSION TAG	UNP Q46944
C	-7	LEU	-	EXPRESSION TAG	UNP Q46944
C	-6	VAL	-	EXPRESSION TAG	UNP Q46944
C	-5	PRO	-	EXPRESSION TAG	UNP Q46944
C	-4	ARG	-	EXPRESSION TAG	UNP Q46944
C	-3	GLY	-	EXPRESSION TAG	UNP Q46944
C	-2	SER	-	EXPRESSION TAG	UNP Q46944
C	-1	SER	-	EXPRESSION TAG	UNP Q46944
C	0	MET	-	EXPRESSION TAG	UNP Q46944
C	1	GLY	-	EXPRESSION TAG	UNP Q46944
C	49	PHE	TYR	ENGINEERED MUTATION	UNP Q46944
C	69	LYS	LEU	ENGINEERED MUTATION	UNP Q46944
D	-20	MET	-	EXPRESSION TAG	UNP Q46944
D	-19	ALA	-	EXPRESSION TAG	UNP Q46944
D	-18	SER	-	EXPRESSION TAG	UNP Q46944
D	-17	SER	-	EXPRESSION TAG	UNP Q46944
D	-16	HIS	-	EXPRESSION TAG	UNP Q46944
D	-15	HIS	-	EXPRESSION TAG	UNP Q46944
D	-14	HIS	-	EXPRESSION TAG	UNP Q46944
D	-13	HIS	-	EXPRESSION TAG	UNP Q46944
D	-12	HIS	-	EXPRESSION TAG	UNP Q46944
D	-11	HIS	-	EXPRESSION TAG	UNP Q46944
D	-10	SER	-	EXPRESSION TAG	UNP Q46944
D	-9	SER	-	EXPRESSION TAG	UNP Q46944
D	-8	GLY	-	EXPRESSION TAG	UNP Q46944
D	-7	LEU	-	EXPRESSION TAG	UNP Q46944
D	-6	VAL	-	EXPRESSION TAG	UNP Q46944
D	-5	PRO	-	EXPRESSION TAG	UNP Q46944
D	-4	ARG	-	EXPRESSION TAG	UNP Q46944
D	-3	GLY	-	EXPRESSION TAG	UNP Q46944
D	-2	SER	-	EXPRESSION TAG	UNP Q46944
D	-1	SER	-	EXPRESSION TAG	UNP Q46944
D	0	MET	-	EXPRESSION TAG	UNP Q46944
D	1	GLY	-	EXPRESSION TAG	UNP Q46944
D	49	PHE	TYR	ENGINEERED MUTATION	UNP Q46944
D	69	LYS	LEU	ENGINEERED MUTATION	UNP Q46944
E	-20	MET	-	EXPRESSION TAG	UNP Q46944

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	-19	ALA	-	EXPRESSION TAG	UNP Q46944
E	-18	SER	-	EXPRESSION TAG	UNP Q46944
E	-17	SER	-	EXPRESSION TAG	UNP Q46944
E	-16	HIS	-	EXPRESSION TAG	UNP Q46944
E	-15	HIS	-	EXPRESSION TAG	UNP Q46944
E	-14	HIS	-	EXPRESSION TAG	UNP Q46944
E	-13	HIS	-	EXPRESSION TAG	UNP Q46944
E	-12	HIS	-	EXPRESSION TAG	UNP Q46944
E	-11	HIS	-	EXPRESSION TAG	UNP Q46944
E	-10	SER	-	EXPRESSION TAG	UNP Q46944
E	-9	SER	-	EXPRESSION TAG	UNP Q46944
E	-8	GLY	-	EXPRESSION TAG	UNP Q46944
E	-7	LEU	-	EXPRESSION TAG	UNP Q46944
E	-6	VAL	-	EXPRESSION TAG	UNP Q46944
E	-5	PRO	-	EXPRESSION TAG	UNP Q46944
E	-4	ARG	-	EXPRESSION TAG	UNP Q46944
E	-3	GLY	-	EXPRESSION TAG	UNP Q46944
E	-2	SER	-	EXPRESSION TAG	UNP Q46944
E	-1	SER	-	EXPRESSION TAG	UNP Q46944
E	0	MET	-	EXPRESSION TAG	UNP Q46944
E	1	GLY	-	EXPRESSION TAG	UNP Q46944
E	49	PHE	TYR	ENGINEERED MUTATION	UNP Q46944
E	69	LYS	LEU	ENGINEERED MUTATION	UNP Q46944
F	-20	MET	-	EXPRESSION TAG	UNP Q46944
F	-19	ALA	-	EXPRESSION TAG	UNP Q46944
F	-18	SER	-	EXPRESSION TAG	UNP Q46944
F	-17	SER	-	EXPRESSION TAG	UNP Q46944
F	-16	HIS	-	EXPRESSION TAG	UNP Q46944
F	-15	HIS	-	EXPRESSION TAG	UNP Q46944
F	-14	HIS	-	EXPRESSION TAG	UNP Q46944
F	-13	HIS	-	EXPRESSION TAG	UNP Q46944
F	-12	HIS	-	EXPRESSION TAG	UNP Q46944
F	-11	HIS	-	EXPRESSION TAG	UNP Q46944
F	-10	SER	-	EXPRESSION TAG	UNP Q46944
F	-9	SER	-	EXPRESSION TAG	UNP Q46944
F	-8	GLY	-	EXPRESSION TAG	UNP Q46944
F	-7	LEU	-	EXPRESSION TAG	UNP Q46944
F	-6	VAL	-	EXPRESSION TAG	UNP Q46944
F	-5	PRO	-	EXPRESSION TAG	UNP Q46944
F	-4	ARG	-	EXPRESSION TAG	UNP Q46944
F	-3	GLY	-	EXPRESSION TAG	UNP Q46944
F	-2	SER	-	EXPRESSION TAG	UNP Q46944

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	SER	-	EXPRESSION TAG	UNP Q46944
F	0	MET	-	EXPRESSION TAG	UNP Q46944
F	1	GLY	-	EXPRESSION TAG	UNP Q46944
F	49	PHE	TYR	ENGINEERED MUTATION	UNP Q46944
F	69	LYS	LEU	ENGINEERED MUTATION	UNP Q46944
G	-20	MET	-	EXPRESSION TAG	UNP Q46944
G	-19	ALA	-	EXPRESSION TAG	UNP Q46944
G	-18	SER	-	EXPRESSION TAG	UNP Q46944
G	-17	SER	-	EXPRESSION TAG	UNP Q46944
G	-16	HIS	-	EXPRESSION TAG	UNP Q46944
G	-15	HIS	-	EXPRESSION TAG	UNP Q46944
G	-14	HIS	-	EXPRESSION TAG	UNP Q46944
G	-13	HIS	-	EXPRESSION TAG	UNP Q46944
G	-12	HIS	-	EXPRESSION TAG	UNP Q46944
G	-11	HIS	-	EXPRESSION TAG	UNP Q46944
G	-10	SER	-	EXPRESSION TAG	UNP Q46944
G	-9	SER	-	EXPRESSION TAG	UNP Q46944
G	-8	GLY	-	EXPRESSION TAG	UNP Q46944
G	-7	LEU	-	EXPRESSION TAG	UNP Q46944
G	-6	VAL	-	EXPRESSION TAG	UNP Q46944
G	-5	PRO	-	EXPRESSION TAG	UNP Q46944
G	-4	ARG	-	EXPRESSION TAG	UNP Q46944
G	-3	GLY	-	EXPRESSION TAG	UNP Q46944
G	-2	SER	-	EXPRESSION TAG	UNP Q46944
G	-1	SER	-	EXPRESSION TAG	UNP Q46944
G	0	MET	-	EXPRESSION TAG	UNP Q46944
G	1	GLY	-	EXPRESSION TAG	UNP Q46944
G	49	PHE	TYR	ENGINEERED MUTATION	UNP Q46944
G	69	LYS	LEU	ENGINEERED MUTATION	UNP Q46944
H	-20	MET	-	EXPRESSION TAG	UNP Q46944
H	-19	ALA	-	EXPRESSION TAG	UNP Q46944
H	-18	SER	-	EXPRESSION TAG	UNP Q46944
H	-17	SER	-	EXPRESSION TAG	UNP Q46944
H	-16	HIS	-	EXPRESSION TAG	UNP Q46944
H	-15	HIS	-	EXPRESSION TAG	UNP Q46944
H	-14	HIS	-	EXPRESSION TAG	UNP Q46944
H	-13	HIS	-	EXPRESSION TAG	UNP Q46944
H	-12	HIS	-	EXPRESSION TAG	UNP Q46944
H	-11	HIS	-	EXPRESSION TAG	UNP Q46944
H	-10	SER	-	EXPRESSION TAG	UNP Q46944
H	-9	SER	-	EXPRESSION TAG	UNP Q46944
H	-8	GLY	-	EXPRESSION TAG	UNP Q46944

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	-7	LEU	-	EXPRESSION TAG	UNP Q46944
H	-6	VAL	-	EXPRESSION TAG	UNP Q46944
H	-5	PRO	-	EXPRESSION TAG	UNP Q46944
H	-4	ARG	-	EXPRESSION TAG	UNP Q46944
H	-3	GLY	-	EXPRESSION TAG	UNP Q46944
H	-2	SER	-	EXPRESSION TAG	UNP Q46944
H	-1	SER	-	EXPRESSION TAG	UNP Q46944
H	0	MET	-	EXPRESSION TAG	UNP Q46944
H	1	GLY	-	EXPRESSION TAG	UNP Q46944
H	49	PHE	TYR	ENGINEERED MUTATION	UNP Q46944
H	69	LYS	LEU	ENGINEERED MUTATION	UNP Q46944

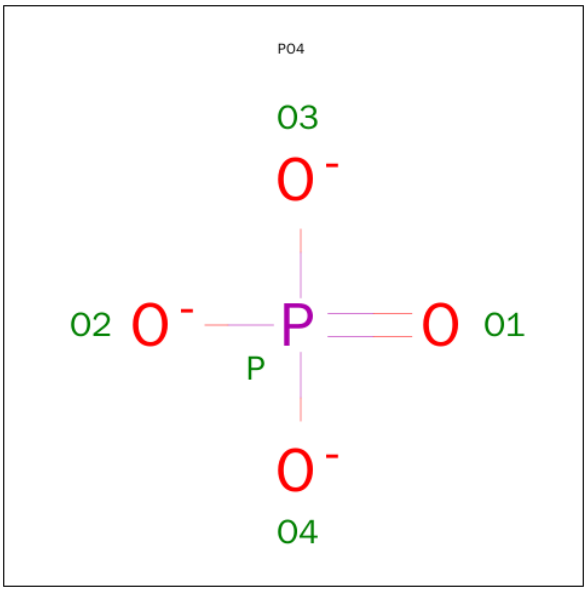
- Molecule 2 is a DNA chain called DNA (5'-D(P*CP*GP*GP*GP*AP*GP*GP*CP*CP*CP *GP*CP*GP*GP*GP*CP*CP*GP*CP*CP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	M	22	Total	C	N	O	P	0	0	0
			453	210	90	131	22			
2	I	22	Total	C	N	O	P	0	0	0
			453	210	90	131	22			
2	K	22	Total	C	N	O	P	0	0	0
			453	210	90	131	22			
2	O	22	Total	C	N	O	P	0	0	0
			453	210	90	131	22			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	N	22	Total	C	N	O	P	0	0	0
			449	209	85	133	22			
3	J	22	Total	C	N	O	P	0	0	0
			449	209	85	133	22			
3	L	22	Total	C	N	O	P	0	0	0
			449	209	85	133	22			
3	P	22	Total	C	N	O	P	0	0	0
			449	209	85	133	22			

- Molecule 4 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O P 5 4 1	0	0
4	B	1	Total O P 5 4 1	0	0
4	B	1	Total O P 5 4 1	0	0
4	D	1	Total O P 5 4 1	0	0
4	D	1	Total O P 5 4 1	0	0
4	E	1	Total O P 5 4 1	0	0
4	F	1	Total O P 5 4 1	0	0
4	F	1	Total O P 5 4 1	0	0
4	G	1	Total O P 5 4 1	0	0
4	H	1	Total O P 5 4 1	0	0
4	H	1	Total O P 5 4 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	51	Total O 51 51	0	0

Continued on next page...

Continued from previous page...

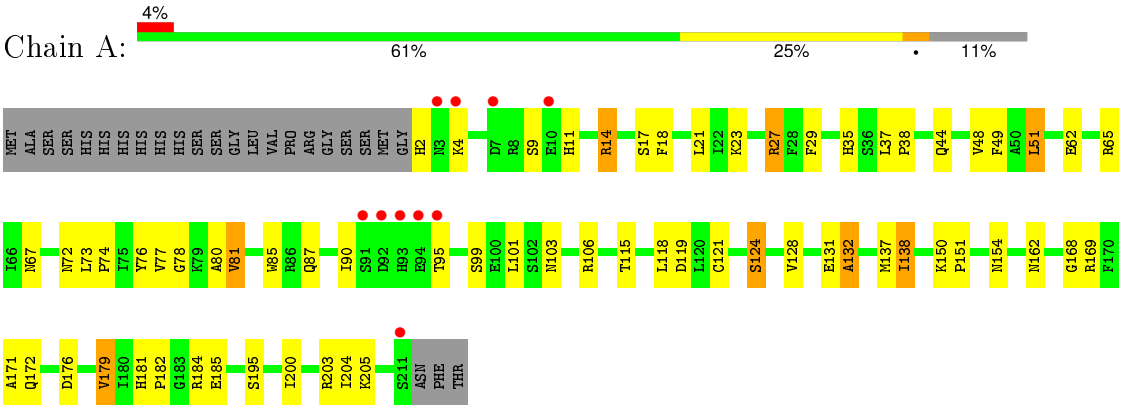
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	56	Total O 56 56	0	0
5	C	60	Total O 60 60	0	0
5	D	35	Total O 35 35	0	0
5	E	46	Total O 46 46	0	0
5	F	59	Total O 59 59	0	0
5	G	37	Total O 37 37	0	0
5	H	29	Total O 29 29	0	0
5	M	11	Total O 11 11	0	0
5	N	10	Total O 10 10	0	0
5	I	5	Total O 5 5	0	0
5	J	12	Total O 12 12	0	0
5	K	11	Total O 11 11	0	0
5	L	19	Total O 19 19	0	0
5	O	7	Total O 7 7	0	0
5	P	18	Total O 18 18	0	0

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 errors 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.

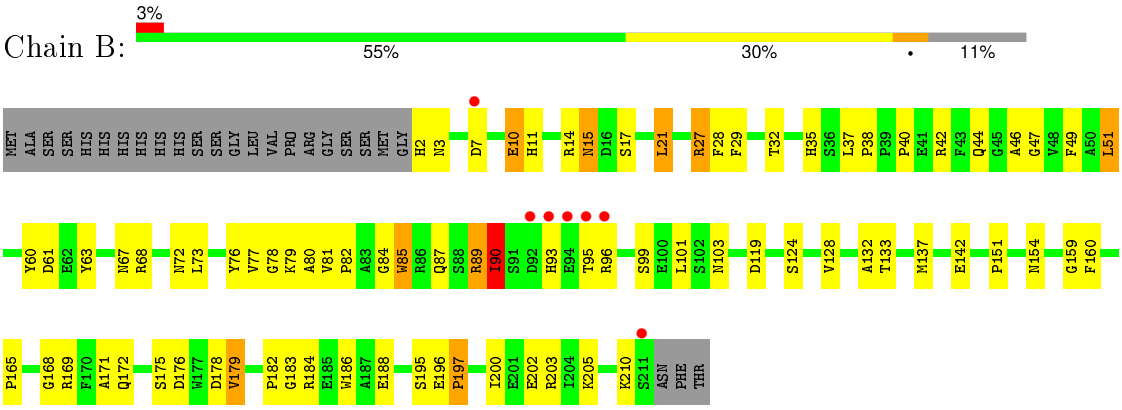
- Molecule 1: Eco29kIR

Chain A:



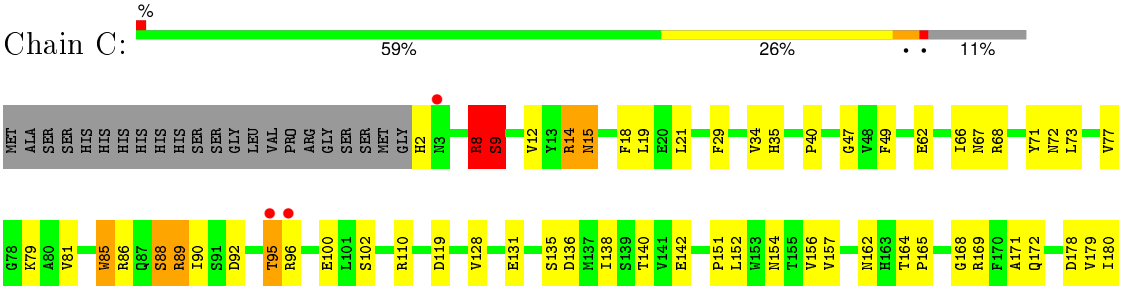
- Molecule 1: Eco29kIR

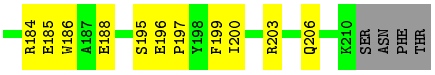
Chain B:



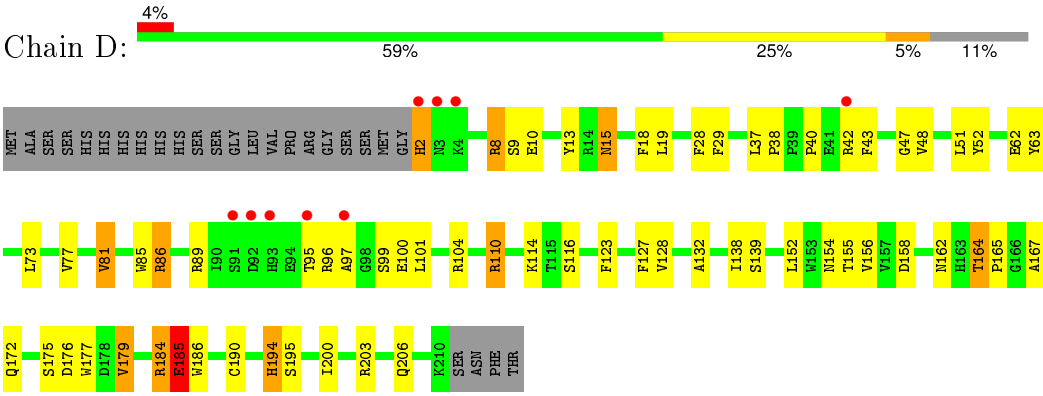
- Molecule 1: Eco29kIR

Chain C:

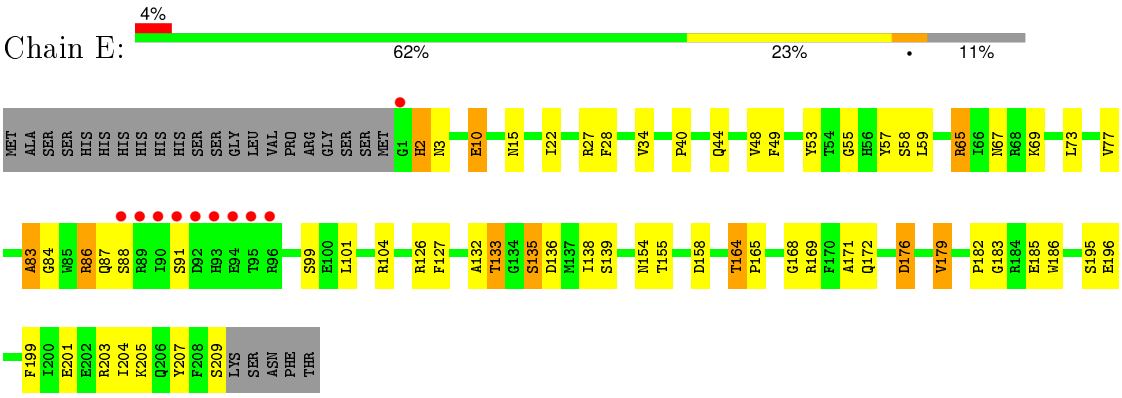




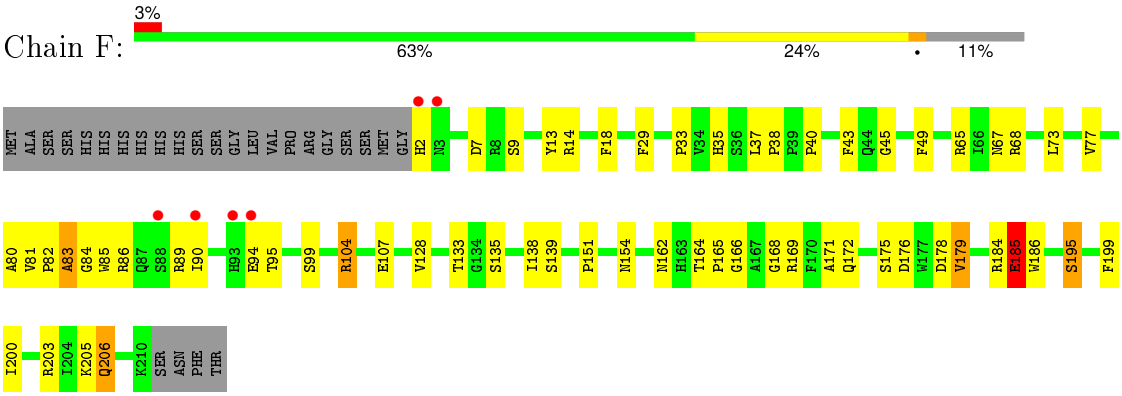
- Molecule 1: Eco29kIR



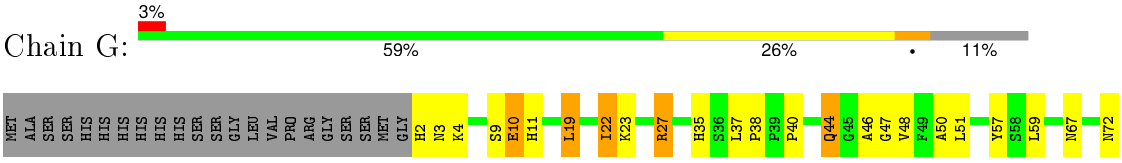
- Molecule 1: Eco29kIR

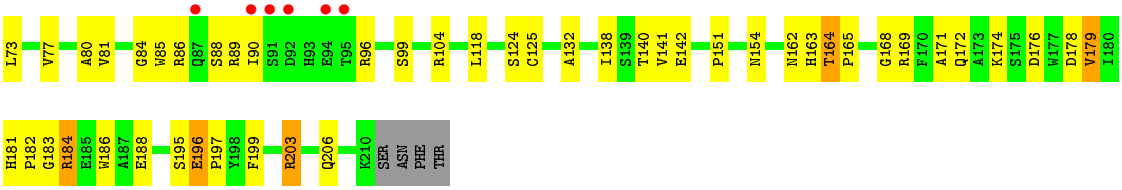


- Molecule 1: Eco29kIR

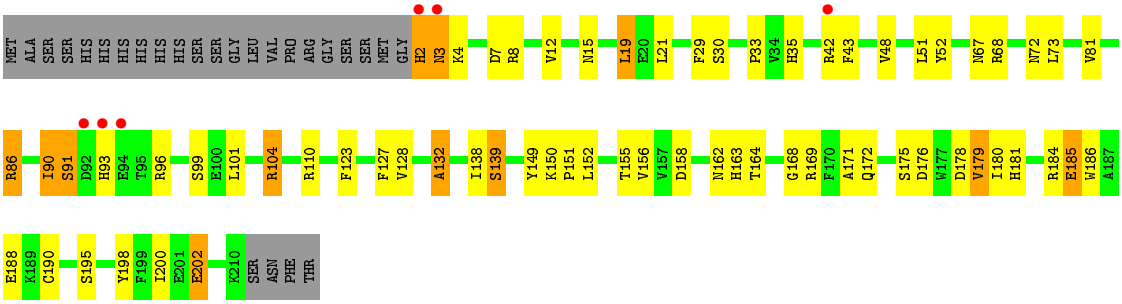


- Molecule 1: Eco29kIR

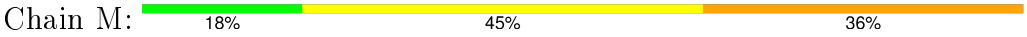




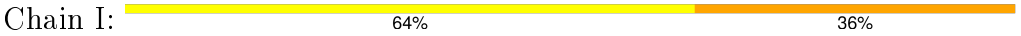
• Molecule 1: Eco29kIR



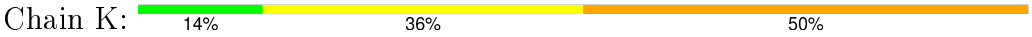
• Molecule 2: DNA (5'-D(P*CP*GP*GP*GP*AP*GP*GP*CP*CP*CP*GP*CP*GP*GP*GP*CP*CP*GP*CP*CP*GP*C)-3')



• Molecule 2: DNA (5'-D(P*CP*GP*GP*GP*AP*GP*GP*CP*CP*CP*GP*CP*GP*GP*GP*CP*CP*GP*CP*CP*GP*C)-3')



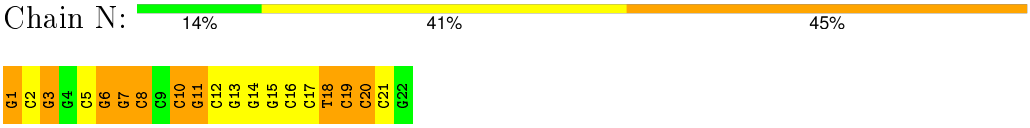
• Molecule 2: DNA (5'-D(P*CP*GP*GP*GP*AP*GP*GP*CP*CP*CP*GP*CP*GP*GP*GP*CP*CP*GP*CP*CP*GP*C)-3')



• Molecule 2: DNA (5'-D(P*CP*GP*GP*GP*AP*GP*GP*CP*CP*CP*GP*CP*GP*GP*GP*CP*CP*GP*CP*CP*GP*C)-3')

• Molecule 3: DNA (5'-D(P*GP*CP*GP*GP*CP*GP*GP*CP*CP*CP*GP*CP*GP*GP*GP*CP*CP*TP*CP*CP*CP*G)-3')

Chain N:

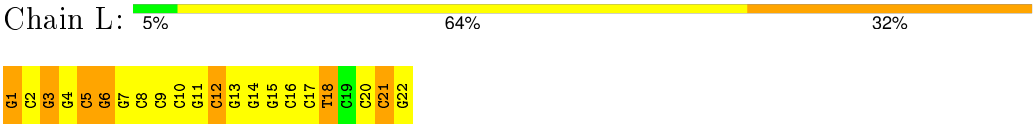


• Molecule 3: DNA (5'-D(P*GP*CP*GP*GP*CP*GP*GP*CP*CP*CP*GP*CP*GP*GP*GP*CP*CP*TP*CP*CP*CP*G)-3')

Chain J:

• Molecule 3: DNA (5'-D(P*GP*CP*GP*GP*CP*GP*GP*CP*CP*CP*GP*CP*GP*GP*GP*CP*CP*TP*CP*CP*CP*G)-3')

Chain L:



• Molecule 3: DNA (5'-D(P*GP*CP*GP*GP*CP*GP*GP*CP*CP*CP*GP*CP*GP*GP*GP*CP*CP*TP*CP*CP*CP*G)-3')

Chain P:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	99.87Å 101.35Å 142.35Å 90.00° 109.59° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 49.93 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.5 (50.00-2.80) 99.5 (49.93-2.80)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.58 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.5.0088	Depositor
R, R_{free}	0.206 , 0.279 0.213 , 0.286	Depositor DCC
R_{free} test set	3346 reflections (5.35%)	DCC
Wilson B-factor (Å^2)	38.7	Xtriage
Anisotropy	0.832	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 49.4	EDS
Estimated twinning fraction	0.000 for h,-k,-h-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Outliers	11 of 65945 reflections (0.017%)	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	17279	wwPDB-VP
Average B, all atoms (Å^2)	46.0	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 56.37 % 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 2.7645e-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, 0.375 respectively for untwinned datasets, and 0.333, 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: PO4

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	0.80	0/1720	0.82	2/2331 (0.1%)
1	B	0.83	0/1733	0.80	0/2347
1	C	0.83	0/1682	0.88	3/2280 (0.1%)
1	D	0.77	0/1698	0.82	0/2302
1	E	0.80	0/1680	0.81	1/2281 (0.0%)
1	F	0.85	1/1674 (0.1%)	0.85	0/2267
1	G	0.76	0/1659	0.85	2/2251 (0.1%)
1	H	0.73	1/1691 (0.1%)	0.82	2/2296 (0.1%)
2	I	1.34	2/508 (0.4%)	2.25	28/782 (3.6%)
2	K	1.30	2/508 (0.4%)	2.19	32/782 (4.1%)
2	M	1.30	1/508 (0.2%)	2.13	27/782 (3.5%)
2	O	1.33	2/508 (0.4%)	2.37	32/782 (4.1%)
3	J	1.37	3/502 (0.6%)	2.50	40/772 (5.2%)
3	L	1.28	1/502 (0.2%)	2.21	28/772 (3.6%)
3	N	1.43	2/502 (0.4%)	2.35	30/772 (3.9%)
3	P	1.31	2/502 (0.4%)	2.25	28/772 (3.6%)
All	All	0.95	17/17577 (0.1%)	1.36	255/24571 (1.0%)

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	H	0	2

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	I	9	DC	C3'-O3'	-9.98	1.30	1.44
2	M	12	DC	C3'-O3'	-7.86	1.33	1.44

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	N	11	DG	N7-C5	-7.29	1.34	1.39
3	P	9	DC	C3'-O3'	-6.57	1.35	1.44
2	K	9	DC	C3'-O3'	-6.54	1.35	1.44

The worst 5 of 255 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	J	1	DG	O4'-C1'-N9	21.01	122.70	108.00
3	N	13	DG	O4'-C1'-N9	18.28	120.79	108.00
3	L	13	DG	O4'-C1'-N9	17.96	120.57	108.00
3	P	1	DG	O4'-C1'-N9	17.57	120.30	108.00
2	K	13	DG	O4'-C1'-N9	15.35	118.74	108.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	H	162	ASN	Peptide
1	H	4	LYS	Peptide

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	A	1671	0	1573	62	0
1	B	1683	0	1585	78	1
1	C	1635	0	1522	76	0
1	D	1649	0	1540	73	0
1	E	1631	0	1513	54	0
1	F	1627	0	1516	62	1
1	G	1612	0	1480	67	0
1	H	1642	0	1526	57	0
2	I	453	0	243	27	0
2	K	453	0	243	25	0
2	M	453	0	243	13	0
2	O	453	0	243	15	0
3	J	449	0	244	18	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	L	449	0	244	17	0
3	N	449	0	244	18	0
3	P	449	0	244	18	0
4	A	5	0	0	0	0
4	B	10	0	0	0	0
4	D	10	0	0	2	0
4	E	5	0	0	0	0
4	F	10	0	0	1	0
4	G	5	0	0	0	0
4	H	10	0	0	3	0
5	A	51	0	0	1	0
5	B	56	0	0	2	0
5	C	60	0	0	8	0
5	D	35	0	0	3	0
5	E	46	0	0	3	0
5	F	59	0	0	9	0
5	G	37	0	0	1	0
5	H	29	0	0	2	0
5	I	5	0	0	0	0
5	J	12	0	0	0	0
5	K	11	0	0	1	0
5	L	19	0	0	5	0
5	M	11	0	0	1	0
5	N	10	0	0	2	0
5	O	7	0	0	2	0
5	P	18	0	0	2	0
All	All	17279	0	14203	559	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 18.

The worst 5 of 559 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)
1:C:171:ALA:CB	1:G:168:GLY:HA2	1.76	1.15
2:I:22:DC:H5	2:K:1:DC:O5'	1.29	1.14
1:C:171:ALA:HB3	1:G:168:GLY:HA2	1.17	1.10
3:J:1:DG:H2"	3:J:2:DC:H5	1.12	1.09
1:G:196:GLU:HG3	1:G:197:PRO:HD2	1.31	1.07

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:61:ASP:OD1	1:F:133:THR:OG1[2_656]	2.16	0.04

5.3 Torsion angles

5.3.1 Protein backbone

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	208/235 (88%)	183 (88%)	21 (10%)	4 (2%)	10	32
1	B	208/235 (88%)	189 (91%)	16 (8%)	3 (1%)	14	42
1	C	207/235 (88%)	182 (88%)	20 (10%)	5 (2%)	7	25
1	D	207/235 (88%)	183 (88%)	20 (10%)	4 (2%)	10	32
1	E	207/235 (88%)	183 (88%)	18 (9%)	6 (3%)	6	19
1	F	207/235 (88%)	183 (88%)	18 (9%)	6 (3%)	6	19
1	G	207/235 (88%)	184 (89%)	16 (8%)	7 (3%)	5	16
1	H	207/235 (88%)	176 (85%)	24 (12%)	7 (3%)	5	16
All	All	1658/1880 (88%)	1463 (88%)	153 (9%)	42 (2%)	7	24

5 of 42 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	132	ALA
1	B	90	ILE
1	B	132	ALA
1	C	90	ILE
1	D	10	GLU

5.3.2 Protein sidechains

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	171/201 (85%)	160 (94%)	11 (6%)	22 52
1	B	174/201 (87%)	157 (90%)	17 (10%)	10 28
1	C	165/201 (82%)	147 (89%)	18 (11%)	8 23
1	D	167/201 (83%)	147 (88%)	20 (12%)	6 19
1	E	163/201 (81%)	150 (92%)	13 (8%)	15 40
1	F	164/201 (82%)	153 (93%)	11 (7%)	20 50
1	G	160/201 (80%)	146 (91%)	14 (9%)	12 35
1	H	166/201 (83%)	155 (93%)	11 (7%)	21 51
All	All	1330/1608 (83%)	1215 (91%)	115 (9%)	13 36

5 of 115 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	65	ARG
1	D	194	HIS
1	H	21	LEU
1	D	81	VAL
1	D	139	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 63 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	162	ASN
1	E	67	ASN
1	H	67	ASN
1	D	172	GLN
1	E	2	HIS

5.3.3 RNA

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

11 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
4	PO4	A	215	-	4,4,4	0.55	0	6,6,6	0.27	0
4	PO4	B	215	-	4,4,4	0.41	0	6,6,6	0.27	0
4	PO4	B	216	-	4,4,4	0.61	0	6,6,6	0.29	0
4	PO4	D	215	-	4,4,4	0.48	0	6,6,6	0.28	0
4	PO4	D	216	-	4,4,4	0.70	0	6,6,6	0.25	0
4	PO4	E	215	-	4,4,4	0.40	0	6,6,6	0.28	0
4	PO4	F	215	-	4,4,4	0.47	0	6,6,6	0.28	0
4	PO4	F	216	-	4,4,4	0.41	0	6,6,6	0.28	0
4	PO4	G	215	-	4,4,4	0.54	0	6,6,6	0.27	0
4	PO4	H	215	-	4,4,4	0.49	0	6,6,6	0.29	0
4	PO4	H	216	-	4,4,4	0.65	0	6,6,6	0.28	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
4	PO4	A	215	-	-	0/0/0/0	0/0/0/0
4	PO4	B	215	-	-	0/0/0/0	0/0/0/0
4	PO4	B	216	-	-	0/0/0/0	0/0/0/0
4	PO4	D	215	-	-	0/0/0/0	0/0/0/0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	PO4	D	216	-	-	0/0/0/0	0/0/0/0
4	PO4	E	215	-	-	0/0/0/0	0/0/0/0
4	PO4	F	215	-	-	0/0/0/0	0/0/0/0
4	PO4	F	216	-	-	0/0/0/0	0/0/0/0
4	PO4	G	215	-	-	0/0/0/0	0/0/0/0
4	PO4	H	215	-	-	0/0/0/0	0/0/0/0
4	PO4	H	216	-	-	0/0/0/0	0/0/0/0

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.

4 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	215	PO4	1	0
4	D	216	PO4	1	0
4	F	216	PO4	1	0
4	H	216	PO4	3	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	210/235 (89%)	-0.12	10 (4%) 34 23	21, 36, 79, 99	0
1	B	210/235 (89%)	-0.09	7 (3%) 50 38	18, 34, 75, 106	0
1	C	209/235 (88%)	-0.03	3 (1%) 78 69	20, 33, 69, 102	0
1	D	209/235 (88%)	0.03	9 (4%) 39 27	22, 45, 77, 94	0
1	E	209/235 (88%)	0.06	10 (4%) 34 23	22, 38, 76, 123	0
1	F	209/235 (88%)	-0.03	6 (2%) 55 43	20, 32, 77, 95	0
1	G	209/235 (88%)	-0.00	6 (2%) 55 43	21, 38, 73, 103	0
1	H	209/235 (88%)	0.02	6 (2%) 55 43	29, 48, 77, 97	0
2	I	22/22 (100%)	-0.03	0 100 100	22, 69, 113, 127	0
2	K	22/22 (100%)	0.03	0 100 100	26, 58, 109, 125	0
2	M	22/22 (100%)	0.17	0 100 100	30, 64, 111, 134	0
2	O	22/22 (100%)	0.10	2 (9%) 11 6	23, 69, 135, 146	0
3	J	22/22 (100%)	0.11	0 100 100	21, 64, 121, 139	0
3	L	22/22 (100%)	0.06	0 100 100	21, 60, 98, 122	0
3	N	22/22 (100%)	0.05	0 100 100	23, 63, 114, 134	0
3	P	22/22 (100%)	0.13	1 (4%) 37 26	25, 64, 123, 140	0
All	All	1850/2056 (89%)	-0.01	60 (3%) 51 39	18, 38, 91, 146	0

The worst 5 of 60 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	93	HIS	9.4
1	C	95	THR	7.5
1	E	93	HIS	7.1
1	D	3	ASN	6.8
1	E	88	SER	6.8

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

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

6.3 Carbohydrates (i)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
4	PO4	F	216	5/5	0.88	0.38	10.35	104,104,105,105	0
4	PO4	H	216	5/5	0.89	0.16	-0.64	66,66,67,68	0
4	PO4	D	215	5/5	0.86	0.21	-	100,101,101,101	0
4	PO4	B	216	5/5	0.94	0.15	-	73,73,74,75	0
4	PO4	D	216	5/5	0.92	0.16	-	75,75,76,77	0
4	PO4	E	215	5/5	0.87	0.31	-	100,100,101,101	0
4	PO4	A	215	5/5	0.88	0.22	-	75,76,77,78	0
4	PO4	G	215	5/5	0.87	0.26	-	101,101,101,102	0
4	PO4	B	215	5/5	0.72	0.39	-	109,110,110,110	0
4	PO4	H	215	5/5	0.93	0.21	-	89,89,90,91	0
4	PO4	F	215	5/5	0.84	0.29	-	115,115,115,115	0

6.5 Other polymers (i)

There are no such residues in this entry.