wwPDB X-ray Structure Validation Summary Report (i)

Oct 2, 2021 – 02:10 PM EDT

PDB ID	:	3KLE
Title	:	Crystal structure of AZT-resistant HIV-1 Reverse Transcriptase crosslinked to a DSDNA with a bound excision product, AZTPPPPA
Authors	:	Tu, X.; Das, K.; Sarafianos, S.G.; Arnold, E.
Deposited on	:	2009-11-07
Resolution	:	3.20 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 (i)) were used in the production of this report:

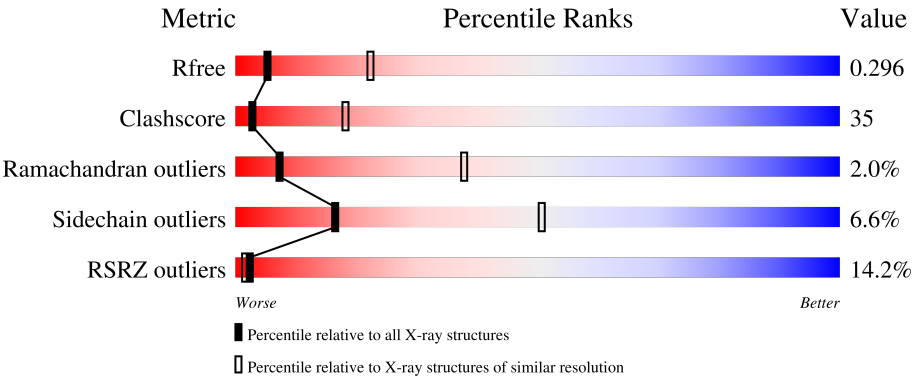
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 3.20 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}	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	A	562	.% 41% 53% 6%
1	E	562	.% 40% 53% 6%
1	I	562	16% 39% 55% 6%
1	M	562	33% 41% 53% 6%
2	B	437	3% 46% 46% 5% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	437	3% 41% 48% 6%
2	J	437	20% 43% 48% 5%
2	N	437	38% 43% 48% 5%
3	C	27	33% 56% 7%
3	G	27	26% 63% 7%
3	K	27	30% 59% 7%
3	O	27	15% 33% 56% 7%
4	D	21	43% 43% 14%
4	H	21	52% 33% 14%
4	L	21	5% 38% 52% 10%
4	P	21	10% 38% 52% 10%
5	Q	2	50% 50%

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 36120 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 Reverse transcriptase/ribonuclease H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	560	Total	C	N	O	S	0	0	0
			4560	2953	762	838	7			
1	E	560	Total	C	N	O	S	0	0	0
			4560	2953	762	838	7			
1	I	560	Total	C	N	O	S	0	0	0
			4560	2953	762	838	7			
1	M	560	Total	C	N	O	S	0	0	0
			4560	2953	762	838	7			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	expression tag	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	41	LEU	MET	engineered mutation	UNP P03366
A	67	ASN	ASP	engineered mutation	UNP P03366
A	70	ARG	LYS	engineered mutation	UNP P03366
A	215	TYR	THR	engineered mutation	UNP P03366
A	219	GLN	LYS	engineered mutation	UNP P03366
A	258	CYS	GLN	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366
A	559	VAL	ILE	see remark 999	UNP P03366
E	-1	MET	-	expression tag	UNP P03366
E	0	VAL	-	expression tag	UNP P03366
E	41	LEU	MET	engineered mutation	UNP P03366
E	67	ASN	ASP	engineered mutation	UNP P03366
E	70	ARG	LYS	engineered mutation	UNP P03366
E	215	TYR	THR	engineered mutation	UNP P03366
E	219	GLN	LYS	engineered mutation	UNP P03366
E	258	CYS	GLN	engineered mutation	UNP P03366
E	280	SER	CYS	engineered mutation	UNP P03366
E	559	VAL	ILE	see remark 999	UNP P03366
I	-1	MET	-	expression tag	UNP P03366

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
I	0	VAL	-	expression tag	UNP P03366
I	41	LEU	MET	engineered mutation	UNP P03366
I	67	ASN	ASP	engineered mutation	UNP P03366
I	70	ARG	LYS	engineered mutation	UNP P03366
I	215	TYR	THR	engineered mutation	UNP P03366
I	219	GLN	LYS	engineered mutation	UNP P03366
I	258	CYS	GLN	engineered mutation	UNP P03366
I	280	SER	CYS	engineered mutation	UNP P03366
I	559	VAL	ILE	see remark 999	UNP P03366
M	-1	MET	-	expression tag	UNP P03366
M	0	VAL	-	expression tag	UNP P03366
M	41	LEU	MET	engineered mutation	UNP P03366
M	67	ASN	ASP	engineered mutation	UNP P03366
M	70	ARG	LYS	engineered mutation	UNP P03366
M	215	TYR	THR	engineered mutation	UNP P03366
M	219	GLN	LYS	engineered mutation	UNP P03366
M	258	CYS	GLN	engineered mutation	UNP P03366
M	280	SER	CYS	engineered mutation	UNP P03366
M	559	VAL	ILE	see remark 999	UNP P03366

- Molecule 2 is a protein called p51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	425	Total	C	N	O	S	0	0	0
			3462	2248	576	632	6			
2	F	419	Total	C	N	O	S	0	0	0
			3434	2232	570	626	6			
2	J	419	Total	C	N	O	S	0	0	0
			3434	2232	570	626	6			
2	N	419	Total	C	N	O	S	0	0	0
			3434	2232	570	626	6			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	UNP P03366
B	429	GLY	-	expression tag	UNP P03366
B	430	GLY	-	expression tag	UNP P03366
B	431	HIS	-	expression tag	UNP P03366
B	432	HIS	-	expression tag	UNP P03366
B	433	HIS	-	expression tag	UNP P03366
B	434	HIS	-	expression tag	UNP P03366

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	435	HIS	-	expression tag	UNP P03366
B	436	HIS	-	expression tag	UNP P03366
B	437	HIS	-	expression tag	UNP P03366
F	280	SER	CYS	engineered mutation	UNP P03366
F	429	GLY	-	expression tag	UNP P03366
F	430	GLY	-	expression tag	UNP P03366
F	431	HIS	-	expression tag	UNP P03366
F	432	HIS	-	expression tag	UNP P03366
F	433	HIS	-	expression tag	UNP P03366
F	434	HIS	-	expression tag	UNP P03366
F	435	HIS	-	expression tag	UNP P03366
F	436	HIS	-	expression tag	UNP P03366
F	437	HIS	-	expression tag	UNP P03366
J	280	SER	CYS	engineered mutation	UNP P03366
J	429	GLY	-	expression tag	UNP P03366
J	430	GLY	-	expression tag	UNP P03366
J	431	HIS	-	expression tag	UNP P03366
J	432	HIS	-	expression tag	UNP P03366
J	433	HIS	-	expression tag	UNP P03366
J	434	HIS	-	expression tag	UNP P03366
J	435	HIS	-	expression tag	UNP P03366
J	436	HIS	-	expression tag	UNP P03366
J	437	HIS	-	expression tag	UNP P03366
N	280	SER	CYS	engineered mutation	UNP P03366
N	429	GLY	-	expression tag	UNP P03366
N	430	GLY	-	expression tag	UNP P03366
N	431	HIS	-	expression tag	UNP P03366
N	432	HIS	-	expression tag	UNP P03366
N	433	HIS	-	expression tag	UNP P03366
N	434	HIS	-	expression tag	UNP P03366
N	435	HIS	-	expression tag	UNP P03366
N	436	HIS	-	expression tag	UNP P03366
N	437	HIS	-	expression tag	UNP P03366

- Molecule 3 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	25	Total	C	N	O	P	0	0	0
			515	243	102	146	24			
3	G	25	Total	C	N	O	P	0	0	0
			515	243	102	146	24			
3	K	25	Total	C	N	O	P	0	0	0
			515	243	102	146	24			

Continued on next page...

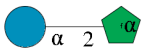
Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	O	25	Total	C	N	O	P	0	0	0
			515	243	102	146	24			

- Molecule 4 is a DNA chain called DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(2DA))-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	D	21	Total	C	N	O	P	S	0	0	0
			427	205	77	124	20	1			
4	H	21	Total	C	N	O	P	S	0	0	0
			427	205	77	124	20	1			
4	L	21	Total	C	N	O	P	S	0	0	0
			427	205	77	124	20	1			
4	P	21	Total	C	N	O	P	S	0	0	0
			427	205	77	124	20	1			

- Molecule 5 is an oligosaccharide called alpha-D-fructofuranose-(2-1)-alpha-D-glucopyranose.

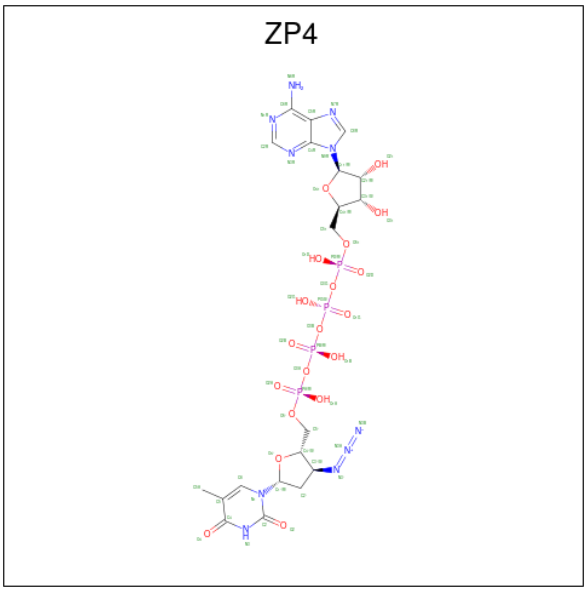


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
5	Q	2	Total	C	O	0	0	0
			23	12	11			

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

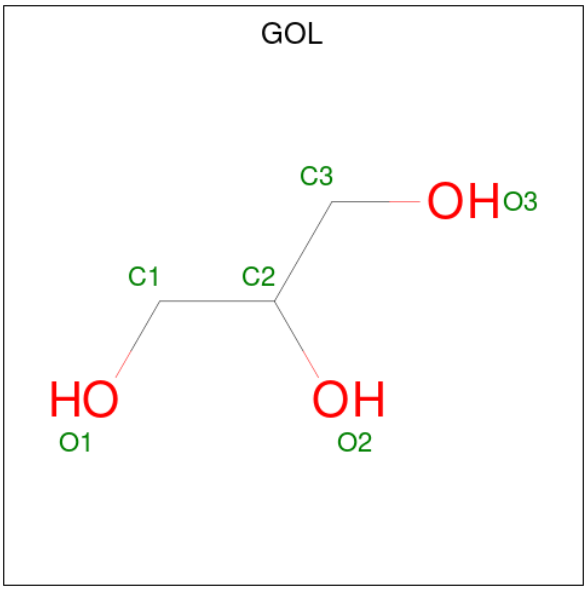
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	2	Total	Mg	0	0
			2	2		
6	E	2	Total	Mg	0	0
			2	2		
6	I	2	Total	Mg	0	0
			2	2		
6	M	2	Total	Mg	0	0
			2	2		

- Molecule 7 is [[[[(2R,3S,4R,5R)-5-(6-aminopurin-9-yl)-3,4-dihydroxy-oxolan-2-yl]methoxy-hydroxy-phosphoryl]oxy-hydroxy-phosphoryl]oxy-hydroxy-phosphoryl] [(2S,3S,5R)-3-azido-5-(5-methyl-2,4-dioxo-pyrimidin-1-yl)oxolan-2-yl]methyl hydrogen phosphate (three-letter code: ZP4) (formula: $C_{20}H_{28}N_{10}O_{19}P_4$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	A	1	Total	C	N	O	P	0	0
			53	20	10	19	4		
7	E	1	Total	C	N	O	P	0	0
			53	20	10	19	4		
7	I	1	Total	C	N	O	P	0	0
			53	20	10	19	4		
7	M	1	Total	C	N	O	P	0	0
			53	20	10	19	4		

- Molecule 8 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	B	1	Total C O 6 3 3	0	0
8	F	1	Total C O 6 3 3	0	0

- Molecule 9 is water.

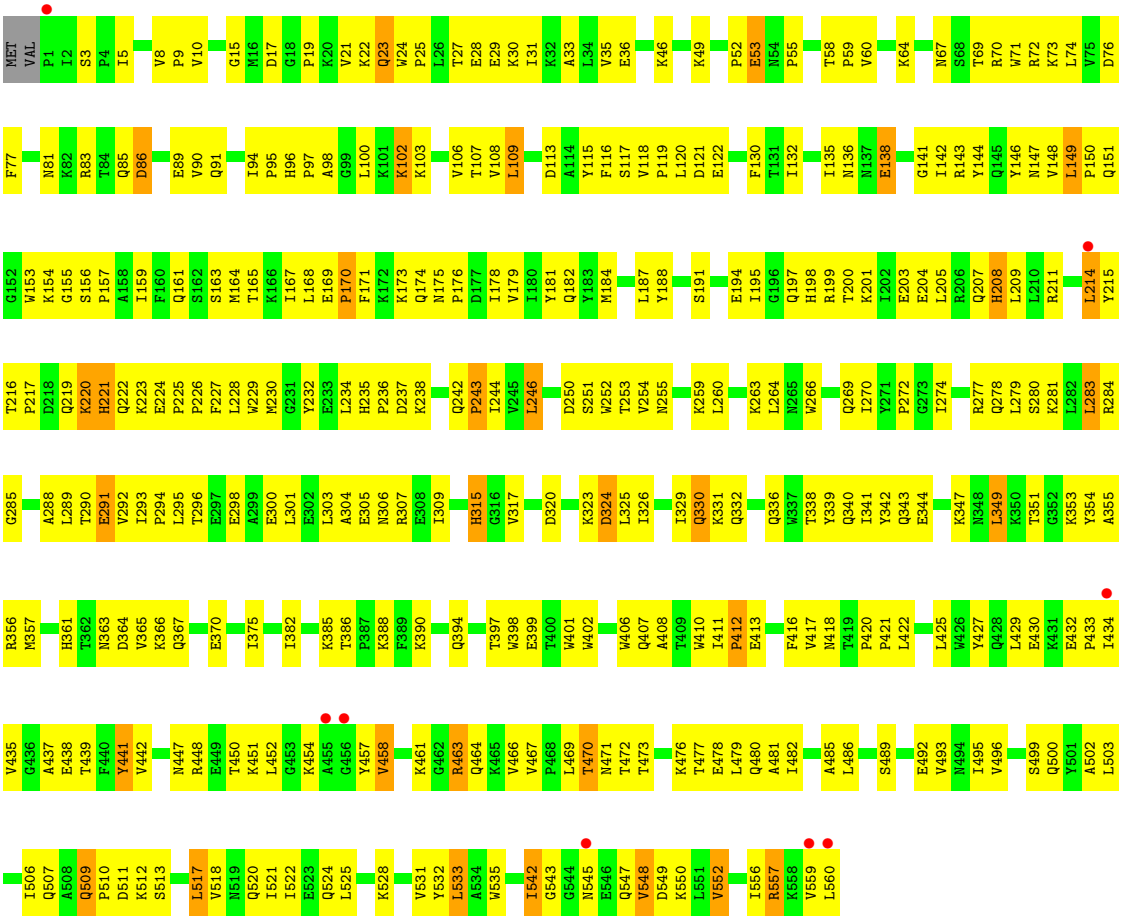
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	A	18	Total O 18 18	0	0
9	B	10	Total O 10 10	0	0
9	E	12	Total O 12 12	0	0
9	F	10	Total O 10 10	0	0
9	I	5	Total O 5 5	0	0
9	J	14	Total O 14 14	0	0
9	M	3	Total O 3 3	0	0
9	N	8	Total O 8 8	0	0
9	C	2	Total O 2 2	0	0
9	D	2	Total O 2 2	0	0
9	G	1	Total O 1 1	0	0
9	H	1	Total O 1 1	0	0
9	K	1	Total O 1 1	0	0
9	L	1	Total O 1 1	0	0
9	O	4	Total O 4 4	0	0
9	P	1	Total O 1 1	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.

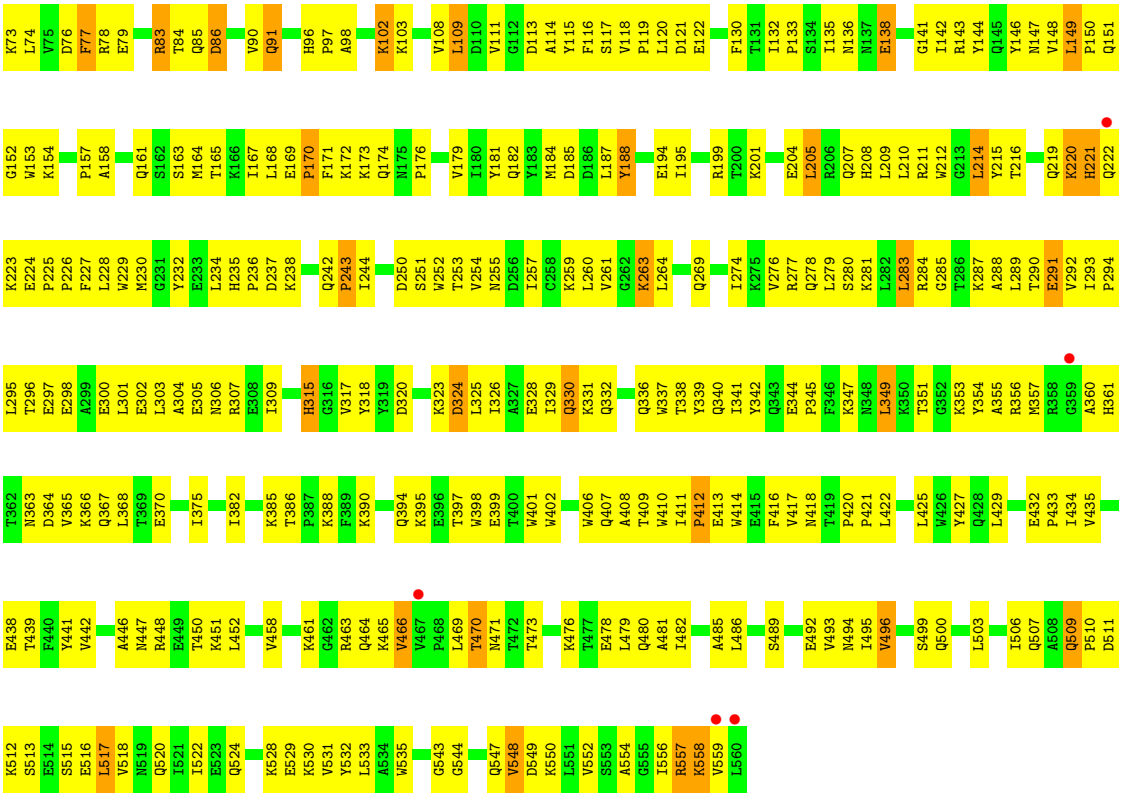
• Molecule 1: Reverse transcriptase/ribonuclease H

Chain A:



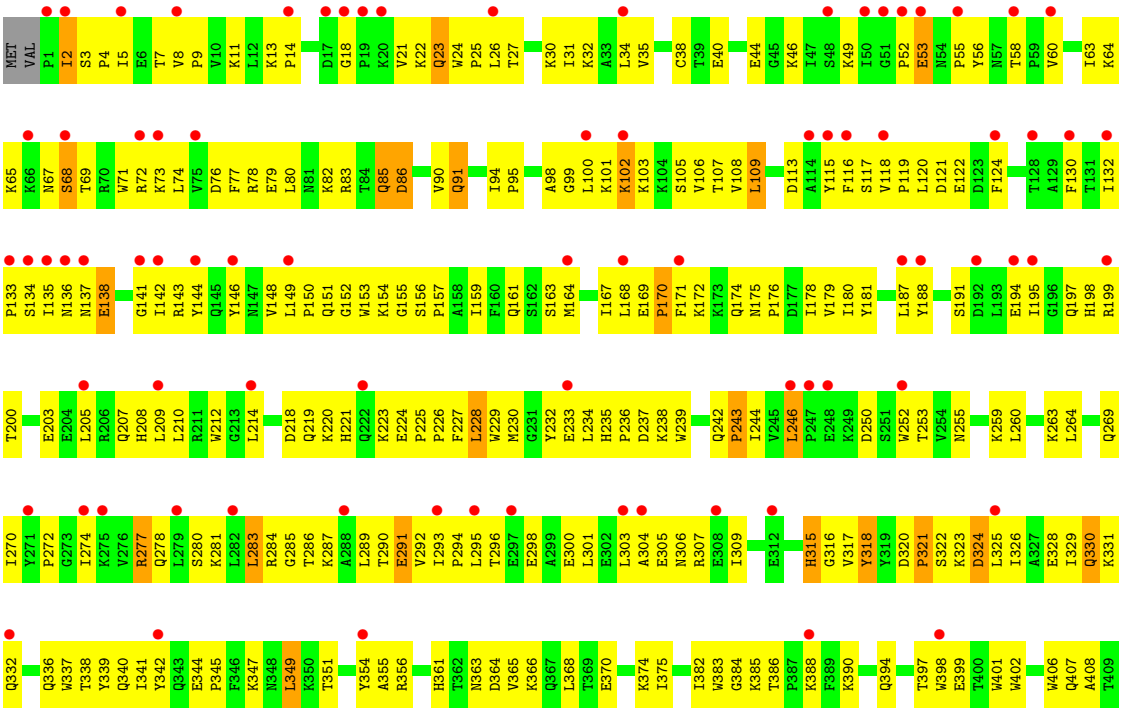
• Molecule 1: Reverse transcriptase/ribonuclease H

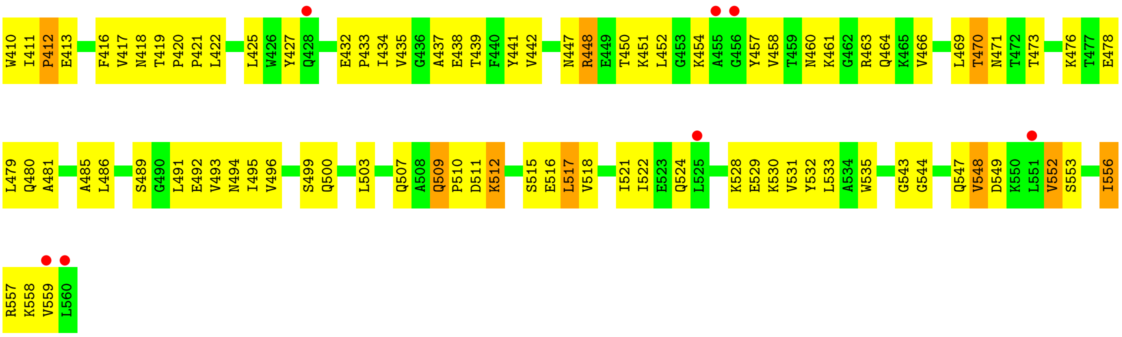
Chain E:



• Molecule 1: Reverse transcriptase/ribonuclease H

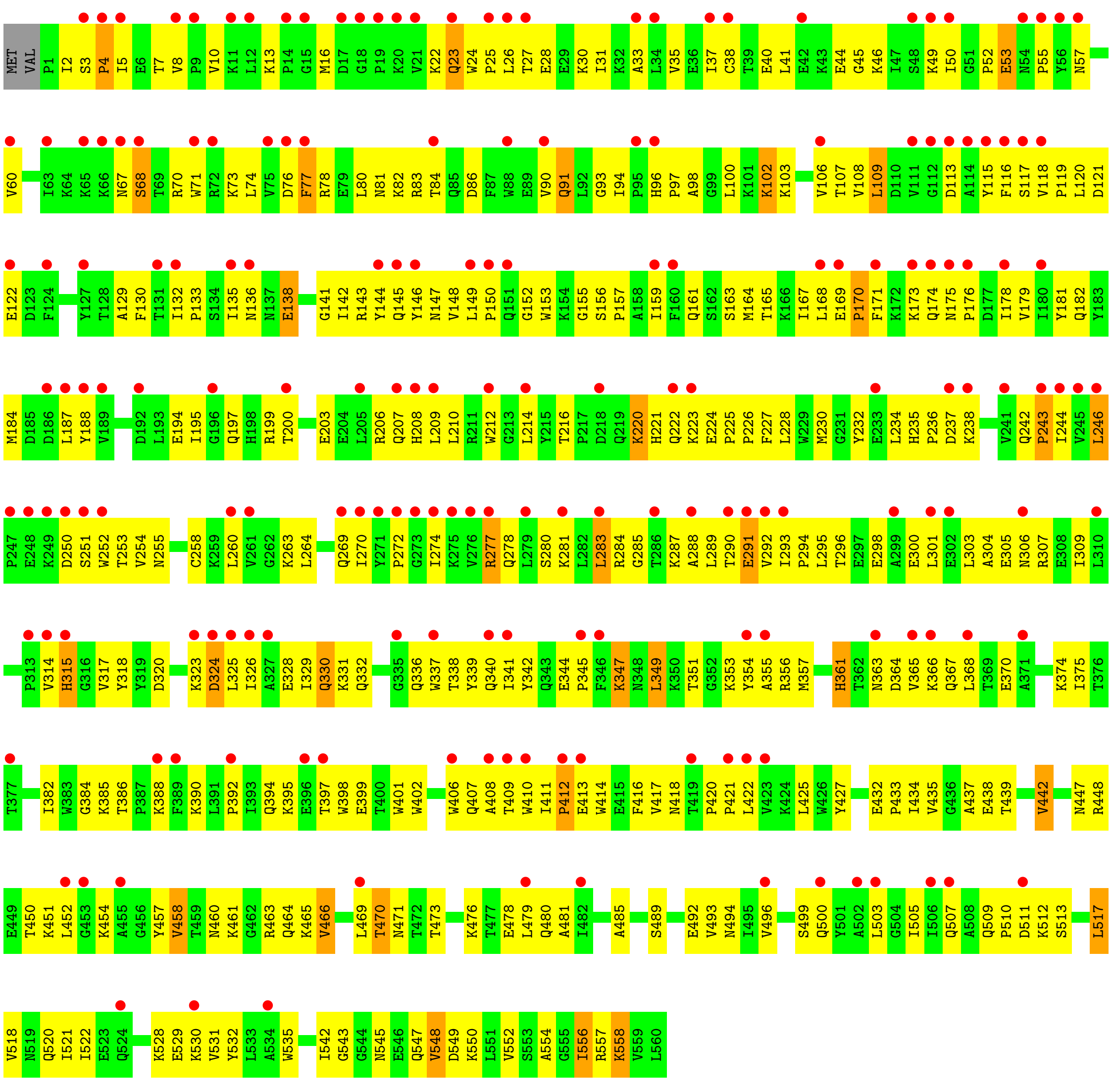
Chain I:





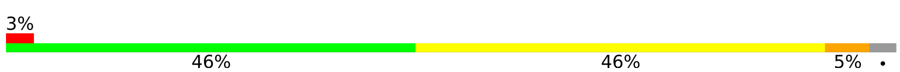
• Molecule 1: Reverse transcriptase/ribonuclease H

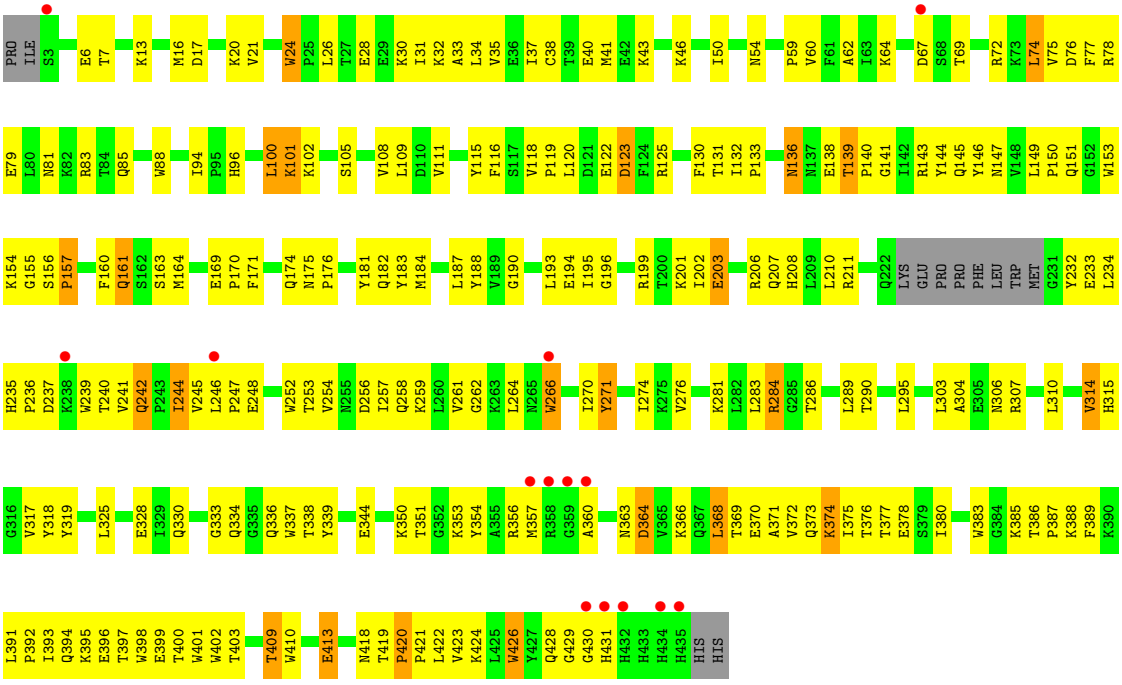
Chain M:



• Molecule 2: p51 RT

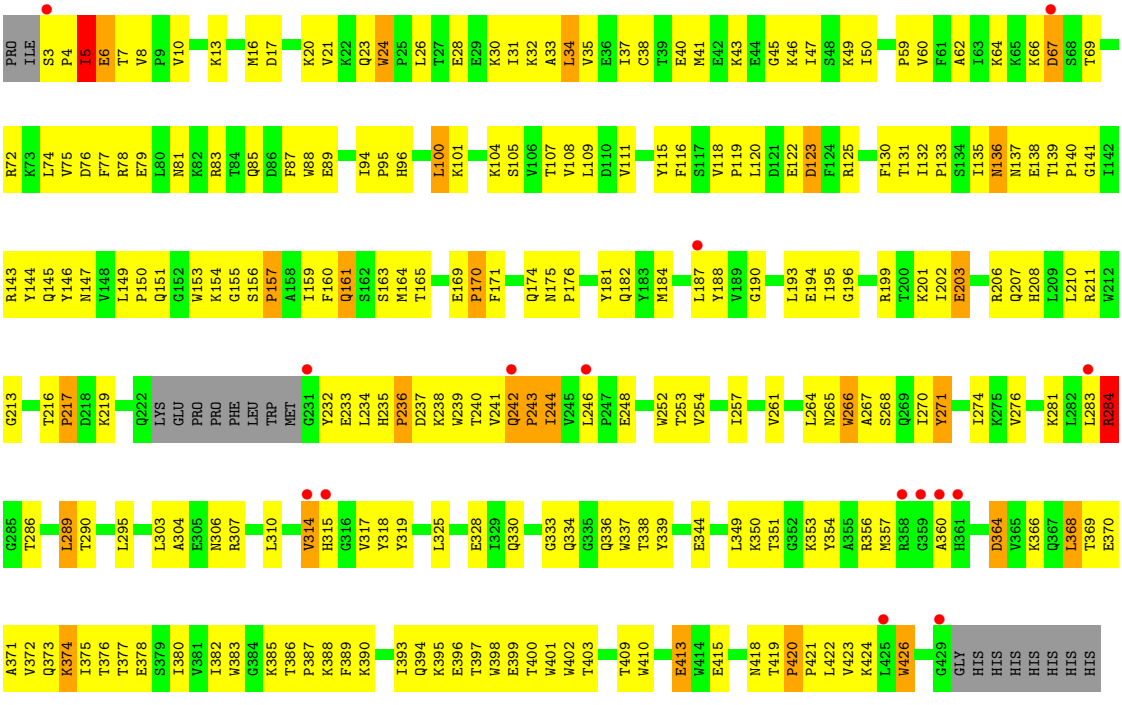
Chain B:





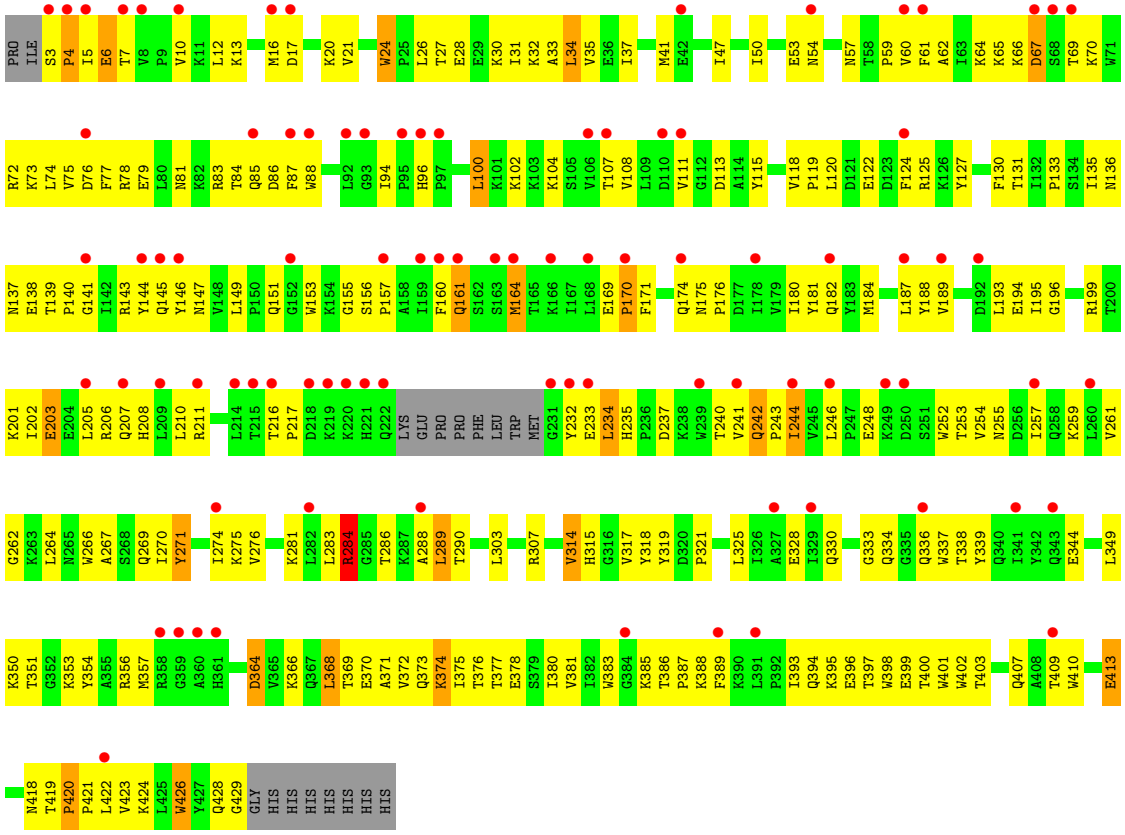
• Molecule 2: p51 RT

Chain F:



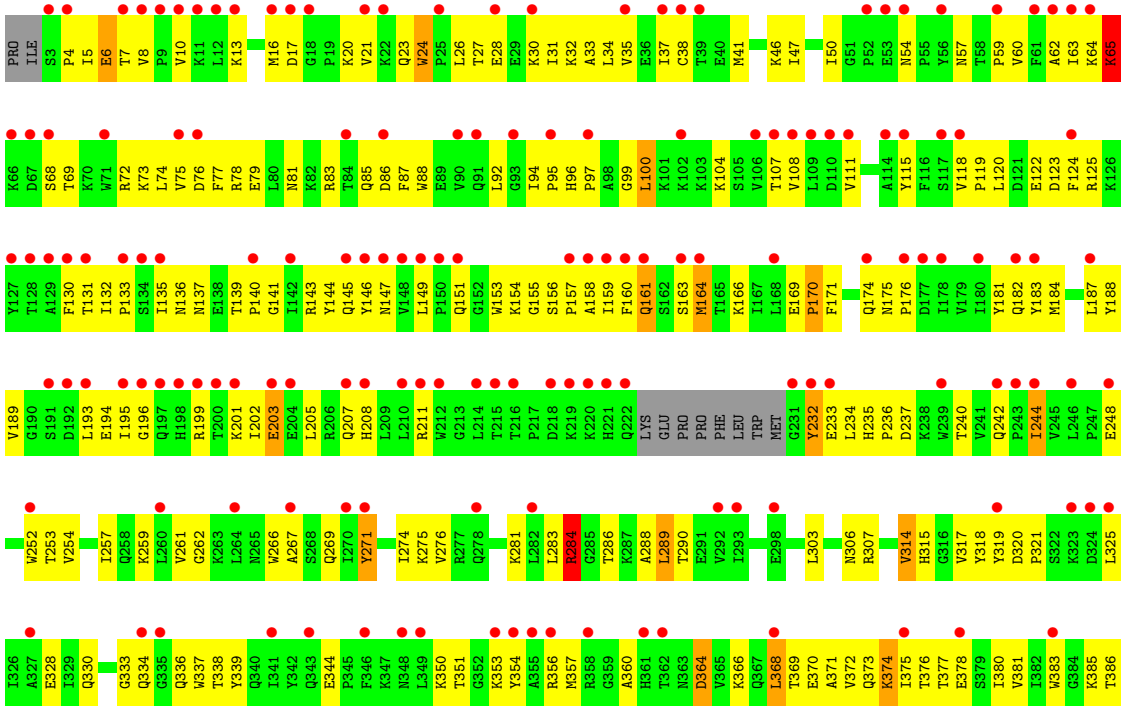
• Molecule 2: p51 RT

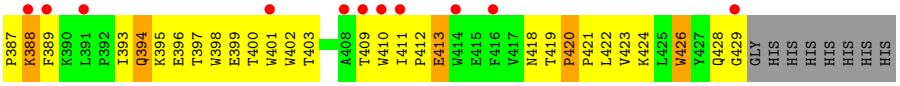
Chain J:



• Molecule 2: p51 RT

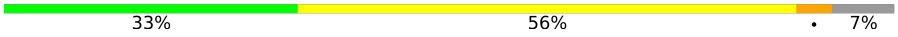
Chain N:

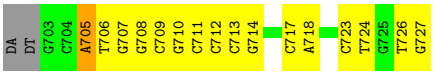




• Molecule 3: DNA (25-MER)

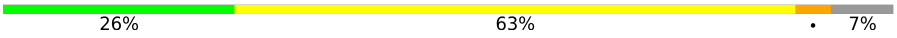
Chain C:

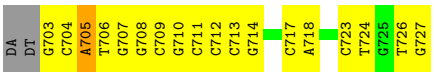




• Molecule 3: DNA (25-MER)

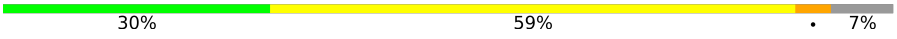
Chain G:

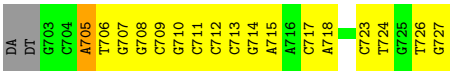




• Molecule 3: DNA (25-MER)

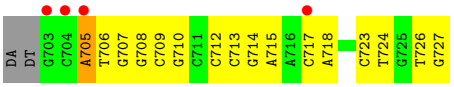
Chain K:





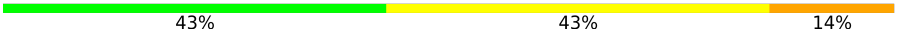
• Molecule 3: DNA (25-MER)

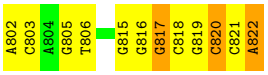
Chain O:



• Molecule 4: DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(2DA))-3')

Chain D:



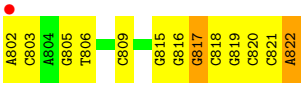


• Molecule 4: DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(2DA))-3')

Chain H:

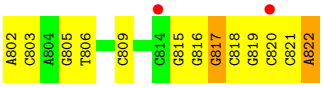
• Molecule 4: DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(2DA))-3')

Chain L:



• Molecule 4: DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(2DA))-3')

Chain P:



• Molecule 5: alpha-D-fructofuranose-(2-1)-alpha-D-glucopyranose

Chain Q:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	78.74Å 283.33Å 155.23Å 90.00° 89.73° 90.00°	Depositor
Resolution (Å)	24.83 – 3.20 24.83 – 3.18	Depositor EDS
% Data completeness (in resolution range)	90.4 (24.83-3.20) 88.8 (24.83-3.18)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.79 (at 3.17Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.280 , 0.308 0.270 , 0.296	Depositor DCC
R_{free} test set	1993 reflections (1.98%)	wwPDB-VP
Wilson B-factor (Å^2)	73.0	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 12.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.40$, $< L^2 > = 0.23$	Xtriage
Estimated twinning fraction	0.279 for h,-k,-l	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	36120	wwPDB-VP
Average B, all atoms (Å^2)	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.60% 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: ZP4, 2DA, Z9N, GOL, MG, MRG, GLC

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	0.56	0/4679	0.70	2/6357 (0.0%)
1	E	0.59	0/4679	0.71	2/6357 (0.0%)
1	I	0.42	0/4679	0.61	0/6357
1	M	0.42	0/4679	0.61	2/6357 (0.0%)
2	B	0.63	0/3559	0.71	0/4838
2	F	0.62	0/3531	0.71	0/4800
2	J	0.44	0/3531	0.64	0/4800
2	N	0.42	0/3531	0.62	0/4800
3	C	0.88	0/579	0.93	0/893
3	G	0.96	0/579	0.93	0/893
3	K	0.56	0/579	0.81	0/893
3	O	0.56	0/579	0.81	0/893
4	D	0.92	0/424	0.90	0/649
4	H	0.87	0/424	0.93	0/649
4	L	0.56	0/424	0.88	0/649
4	P	0.53	0/424	0.85	0/649
All	All	0.55	0/36880	0.69	6/50834 (0.0%)

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
3	C	0	1
3	G	0	1
3	K	0	1
3	O	0	1
4	D	0	1
4	H	0	1
All	All	0	6

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	M	442	VAL	CB-CA-C	-5.99	100.03	111.40
1	A	542	ILE	CB-CA-C	-5.82	99.95	111.60
1	M	68	SER	CB-CA-C	-5.54	99.57	110.10
1	A	463	ARG	CB-CA-C	-5.51	99.39	110.40
1	E	492	GLU	N-CA-C	-5.31	96.66	111.00

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	C	705	DA	Sidechain
4	D	820	DC	Sidechain
3	G	705	DA	Sidechain
4	H	820	DC	Sidechain
3	K	705	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	A	4560	0	4627	358	0
1	E	4560	0	4627	359	0
1	I	4560	0	4627	385	0
1	M	4560	0	4628	344	0
2	B	3462	0	3462	236	0
2	F	3434	0	3450	260	0
2	J	3434	0	3450	273	0
2	N	3434	0	3450	253	0
3	C	515	0	280	24	0
3	G	515	0	280	25	0
3	K	515	0	280	27	0
3	O	515	0	280	24	0
4	D	427	0	242	12	0
4	H	427	0	242	11	0
4	L	427	0	242	12	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	P	427	0	243	16	0
5	Q	23	0	10	4	0
6	A	2	0	0	0	0
6	E	2	0	0	0	0
6	I	2	0	0	0	0
6	M	2	0	0	0	0
7	A	53	0	24	10	0
7	E	53	0	24	9	0
7	I	53	0	24	4	0
7	M	53	0	24	7	0
8	B	6	0	8	0	0
8	F	6	0	8	0	0
9	A	18	0	0	0	0
9	B	10	0	0	0	0
9	C	2	0	0	0	0
9	D	2	0	0	0	0
9	E	12	0	0	0	0
9	F	10	0	0	1	0
9	G	1	0	0	0	0
9	H	1	0	0	0	0
9	I	5	0	0	1	0
9	J	14	0	0	5	0
9	K	1	0	0	0	0
9	L	1	0	0	0	0
9	M	3	0	0	0	0
9	N	8	0	0	3	0
9	O	4	0	0	0	0
9	P	1	0	0	1	0
All	All	36120	0	34532	2497	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 35.

The worst 5 of 2497 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:I:441:TYR:CD1	1:I:544:GLY:HA3	1.70	1.25
1:I:286:THR:O	1:I:287:LYS:HG2	1.44	1.16
2:F:125:ARG:HD3	2:F:147:ASN:HA	1.29	1.15
1:I:489:SER:HB2	1:I:493:VAL:HG11	1.29	1.14
2:B:79:GLU:HG3	2:B:83:ARG:HH12	0.98	1.14

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
1	A	558/562 (99%)	469 (84%)	81 (14%)	8 (1%)	11 46
1	E	558/562 (99%)	461 (83%)	85 (15%)	12 (2%)	6 35
1	I	558/562 (99%)	476 (85%)	69 (12%)	13 (2%)	6 34
1	M	558/562 (99%)	474 (85%)	70 (12%)	14 (2%)	5 32
2	B	421/437 (96%)	362 (86%)	53 (13%)	6 (1%)	11 46
2	F	415/437 (95%)	361 (87%)	42 (10%)	12 (3%)	4 28
2	J	415/437 (95%)	366 (88%)	43 (10%)	6 (1%)	11 46
2	N	415/437 (95%)	360 (87%)	47 (11%)	8 (2%)	8 39
All	All	3898/3996 (98%)	3329 (85%)	490 (13%)	79 (2%)	7 38

5 of 79 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	296	THR
2	B	314	VAL
1	E	77	PHE
2	F	314	VAL
1	A	448	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	
1	A	500/502 (100%)	465 (93%)	35 (7%)	15	48
1	E	500/502 (100%)	470 (94%)	30 (6%)	19	54
1	I	500/502 (100%)	471 (94%)	29 (6%)	20	55
1	M	500/502 (100%)	475 (95%)	25 (5%)	24	60
2	B	375/397 (94%)	348 (93%)	27 (7%)	14	47
2	F	375/397 (94%)	347 (92%)	28 (8%)	13	45
2	J	375/397 (94%)	348 (93%)	27 (7%)	14	47
2	N	375/397 (94%)	345 (92%)	30 (8%)	12	42
All	All	3500/3596 (97%)	3269 (93%)	231 (7%)	16	51

5 of 231 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	368	LEU
2	N	286	THR
1	I	509	GLN
2	N	275	LYS
1	M	556	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 121 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	278	GLN
2	N	81	ASN
1	I	340	GLN
1	M	547	GLN
2	N	315	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

8 non-standard protein/DNA/RNA residues 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	MRG	H	817	1,3,4	22,28,29	1.83	2 (9%)	23,39,42	3.92	7 (30%)
4	2DA	L	822	3,4	17,22,23	0.64	0	13,31,34	1.07	1 (7%)
4	MRG	L	817	1,3,4	22,28,29	1.52	3 (13%)	23,39,42	3.92	8 (34%)
4	2DA	P	822	3,4	17,22,23	0.59	0	13,31,34	1.02	1 (7%)
4	MRG	D	817	1,3,4	22,28,29	1.69	3 (13%)	23,39,42	3.93	8 (34%)
4	2DA	D	822	3,4	17,22,23	0.93	0	13,31,34	1.24	2 (15%)
4	MRG	P	817	3,4	22,28,29	1.65	3 (13%)	23,39,42	3.93	8 (34%)
4	2DA	H	822	3,4	17,22,23	0.77	0	13,31,34	1.12	1 (7%)

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	MRG	H	817	1,3,4	-	3/8/26/27	0/3/3/3
4	2DA	L	822	3,4	-	0/3/18/19	0/3/3/3
4	MRG	L	817	1,3,4	-	3/8/26/27	0/3/3/3
4	2DA	P	822	3,4	-	0/3/18/19	0/3/3/3
4	MRG	D	817	1,3,4	-	3/8/26/27	0/3/3/3
4	2DA	D	822	3,4	-	0/3/18/19	0/3/3/3
4	MRG	P	817	3,4	-	4/8/26/27	0/3/3/3
4	2DA	H	822	3,4	-	0/3/18/19	0/3/3/3

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	H	817	MRG	C21-N2	-6.25	1.33	1.45
4	D	817	MRG	C21-N2	-5.67	1.34	1.45
4	P	817	MRG	C21-N2	-5.61	1.34	1.45
4	L	817	MRG	C21-N2	-5.14	1.35	1.45
4	P	817	MRG	C6-N1	4.08	1.40	1.33

The worst 5 of 36 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	817	MRG	C21-N2-C2	-13.79	99.92	123.75
4	P	817	MRG	C21-N2-C2	-13.71	100.06	123.75
4	L	817	MRG	C21-N2-C2	-13.68	100.11	123.75
4	H	817	MRG	C21-N2-C2	-13.48	100.45	123.75
4	H	817	MRG	C5-C6-N1	-9.09	111.00	123.43

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	D	817	MRG	N3-C2-N2-C21
4	D	817	MRG	N1-C2-N2-C21
4	D	817	MRG	N2-C21-C22-C23
4	H	817	MRG	N3-C2-N2-C21
4	H	817	MRG	N1-C2-N2-C21

There are no ring outliers.

8 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	H	817	MRG	3	0
4	L	822	2DA	1	0
4	L	817	MRG	3	0
4	P	822	2DA	3	0
4	D	817	MRG	2	0
4	D	822	2DA	1	0
4	P	817	MRG	3	0
4	H	822	2DA	1	0

5.5 Carbohydrates (i)

2 monosaccharides 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	$\#\|Z\|>2$
5	GLC	Q	1	5	11,11,12	0.84	1 (9%)	15,15,17	1.32	3 (20%)
5	Z9N	Q	2	5	11,12,12	0.99	1 (9%)	10,18,18	0.97	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
5	GLC	Q	1	5	-	2/2/19/22	0/1/1/1
5	Z9N	Q	2	5	-	0/5/24/24	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	Q	2	Z9N	O2-C2	2.56	1.45	1.40
5	Q	1	GLC	C2-C3	2.08	1.55	1.52

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	Q	1	GLC	C1-C2-C3	3.11	113.48	109.67
5	Q	1	GLC	O5-C5-C6	2.27	110.76	107.20
5	Q	1	GLC	C1-O5-C5	2.23	115.22	112.19

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	Q	1	GLC	O5-C5-C6-O6
5	Q	1	GLC	C4-C5-C6-O6

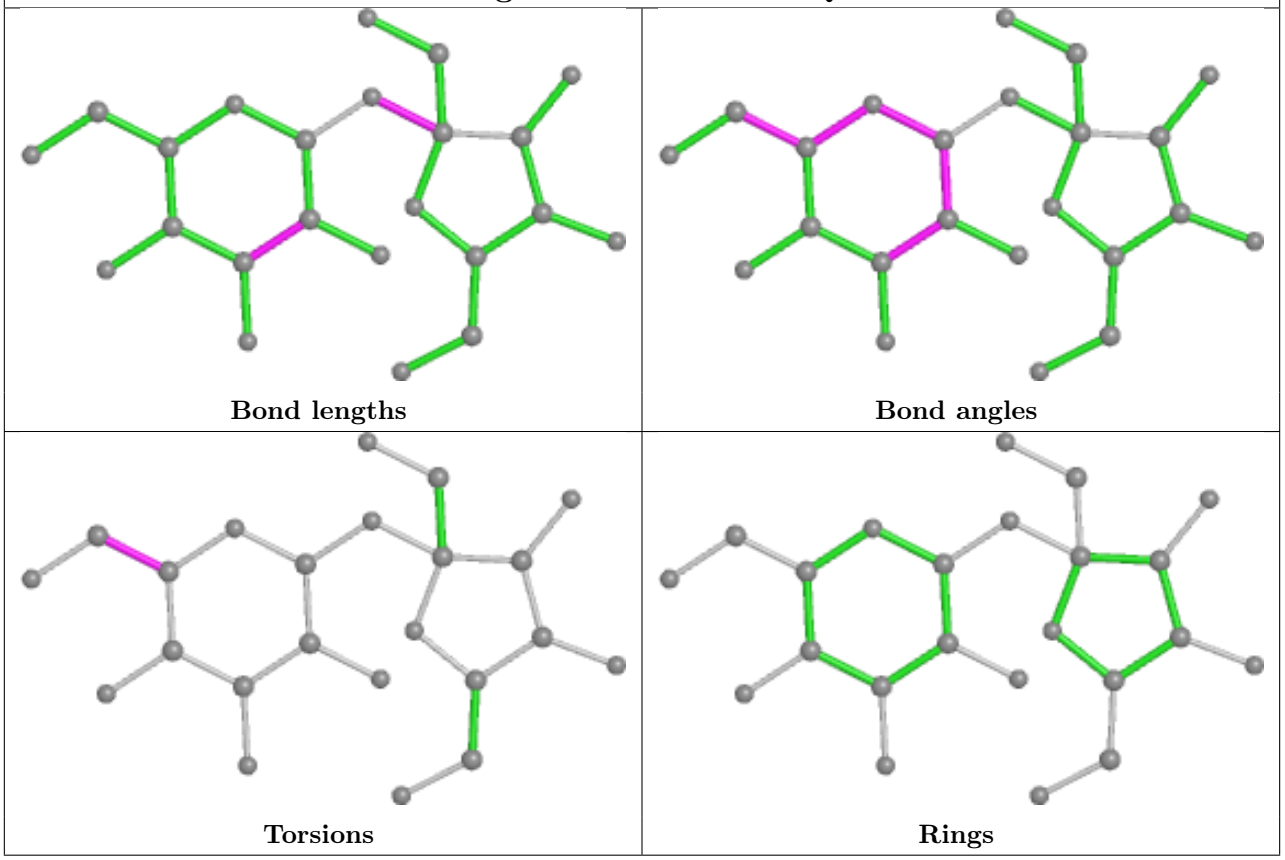
There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	Q	1	GLC	4	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

Oligosaccharide Chain Q



5.6 Ligand geometry (i)

Of 14 ligands modelled in this entry, 8 are monoatomic - leaving 6 for Mogul analysis.

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	$\#\|Z\| > 2$
8	GOL	F	438	-	5,5,5	0.52	0	5,5,5	0.34	0
7	ZP4	I	823	6	44,57,57	1.17	5 (11%)	48,88,88	2.47	8 (16%)
7	ZP4	A	823	6	44,57,57	1.22	5 (11%)	48,88,88	2.49	8 (16%)
8	GOL	B	438	-	5,5,5	0.64	0	5,5,5	0.33	0
7	ZP4	M	823	6	44,57,57	1.23	5 (11%)	48,88,88	2.52	6 (12%)
7	ZP4	E	823	6	44,57,57	1.17	6 (13%)	48,88,88	2.53	7 (14%)

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
8	GOL	F	438	-	-	4/4/4/4	-
7	ZP4	I	823	6	-	5/34/69/69	0/5/5/5
7	ZP4	A	823	6	-	6/34/69/69	0/5/5/5
8	GOL	B	438	-	-	4/4/4/4	-
7	ZP4	M	823	6	-	6/34/69/69	0/5/5/5
7	ZP4	E	823	6	-	6/34/69/69	0/5/5/5

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	M	823	ZP4	C4-N3	3.92	1.39	1.33
7	I	823	ZP4	C4-N3	3.27	1.38	1.33
7	A	823	ZP4	C5R-C4R	3.16	1.49	1.40
7	A	823	ZP4	C2R-N3R	3.15	1.37	1.32
7	I	823	ZP4	C5R-C4R	3.06	1.49	1.40

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	M	823	ZP4	C4-N3-C2	14.39	127.29	115.14
7	I	823	ZP4	C4-N3-C2	14.23	127.16	115.14
7	E	823	ZP4	C4-N3-C2	14.23	127.15	115.14
7	A	823	ZP4	C4-N3-C2	13.94	126.91	115.14
7	M	823	ZP4	N3R-C2R-N1R	-3.79	122.75	128.68

There are no chirality outliers.

5 of 31 torsion outliers are listed below:

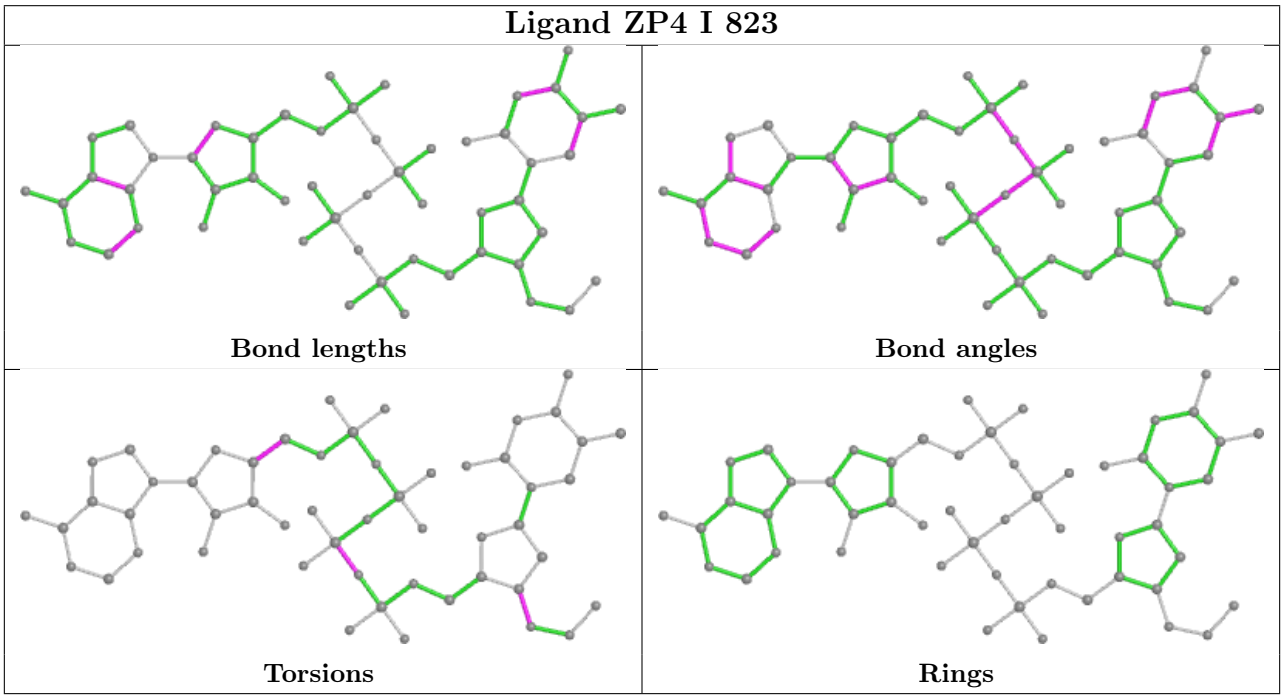
Mol	Chain	Res	Type	Atoms
7	A	823	ZP4	C4'-C3'-N3'-N3A
7	E	823	ZP4	C4'-C3'-N3'-N3A
7	I	823	ZP4	C4'-C3'-N3'-N3A
7	M	823	ZP4	C4'-C3'-N3'-N3A
8	B	438	GOL	O1-C1-C2-C3

There are no ring outliers.

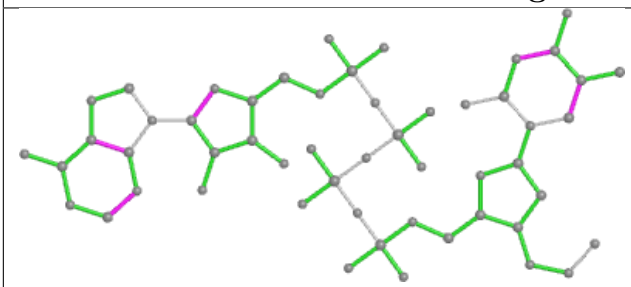
4 monomers are involved in 30 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	I	823	ZP4	4	0
7	A	823	ZP4	10	0
7	M	823	ZP4	7	0
7	E	823	ZP4	9	0

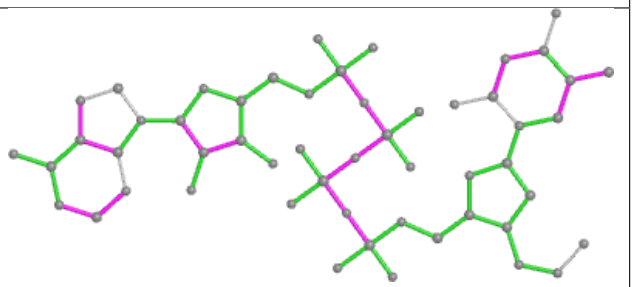
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



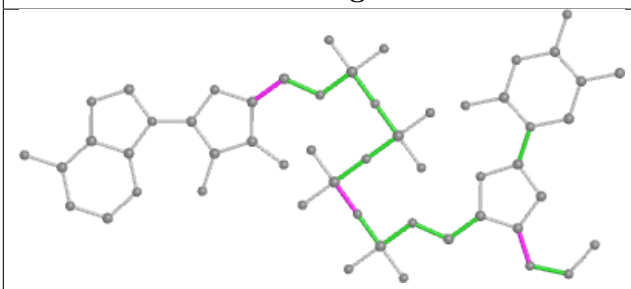
Ligand ZP4 A 823



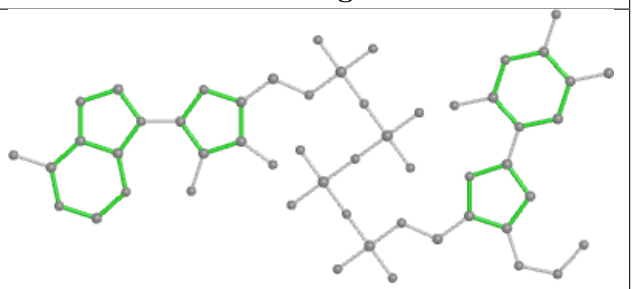
Bond lengths



Bond angles

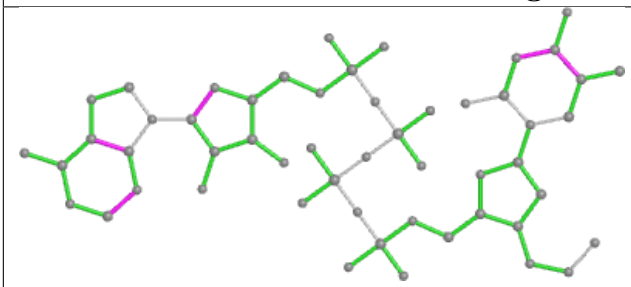


Torsions

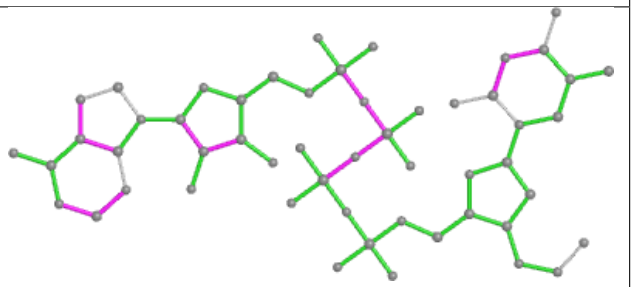


Rings

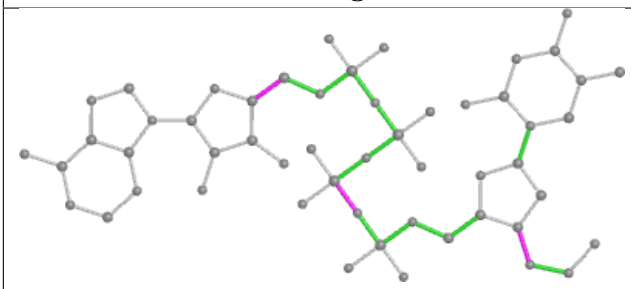
Ligand ZP4 M 823



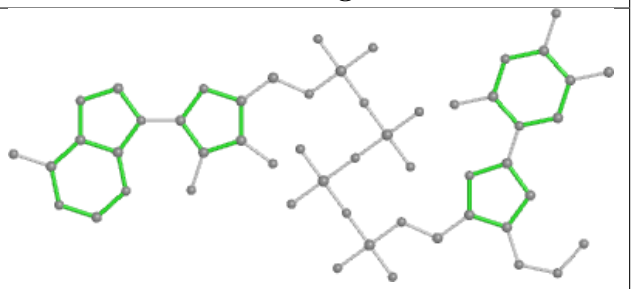
Bond lengths



Bond angles

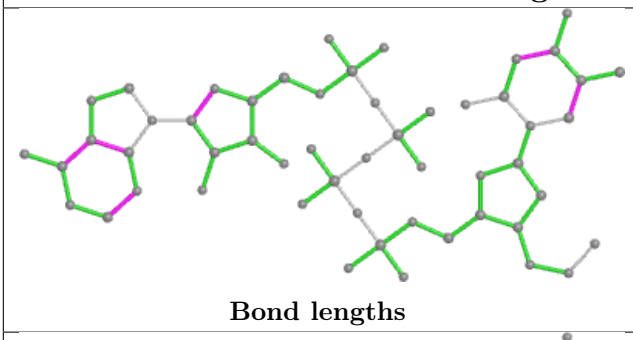


Torsions

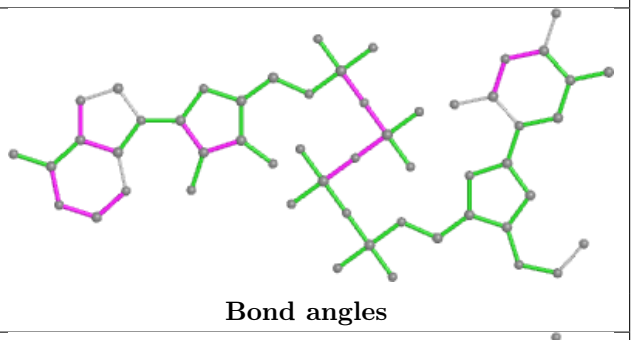


Rings

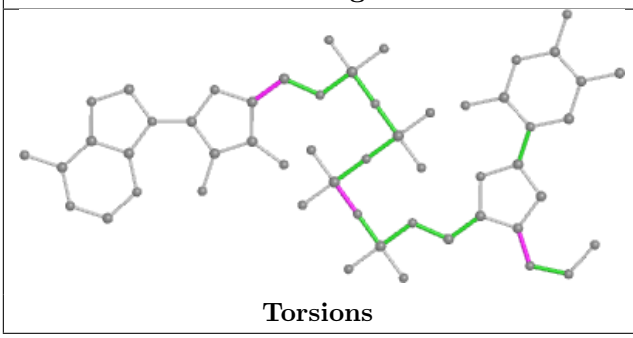
Ligand ZP4 E 823



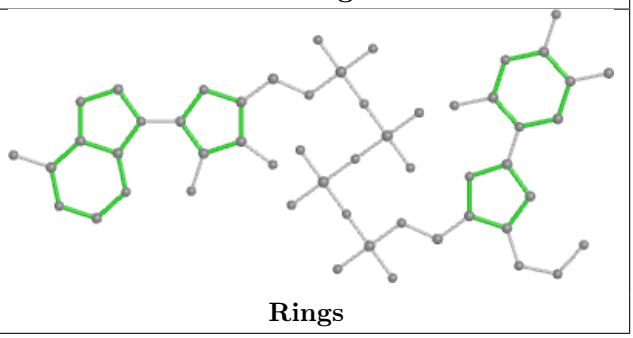
Bond lengths



Bond angles



Torsions



Rings

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	560/562 (99%)	0.16	8 (1%) 75 63	31, 60, 76, 87	0
1	E	560/562 (99%)	0.18	8 (1%) 75 63	35, 62, 78, 88	0
1	I	560/562 (99%)	0.81	88 (15%) 2 1	100, 115, 121, 124	0
1	M	560/562 (99%)	1.72	187 (33%) 0 0	109, 129, 134, 137	0
2	B	425/437 (97%)	0.32	14 (3%) 46 30	22, 54, 84, 97	0
2	F	419/437 (95%)	0.30	15 (3%) 42 27	29, 58, 84, 91	0
2	J	419/437 (95%)	1.08	89 (21%) 0 1	102, 115, 121, 126	0
2	N	419/437 (95%)	2.05	167 (39%) 0 0	117, 129, 133, 137	0
3	C	25/27 (92%)	-0.08	0 100 100	29, 62, 82, 92	0
3	G	25/27 (92%)	-0.14	0 100 100	32, 65, 84, 89	0
3	K	25/27 (92%)	-0.32	0 100 100	111, 117, 123, 125	0
3	O	25/27 (92%)	0.67	4 (16%) 1 1	127, 130, 135, 139	0
4	D	19/21 (90%)	-0.08	0 100 100	40, 58, 87, 89	0
4	H	19/21 (90%)	-0.14	0 100 100	42, 62, 88, 96	0
4	L	19/21 (90%)	-0.09	1 (5%) 26 14	105, 115, 123, 123	0
4	P	19/21 (90%)	0.40	2 (10%) 6 3	124, 131, 134, 134	0
All	All	4098/4188 (97%)	0.78	583 (14%) 2 1	22, 94, 131, 139	0

The worst 5 of 583 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	N	325	LEU	15.1
2	N	21	VAL	13.1
2	N	108	VAL	11.0
2	N	159	ILE	11.0
2	N	3	SER	10.5

6.2 Non-standard residues in protein, DNA, RNA chains (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	2DA	P	822	20/21	0.46	0.37	131,132,133,133	0
4	MRG	P	817	26/27	0.76	0.16	125,132,138,143	0
4	MRG	L	817	26/27	0.79	0.20	117,120,123,124	0
4	2DA	L	822	20/21	0.90	0.18	112,114,115,115	0
4	MRG	H	817	26/27	0.95	0.21	49,55,62,63	0
4	MRG	D	817	26/27	0.96	0.24	46,51,54,55	0
4	2DA	H	822	20/21	0.96	0.21	46,52,61,61	0
4	2DA	D	822	20/21	0.96	0.22	41,45,48,50	0

6.3 Carbohydrates (i)

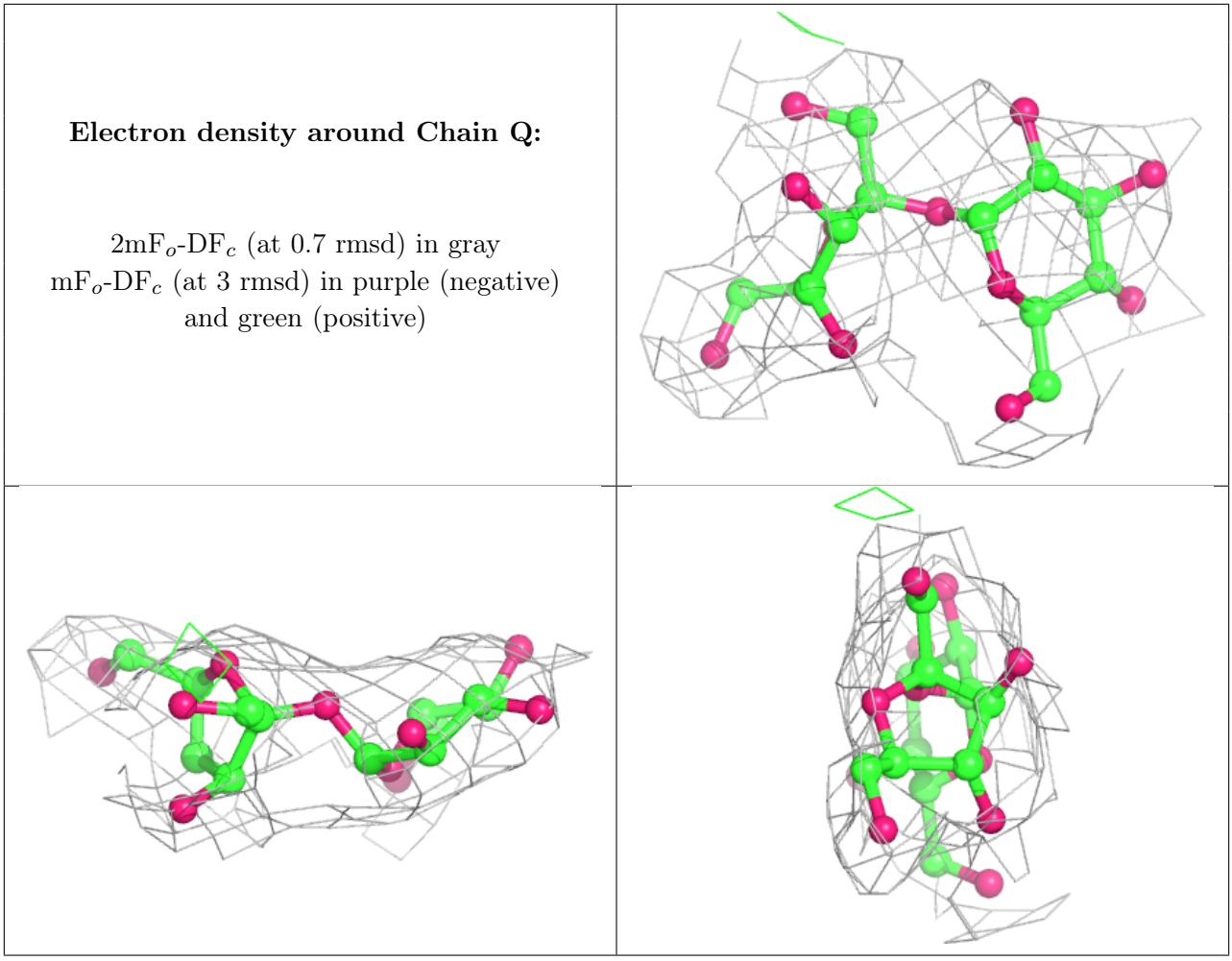
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
5	GLC	Q	1	11/12	0.78	0.22	96,99,100,100	0
5	Z9N	Q	2	12/12	0.78	0.20	94,99,100,101	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around Chain Q:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



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	M	600	1/1	0.43	0.15	123,123,123,123	0
8	GOL	F	438	6/6	0.68	0.28	69,71,73,74	0
7	ZP4	M	823	53/53	0.71	0.27	128,135,138,139	0
8	GOL	B	438	6/6	0.79	0.37	61,65,67,67	0
6	MG	M	602	1/1	0.82	0.29	94,94,94,94	0
7	ZP4	I	823	53/53	0.89	0.19	107,121,125,126	0
7	ZP4	E	823	53/53	0.91	0.25	50,75,91,94	0
7	ZP4	A	823	53/53	0.92	0.27	45,69,92,94	0
6	MG	A	602	1/1	0.95	0.15	24,24,24,24	0
6	MG	E	602	1/1	0.95	0.22	13,13,13,13	0

Continued on next page...

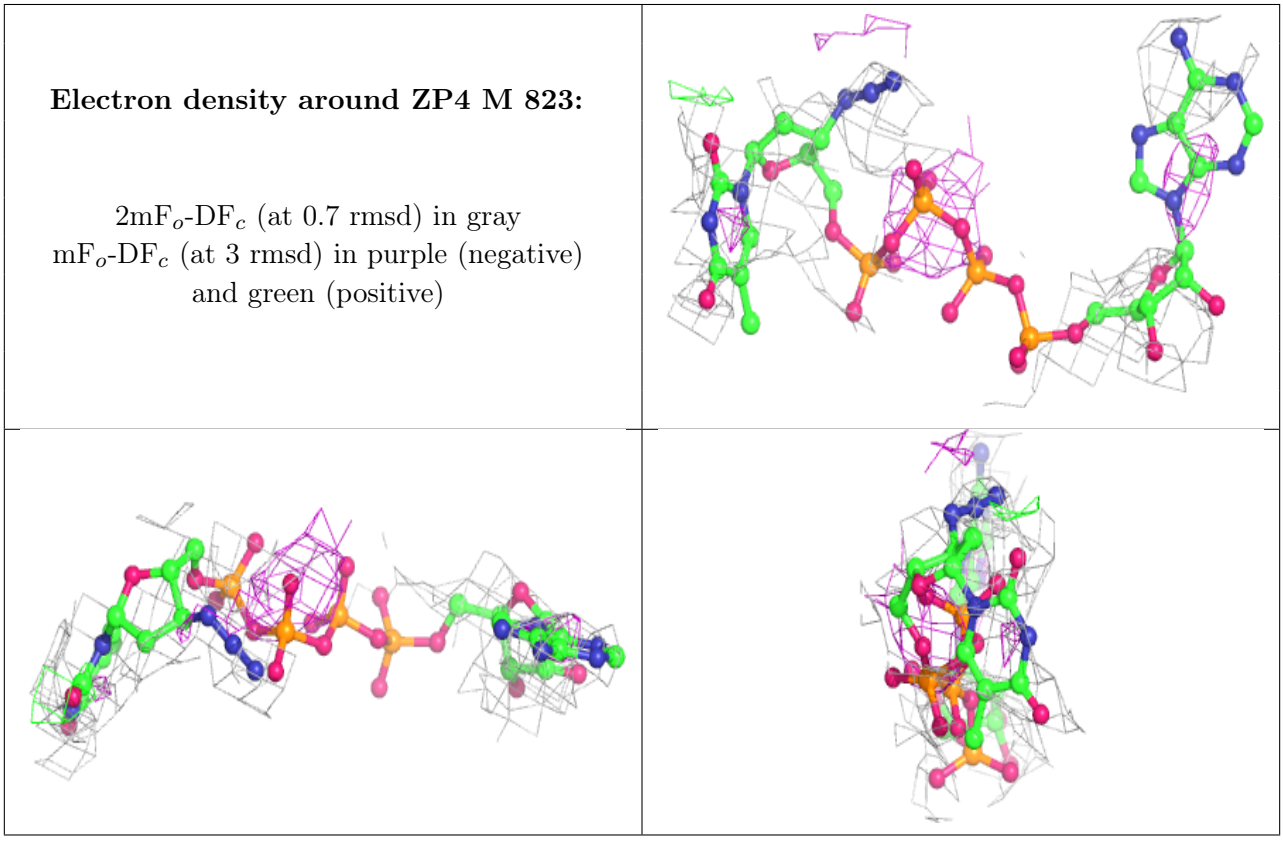
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
6	MG	I	600	1/1	0.96	0.26	96,96,96,96	0
6	MG	I	602	1/1	0.96	0.19	64,64,64,64	0
6	MG	A	600	1/1	0.97	0.18	27,27,27,27	0
6	MG	E	600	1/1	0.99	0.21	28,28,28,28	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

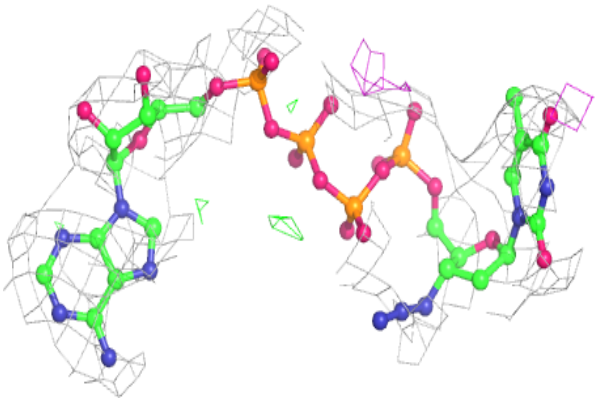
Electron density around ZP4 M 823:

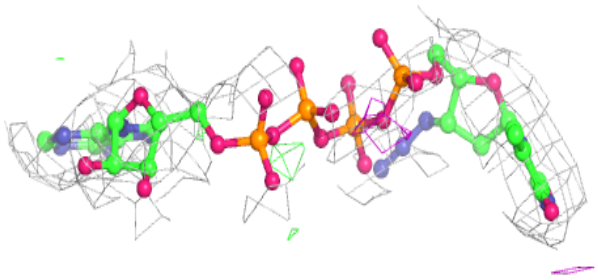
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

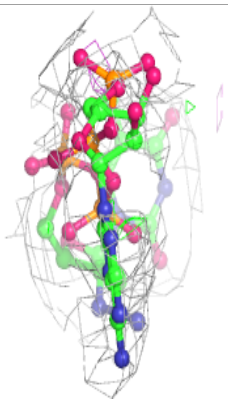


Electron density around ZP4 I 823:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

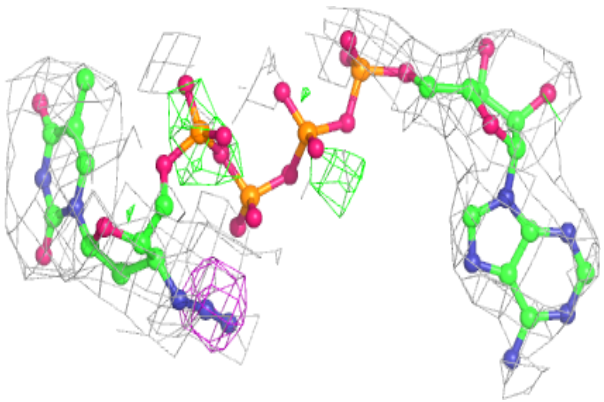


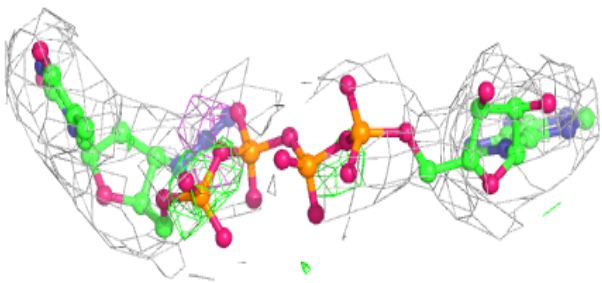


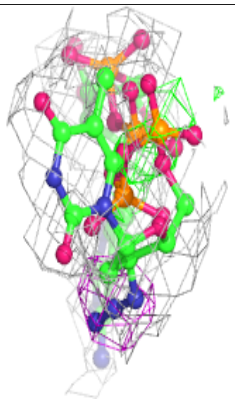


Electron density around ZP4 E 823:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

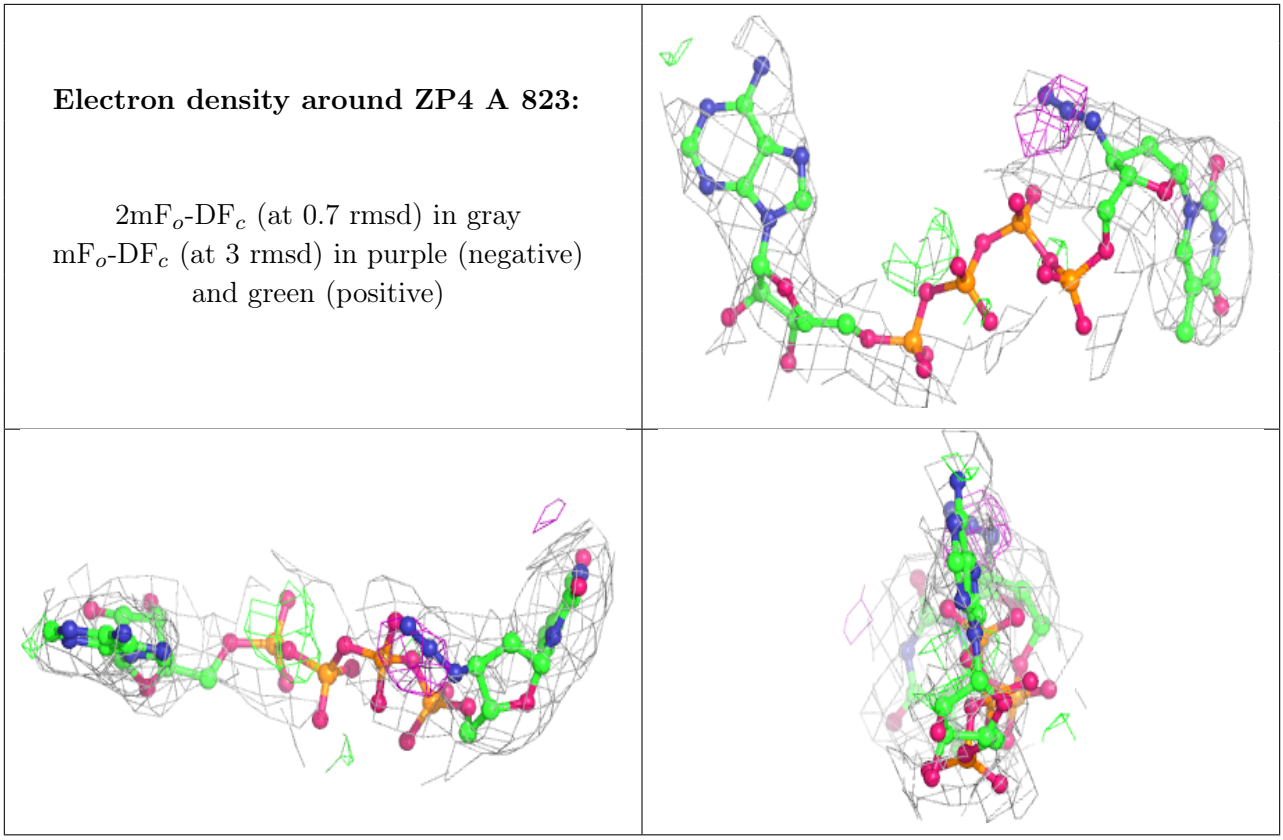






Electron density around ZP4 A 823:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.