wwPDB X-ray Structure Validation Summary Report (i)

Aug 28, 2023 – 10:33 AM EDT

PDB ID	:	3KLF
Title	:	Crystal structure of wild-type 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	:	3.15 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:

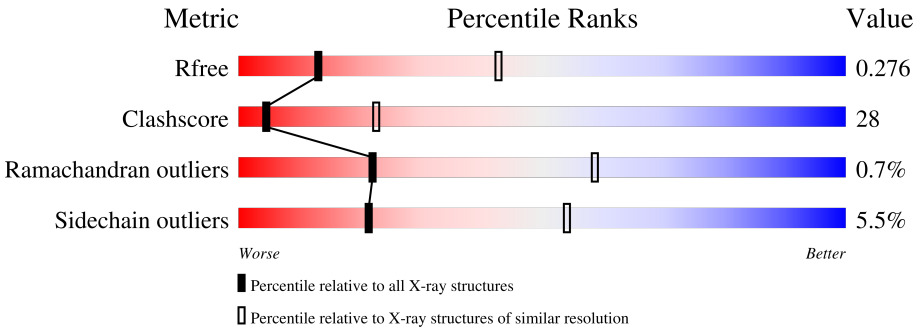
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 3.15 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	1665 (3.20-3.12)
Clashscore	141614	1804 (3.20-3.12)
Ramachandran outliers	138981	1770 (3.20-3.12)
Sidechain outliers	138945	1769 (3.20-3.12)

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%

Mol	Chain	Length	Quality of chain
1	A	557	48% 46% 5%
1	E	557	50% 45%
1	I	557	50% 45%
1	M	557	48% 47%
2	B	444	55% 36% 7%
2	F	444	58% 32% 7%
2	J	444	57% 34% 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	N	444	55% 36% 7%
3	C	27	33% 48% 15%
3	G	27	30% 56% 15%
3	K	27	30% 56% 15%
3	O	27	30% 56% 15%
4	D	21	19% 43% 33% 5%
4	H	21	29% 52% 14% 5%
4	L	21	33% 52% 10% 5%
4	P	21	29% 57% 10% 5%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 35440 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	553	Total	C	N	O	S	0	0	0
			4499	2912	748	831	8			
1	E	553	Total	C	N	O	S	0	0	0
			4499	2912	748	831	8			
1	I	553	Total	C	N	O	S	0	0	0
			4499	2912	748	831	8			
1	M	553	Total	C	N	O	S	0	0	0
			4499	2912	748	831	8			

There are 16 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	258	CYS	GLN	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366
E	-1	MET	-	expression tag	UNP P03366
E	0	VAL	-	expression tag	UNP P03366
E	258	CYS	GLN	engineered mutation	UNP P03366
E	280	SER	CYS	engineered mutation	UNP P03366
I	-1	MET	-	expression tag	UNP P03366
I	0	VAL	-	expression tag	UNP P03366
I	258	CYS	GLN	engineered mutation	UNP P03366
I	280	SER	CYS	engineered mutation	UNP P03366
M	-1	MET	-	expression tag	UNP P03366
M	0	VAL	-	expression tag	UNP P03366
M	258	CYS	GLN	engineered mutation	UNP P03366
M	280	SER	CYS	engineered mutation	UNP P03366

- Molecule 2 is a protein called p51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	413	Total	C	N	O	S	0	0	0
			3405	2215	564	620	6			
2	F	413	Total	C	N	O	S	0	0	0
			3405	2215	564	620	6			
2	J	413	Total	C	N	O	S	0	0	0
			3405	2215	564	620	6			
2	N	413	Total	C	N	O	S	0	0	0
			3405	2215	564	620	6			

There are 68 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-15	MET	-	expression tag	UNP P03366
B	-14	ALA	-	expression tag	UNP P03366
B	-13	HIS	-	expression tag	UNP P03366
B	-12	HIS	-	expression tag	UNP P03366
B	-11	HIS	-	expression tag	UNP P03366
B	-10	HIS	-	expression tag	UNP P03366
B	-9	HIS	-	expression tag	UNP P03366
B	-8	HIS	-	expression tag	UNP P03366
B	-7	ALA	-	expression tag	UNP P03366
B	-6	LEU	-	expression tag	UNP P03366
B	-5	GLU	-	expression tag	UNP P03366
B	-4	VAL	-	expression tag	UNP P03366
B	-3	LEU	-	expression tag	UNP P03366
B	-2	PHE	-	expression tag	UNP P03366
B	-1	GLN	-	expression tag	UNP P03366
B	0	GLY	-	expression tag	UNP P03366
B	280	SER	CYS	engineered mutation	UNP P03366
F	-15	MET	-	expression tag	UNP P03366
F	-14	ALA	-	expression tag	UNP P03366
F	-13	HIS	-	expression tag	UNP P03366
F	-12	HIS	-	expression tag	UNP P03366
F	-11	HIS	-	expression tag	UNP P03366
F	-10	HIS	-	expression tag	UNP P03366
F	-9	HIS	-	expression tag	UNP P03366
F	-8	HIS	-	expression tag	UNP P03366
F	-7	ALA	-	expression tag	UNP P03366
F	-6	LEU	-	expression tag	UNP P03366
F	-5	GLU	-	expression tag	UNP P03366
F	-4	VAL	-	expression tag	UNP P03366
F	-3	LEU	-	expression tag	UNP P03366
F	-2	PHE	-	expression tag	UNP P03366

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	GLN	-	expression tag	UNP P03366
F	0	GLY	-	expression tag	UNP P03366
F	280	SER	CYS	engineered mutation	UNP P03366
J	-15	MET	-	expression tag	UNP P03366
J	-14	ALA	-	expression tag	UNP P03366
J	-13	HIS	-	expression tag	UNP P03366
J	-12	HIS	-	expression tag	UNP P03366
J	-11	HIS	-	expression tag	UNP P03366
J	-10	HIS	-	expression tag	UNP P03366
J	-9	HIS	-	expression tag	UNP P03366
J	-8	HIS	-	expression tag	UNP P03366
J	-7	ALA	-	expression tag	UNP P03366
J	-6	LEU	-	expression tag	UNP P03366
J	-5	GLU	-	expression tag	UNP P03366
J	-4	VAL	-	expression tag	UNP P03366
J	-3	LEU	-	expression tag	UNP P03366
J	-2	PHE	-	expression tag	UNP P03366
J	-1	GLN	-	expression tag	UNP P03366
J	0	GLY	-	expression tag	UNP P03366
J	280	SER	CYS	engineered mutation	UNP P03366
N	-15	MET	-	expression tag	UNP P03366
N	-14	ALA	-	expression tag	UNP P03366
N	-13	HIS	-	expression tag	UNP P03366
N	-12	HIS	-	expression tag	UNP P03366
N	-11	HIS	-	expression tag	UNP P03366
N	-10	HIS	-	expression tag	UNP P03366
N	-9	HIS	-	expression tag	UNP P03366
N	-8	HIS	-	expression tag	UNP P03366
N	-7	ALA	-	expression tag	UNP P03366
N	-6	LEU	-	expression tag	UNP P03366
N	-5	GLU	-	expression tag	UNP P03366
N	-4	VAL	-	expression tag	UNP P03366
N	-3	LEU	-	expression tag	UNP P03366
N	-2	PHE	-	expression tag	UNP P03366
N	-1	GLN	-	expression tag	UNP P03366
N	0	GLY	-	expression tag	UNP P03366
N	280	SER	CYS	engineered mutation	UNP P03366

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	23	Total	C	N	O	P	0	0	0
			473	223	95	133	22			
3	G	23	Total	C	N	O	P	0	0	0
			473	223	95	133	22			
3	K	23	Total	C	N	O	P	0	0	0
			473	223	95	133	22			
3	O	23	Total	C	N	O	P	0	0	0
			473	223	95	133	22			

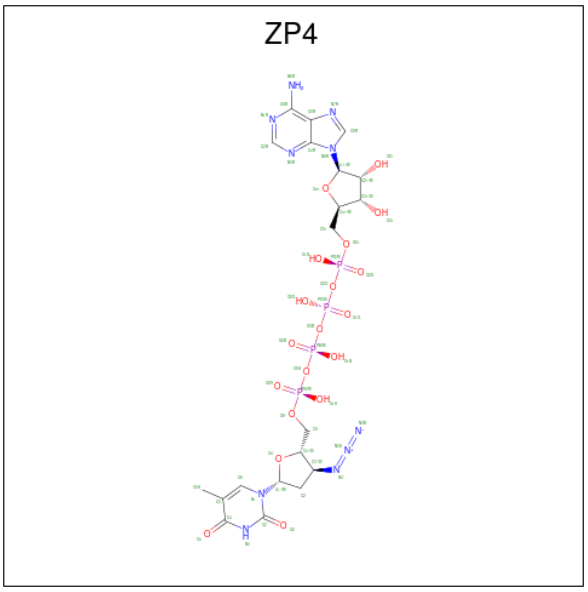
- Molecule 4 is a DNA chain called DNA (5'-D(*A*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(2DA))-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	D	20	Total	C	N	O	P	S	0	0	0
			406	195	72	119	19	1			
4	H	20	Total	C	N	O	P	S	0	0	0
			406	195	72	119	19	1			
4	L	20	Total	C	N	O	P	S	0	0	0
			406	195	72	119	19	1			
4	P	20	Total	C	N	O	P	S	0	0	0
			406	195	72	119	19	1			

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

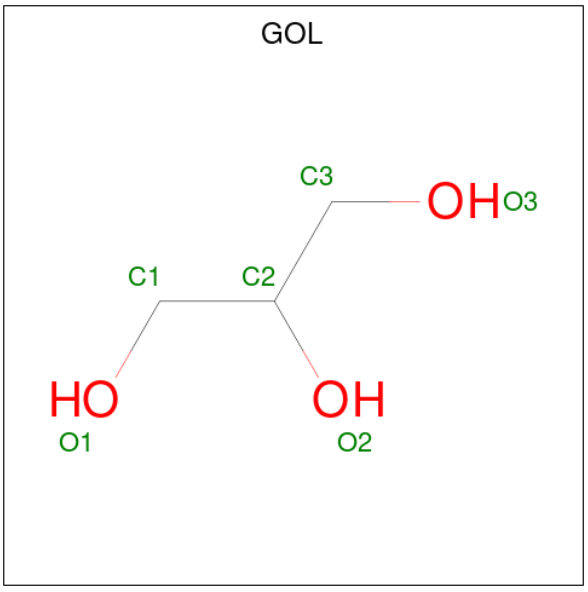
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Mg	0	0
			2	2		
5	E	2	Total	Mg	0	0
			2	2		
5	I	2	Total	Mg	0	0
			2	2		
5	M	2	Total	Mg	0	0
			2	2		

- Molecule 6 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
6	A	1	Total	C	N	O	P	0	0
			53	20	10	19	4		
6	E	1	Total	C	N	O	P	0	0
			53	20	10	19	4		
6	I	1	Total	C	N	O	P	0	0
			53	20	10	19	4		
6	M	1	Total	C	N	O	P	0	0
			53	20	10	19	4		

- Molecule 7 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 8 is water.

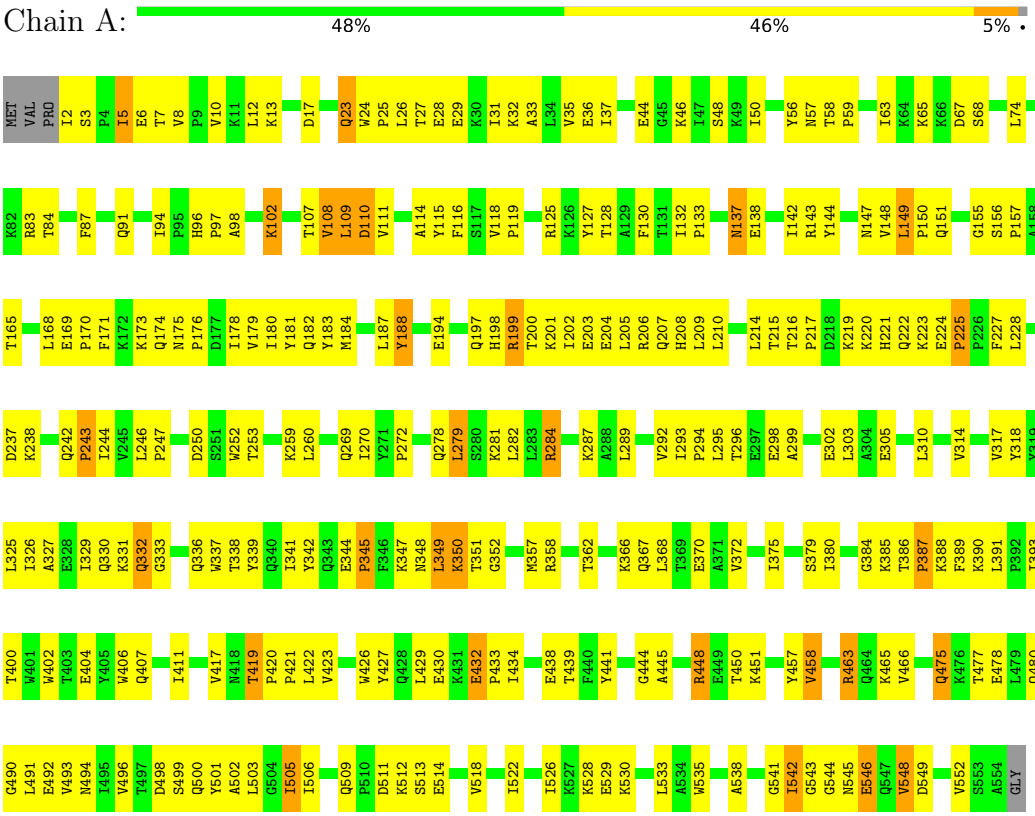
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	12	Total O 12 12	0	0
8	B	9	Total O 9 9	0	0
8	E	11	Total O 11 11	0	0
8	F	12	Total O 12 12	0	0
8	I	10	Total O 10 10	0	0
8	J	5	Total O 5 5	0	0
8	M	3	Total O 3 3	0	0
8	N	5	Total O 5 5	0	0
8	C	1	Total O 1 1	0	0
8	G	3	Total O 3 3	0	0
8	K	1	Total O 1 1	0	0
8	L	2	Total O 2 2	0	0
8	O	2	Total O 2 2	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

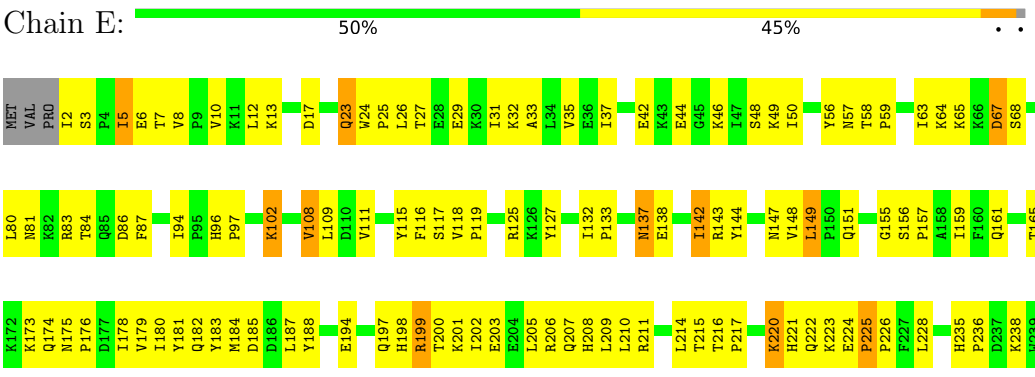
• Molecule 1: Reverse transcriptase/ribonuclease H

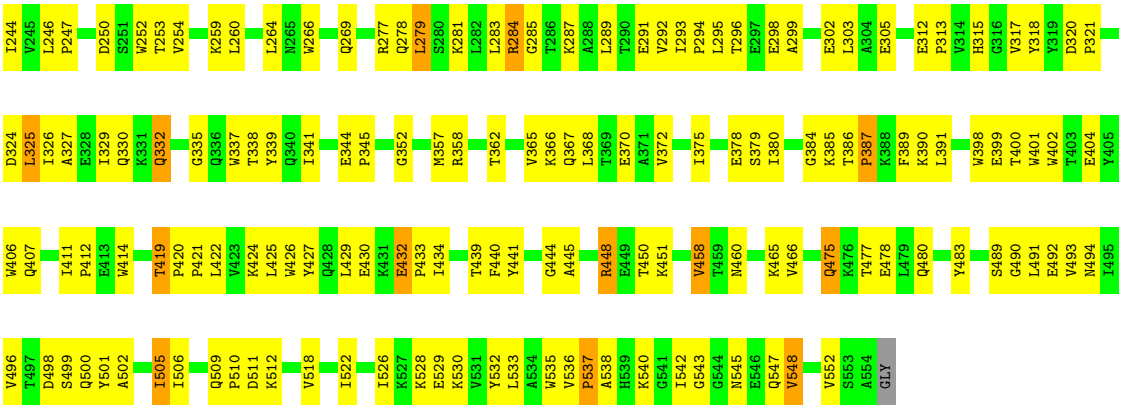
Chain A:



• Molecule 1: Reverse transcriptase/ribonuclease H

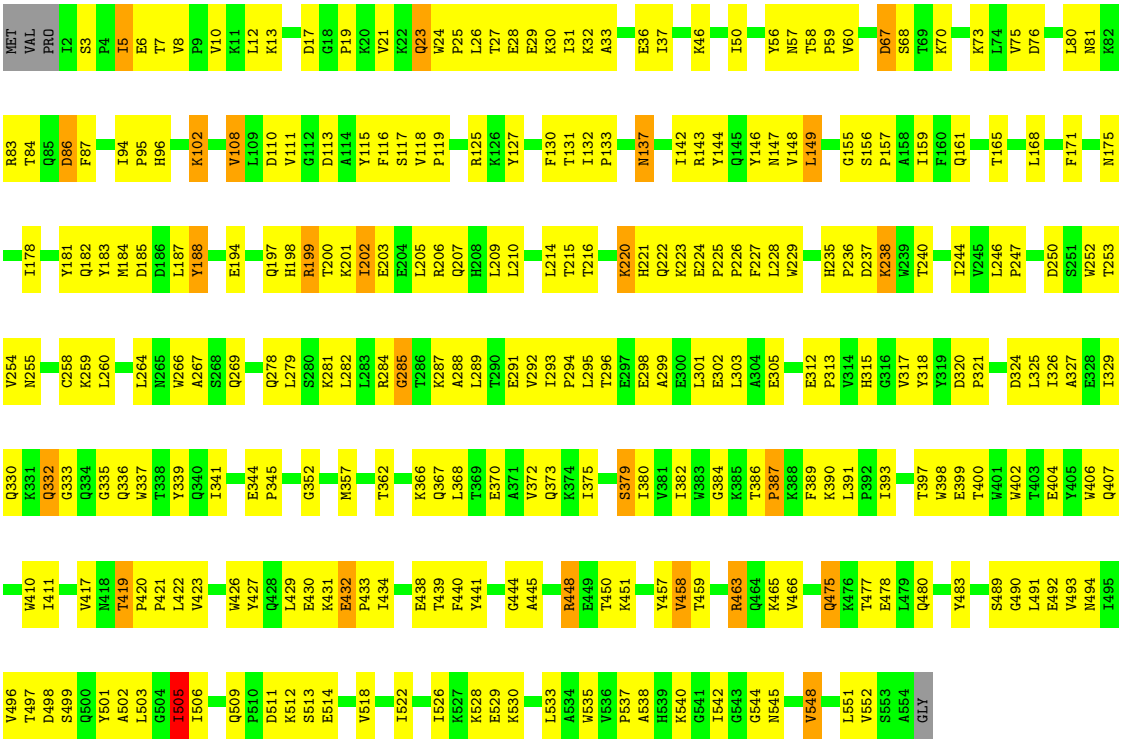
Chain E:





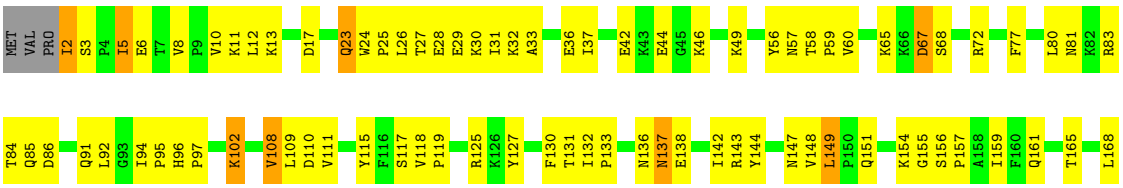
• Molecule 1: Reverse transcriptase/ribonuclease H

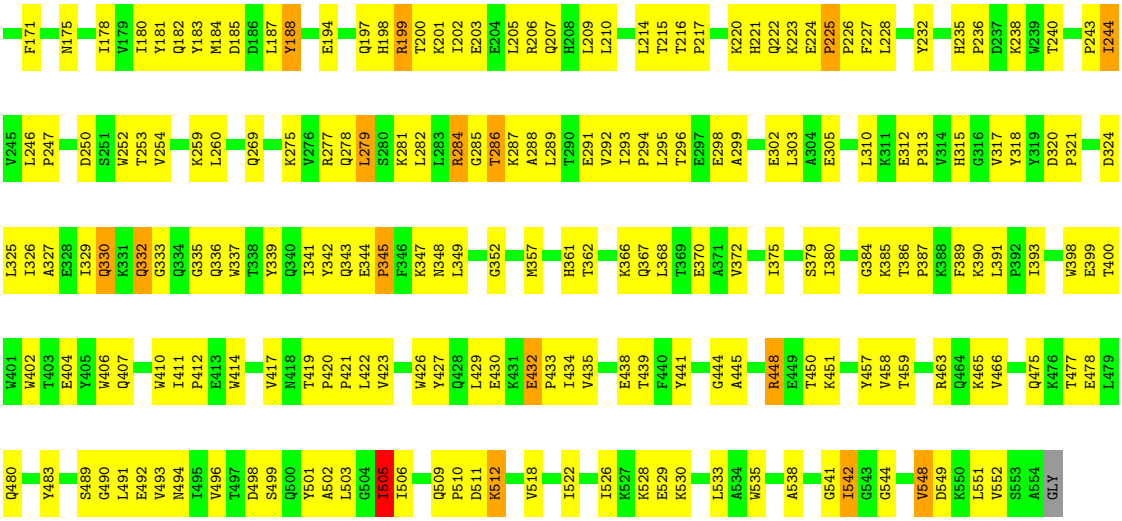
Chain I:



• Molecule 1: Reverse transcriptase/ribonuclease H

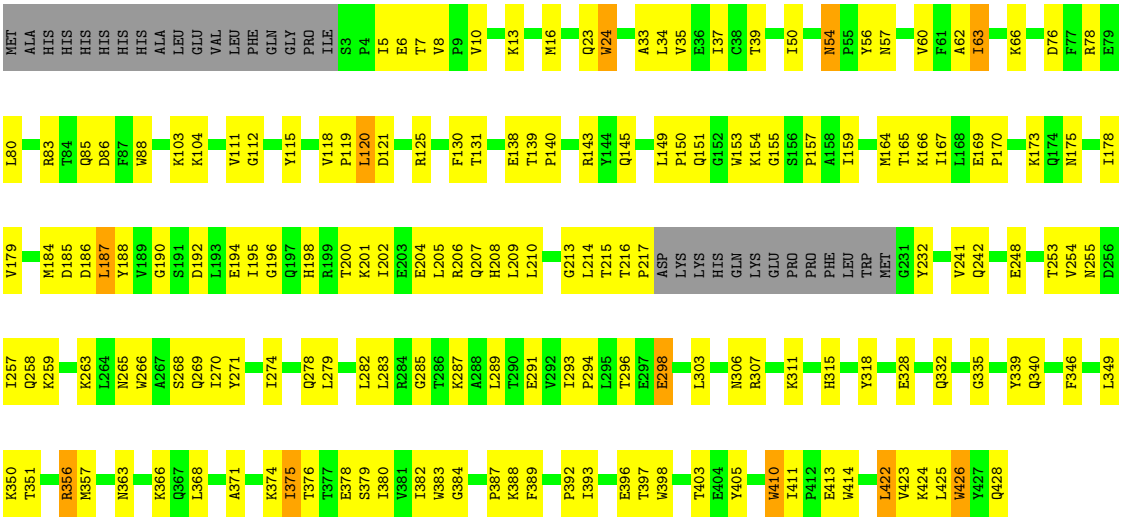
Chain M:





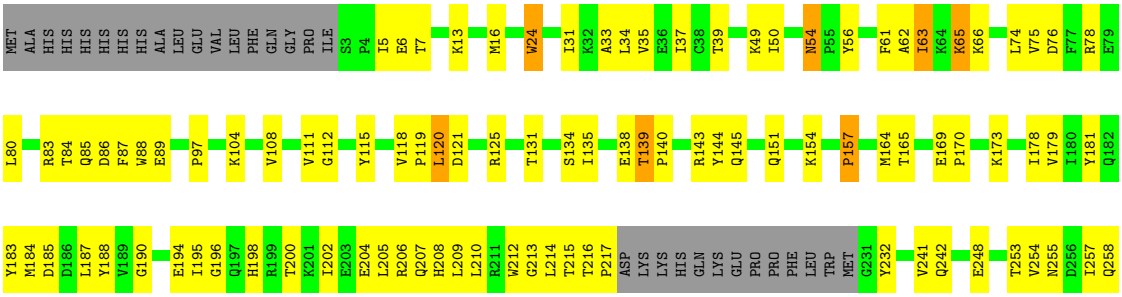
- Molecule 2: p51 RT

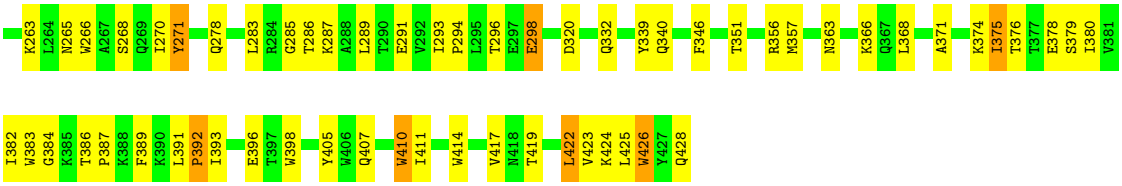
Chain B:



- Molecule 2: p51 RT

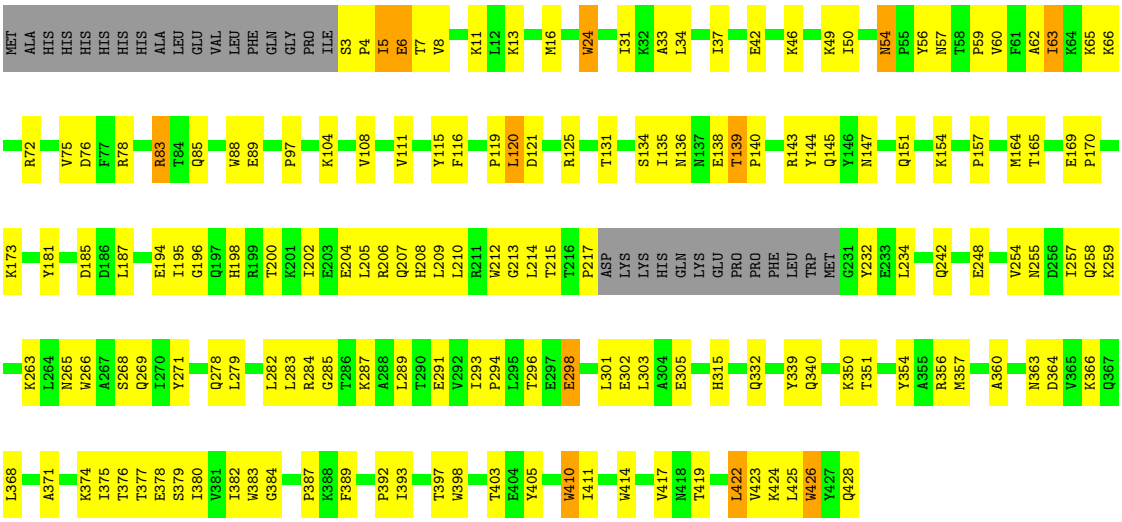
Chain F:





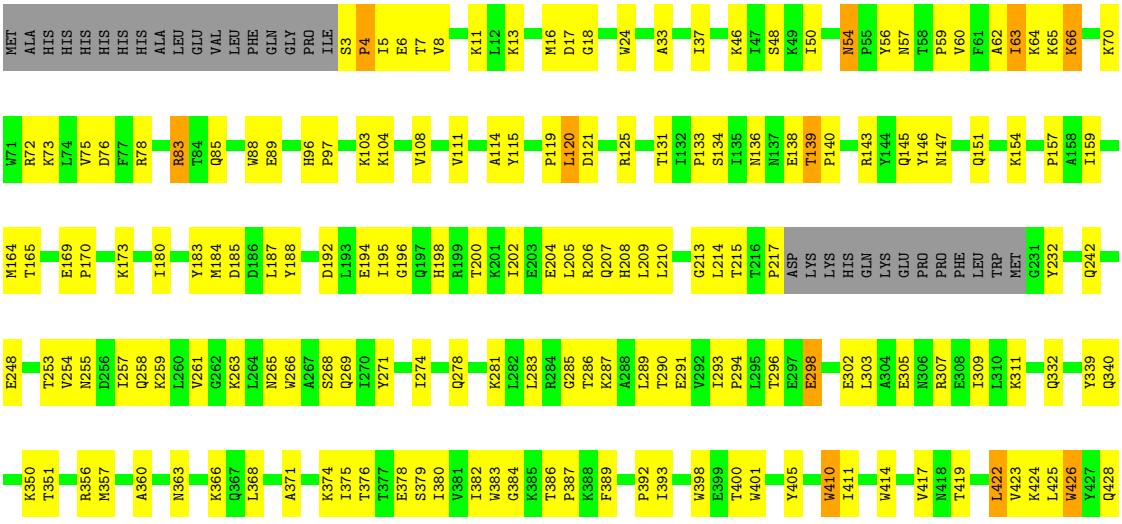
• Molecule 2: p51 RT

Chain J:



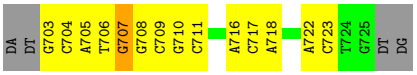
• Molecule 2: p51 RT

Chain N:



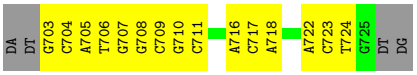
• Molecule 3: DNA (5'-D(*AP*T*GP*CP*AP*TP*GP*GP*CP*GP*CP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*CP*TP*GP*TP*G)-3')

Chain C:



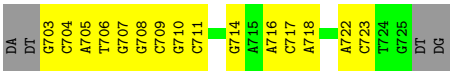
• Molecule 3: DNA (5'-D(*AP*T*GP*CP*AP*TP*GP*GP*CP*GP*CP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*CP*TP*GP*TP*G)-3')

Chain G:



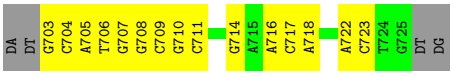
• Molecule 3: DNA (5'-D(*AP*T*GP*CP*AP*TP*GP*GP*CP*GP*CP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*CP*TP*GP*TP*G)-3')

Chain K:



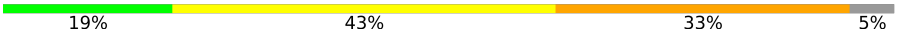
• Molecule 3: DNA (5'-D(*AP*T*GP*CP*AP*TP*GP*GP*CP*GP*CP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*CP*TP*GP*TP*G)-3')

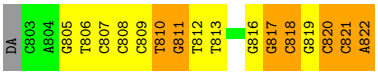
Chain O:



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

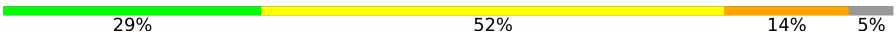
Chain D:

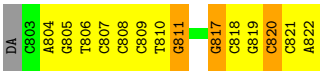




• Molecule 4: DNA (5'-D(*A*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(2DA))-3')

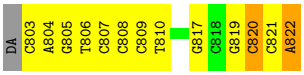
Chain H:





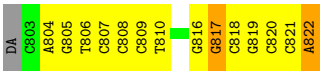
• Molecule 4: DNA (5'-D(*A*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(2DA))-3')

Chain L:



• Molecule 4: DNA (5'-D(*A*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(2DA))-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	78.53Å 274.81Å 152.37Å 90.00° 90.11° 90.00°	Depositor
Resolution (Å)	24.83 – 3.15 24.83 – 3.14	Depositor EDS
% Data completeness (in resolution range)	88.6 (24.83-3.15) 88.2 (24.83-3.14)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.38 (at 3.17Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.264 , 0.288 0.250 , 0.276	Depositor DCC
R_{free} test set	1979 reflections (2.01%)	wwPDB-VP
Wilson B-factor (Å^2)	73.3	Xtriage
Anisotropy	0.001	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 17.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.39$, $< L^2 > = 0.22$	Xtriage
Estimated twinning fraction	0.359 for h,-k,-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	35440	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.71% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, MG, ZP4, 2DA, MRG

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	0.60	0/4616	0.69	0/6271
1	E	0.57	1/4616 (0.0%)	0.66	2/6271 (0.0%)
1	I	0.44	0/4616	0.61	0/6271
1	M	0.40	0/4616	0.60	1/6271 (0.0%)
2	B	0.67	0/3502	0.75	2/4760 (0.0%)
2	F	0.66	0/3502	0.74	1/4760 (0.0%)
2	J	0.47	0/3502	0.66	1/4760 (0.0%)
2	N	0.42	0/3502	0.64	0/4760
3	C	0.78	0/532	0.79	0/820
3	G	0.78	0/532	0.78	0/820
3	K	0.49	0/532	0.72	0/820
3	O	0.46	0/532	0.74	0/820
4	D	0.91	0/400	0.97	1/612 (0.2%)
4	H	0.90	0/400	0.87	0/612
4	L	0.59	0/400	0.78	0/612
4	P	0.50	0/400	0.77	0/612
All	All	0.55	1/36200 (0.0%)	0.68	8/49852 (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
4	D	0	5
4	H	0	2
4	L	0	1
All	All	0	9

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	378	GLU	CG-CD	5.04	1.59	1.51

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	M	286	THR	N-CA-C	5.71	126.42	111.00
2	J	65	LYS	CB-CA-C	-5.67	99.06	110.40
4	D	821	DC	O5'-P-OP1	-5.45	100.80	105.70
2	B	349	LEU	CB-CA-C	5.36	120.38	110.20
1	E	284	ARG	CB-CA-C	-5.35	99.71	110.40

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	C	707	DG	Sidechain
4	D	810	DT	Sidechain
4	D	811	DG	Sidechain
4	D	816	DG	Sidechain
4	D	818	DC	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	4499	0	4556	312	0
1	E	4499	0	4556	310	0
1	I	4499	0	4556	275	0
1	M	4499	0	4556	308	0
2	B	3405	0	3437	163	0
2	F	3405	0	3437	156	0
2	J	3405	0	3437	164	0
2	N	3405	0	3437	175	0
3	C	473	0	257	29	0
3	G	473	0	257	24	0
3	K	473	0	257	22	0
3	O	473	0	257	26	0
4	D	406	0	231	32	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	H	406	0	231	34	0
4	L	406	0	231	20	0
4	P	406	0	231	26	0
5	A	2	0	0	0	0
5	E	2	0	0	0	0
5	I	2	0	0	0	0
5	M	2	0	0	0	0
6	A	53	0	24	4	0
6	E	53	0	24	6	0
6	I	53	0	24	6	0
6	M	53	0	24	3	0
7	B	6	0	8	0	0
7	F	6	0	8	0	0
8	A	12	0	0	3	0
8	B	9	0	0	3	0
8	C	1	0	0	0	0
8	E	11	0	0	5	0
8	F	12	0	0	4	0
8	G	3	0	0	0	0
8	I	10	0	0	5	0
8	J	5	0	0	0	0
8	K	1	0	0	0	0
8	L	2	0	0	0	0
8	M	3	0	0	0	0
8	N	5	0	0	3	0
8	O	2	0	0	0	0
All	All	35440	0	34036	1972	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 28.

The worst 5 of 1972 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:E:511:ASP:OD1	1:E:512:LYS:HG2	1.42	1.19
1:M:339:TYR:CZ	1:M:352:GLY:HA3	1.77	1.19
1:M:511:ASP:OD1	1:M:512:LYS:HG2	1.38	1.18
1:A:175:ASN:HB3	1:A:178:ILE:HD13	1.19	1.18
2:N:66:LYS:NZ	2:N:66:LYS:HB3	1.40	1.18

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	551/557 (99%)	487 (88%)	55 (10%)	9 (2%)	9 40
1	E	551/557 (99%)	491 (89%)	56 (10%)	4 (1%)	22 59
1	I	551/557 (99%)	486 (88%)	60 (11%)	5 (1%)	17 53
1	M	551/557 (99%)	494 (90%)	51 (9%)	6 (1%)	14 48
2	B	409/444 (92%)	367 (90%)	42 (10%)	0	100 100
2	F	409/444 (92%)	363 (89%)	45 (11%)	1 (0%)	47 78
2	J	409/444 (92%)	366 (90%)	42 (10%)	1 (0%)	47 78
2	N	409/444 (92%)	364 (89%)	43 (10%)	2 (0%)	29 65
All	All	3840/4004 (96%)	3418 (89%)	394 (10%)	28 (1%)	22 59

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	542	ILE
1	A	199	ARG
1	E	199	ARG
1	E	537	PRO
2	J	6	GLU

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	494/497 (99%)	467 (94%)	27 (6%)	21 54

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	E	494/497 (99%)	466 (94%)	28 (6%)	20 53
1	I	494/497 (99%)	462 (94%)	32 (6%)	17 48
1	M	494/497 (99%)	466 (94%)	28 (6%)	20 53
2	B	375/403 (93%)	354 (94%)	21 (6%)	21 53
2	F	375/403 (93%)	355 (95%)	20 (5%)	22 55
2	J	375/403 (93%)	358 (96%)	17 (4%)	27 61
2	N	375/403 (93%)	356 (95%)	19 (5%)	24 56
All	All	3476/3600 (97%)	3284 (94%)	192 (6%)	21 54

5 of 192 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	324	ASP
2	J	422	LEU
1	I	379	SER
2	J	54	ASN
1	M	108	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 115 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	269	GLN
2	N	207	GLN
1	I	373	GLN
2	N	175	ASN
1	M	407	GLN

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	2DA	H	822	4,3	17,22,23	0.81	0	13,31,34	1.12	0
4	2DA	L	822	4,3	17,22,23	0.69	0	13,31,34	1.12	2 (15%)
4	MRG	P	817	4,1,3	21,28,29	1.54	3 (14%)	19,39,42	3.25	4 (21%)
4	MRG	D	817	4,1,3	21,28,29	1.91	4 (19%)	19,39,42	3.21	4 (21%)
4	MRG	L	817	4,1,3	21,28,29	1.63	3 (14%)	19,39,42	3.24	5 (26%)
4	2DA	D	822	4,3	17,22,23	0.80	0	13,31,34	0.97	1 (7%)
4	MRG	H	817	4,1,3	21,28,29	1.68	4 (19%)	19,39,42	3.46	5 (26%)
4	2DA	P	822	4,3	17,22,23	0.61	0	13,31,34	1.10	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	2DA	H	822	4,3	-	0/3/18/19	0/3/3/3
4	2DA	L	822	4,3	-	0/3/18/19	0/3/3/3
4	MRG	P	817	4,1,3	-	3/8/26/27	0/3/3/3
4	MRG	D	817	4,1,3	-	3/8/26/27	0/3/3/3
4	MRG	L	817	4,1,3	-	2/8/26/27	0/3/3/3
4	2DA	D	822	4,3	-	0/3/18/19	0/3/3/3
4	MRG	H	817	4,1,3	-	3/8/26/27	0/3/3/3
4	2DA	P	822	4,3	-	0/3/18/19	0/3/3/3

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	817	MRG	C21-N2	-5.70	1.33	1.46
4	L	817	MRG	C21-N2	-5.32	1.34	1.46
4	H	817	MRG	C21-N2	-4.97	1.34	1.46
4	P	817	MRG	C21-N2	-4.93	1.34	1.46
4	D	817	MRG	C5-C6	-4.50	1.38	1.47

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	H	817	MRG	C21-N2-C2	-13.17	98.74	123.41
4	P	817	MRG	C21-N2-C2	-12.61	99.79	123.41
4	L	817	MRG	C21-N2-C2	-12.40	100.18	123.41
4	D	817	MRG	C21-N2-C2	-12.09	100.76	123.41
4	D	817	MRG	C23-C22-C21	-4.38	98.59	112.65

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

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

There are no ring outliers.

7 monomers are involved in 15 short contacts:

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

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	$\#\|Z\| > 2$
6	ZP4	E	823	5	47,57,57	1.26	5 (10%)	58,88,88	1.15	5 (8%)
6	ZP4	A	823	5	47,57,57	1.22	5 (10%)	58,88,88	1.18	5 (8%)
7	GOL	F	429	-	5,5,5	0.77	0	5,5,5	0.76	0
6	ZP4	M	823	5	47,57,57	1.19	4 (8%)	58,88,88	1.14	6 (10%)
7	GOL	B	429	-	5,5,5	0.71	0	5,5,5	0.56	0
6	ZP4	I	823	5	47,57,57	1.22	5 (10%)	58,88,88	1.22	7 (12%)

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
6	ZP4	E	823	5	-	11/37/69/69	0/5/5/5
6	ZP4	A	823	5	-	5/37/69/69	0/5/5/5
7	GOL	F	429	-	-	4/4/4/4	-
6	ZP4	M	823	5	-	7/37/69/69	0/5/5/5
7	GOL	B	429	-	-	2/4/4/4	-
6	ZP4	I	823	5	-	9/37/69/69	0/5/5/5

The worst 5 of 19 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	I	823	ZP4	O41-C11	4.37	1.47	1.41
6	M	823	ZP4	O41-C11	3.83	1.46	1.41
6	A	823	ZP4	C5R-C4R	3.66	1.50	1.40
6	E	823	ZP4	O41-C11	3.64	1.46	1.41
6	M	823	ZP4	C5R-C4R	3.57	1.50	1.40

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	E	823	ZP4	N3R-C2R-N1R	-4.38	121.83	128.68
6	M	823	ZP4	N3R-C2R-N1R	-4.04	122.36	128.68
6	A	823	ZP4	N3R-C2R-N1R	-4.03	122.37	128.68
6	I	823	ZP4	N3R-C2R-N1R	-4.00	122.43	128.68
6	A	823	ZP4	C31-C21-C11	3.13	105.68	100.98

There are no chirality outliers.

5 of 38 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	823	ZP4	C5'-O5'-PA-O1A
6	E	823	ZP4	C5'-O5'-PA-O1A
6	E	823	ZP4	C51-O51-PD-O2D
6	E	823	ZP4	C31-C41-C51-O51
6	I	823	ZP4	C5'-O5'-PA-O1A

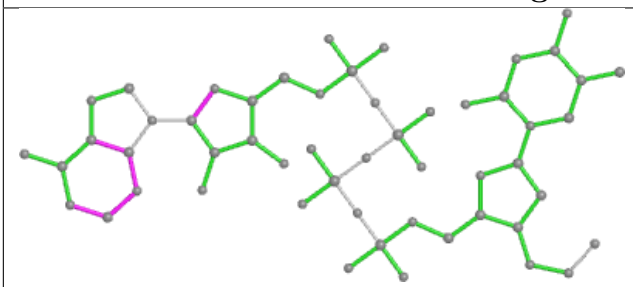
There are no ring outliers.

4 monomers are involved in 19 short contacts:

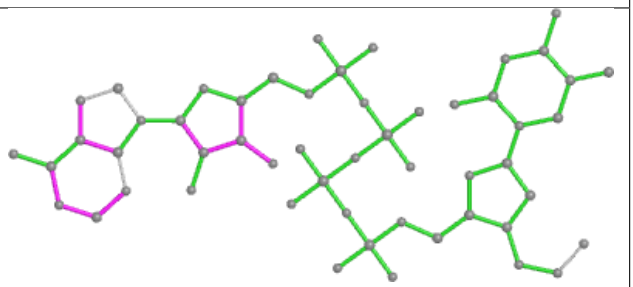
Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	E	823	ZP4	6	0
6	A	823	ZP4	4	0
6	M	823	ZP4	3	0
6	I	823	ZP4	6	0

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.

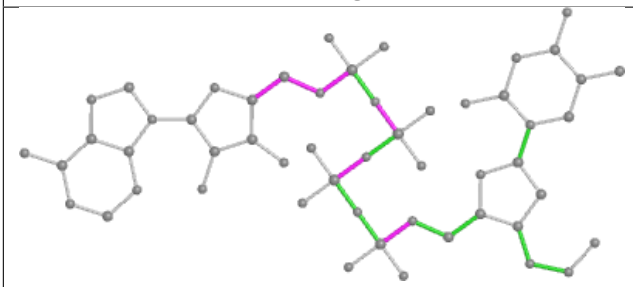
Ligand ZP4 E 823



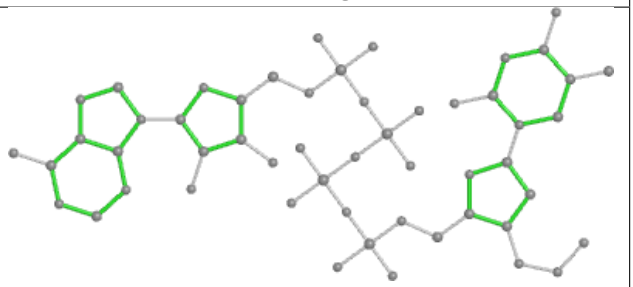
Bond lengths



Bond angles

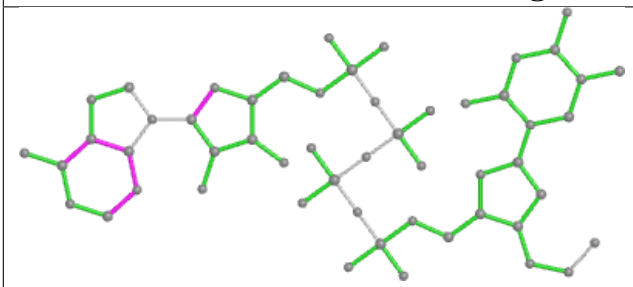


Torsions

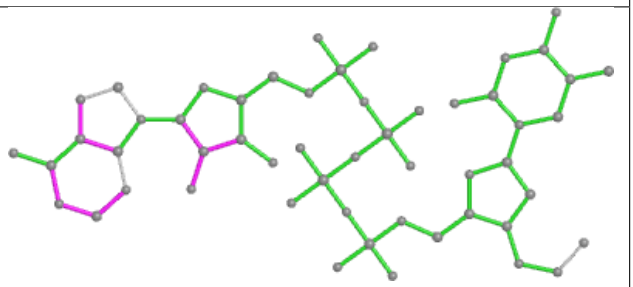


Rings

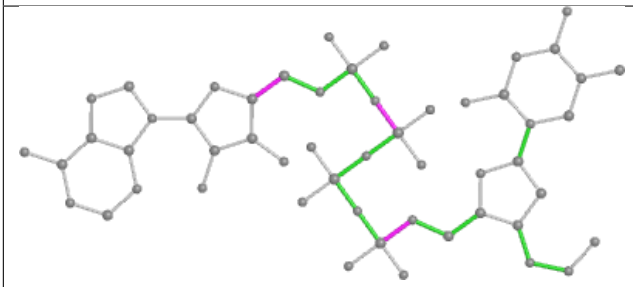
Ligand ZP4 A 823



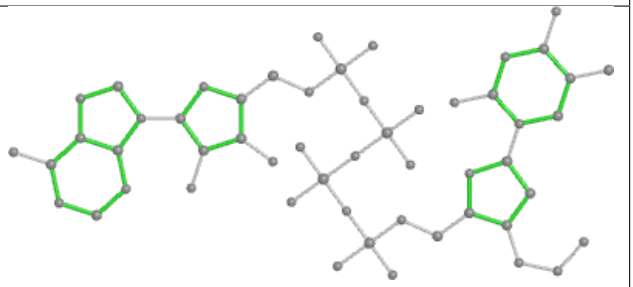
Bond lengths



Bond angles

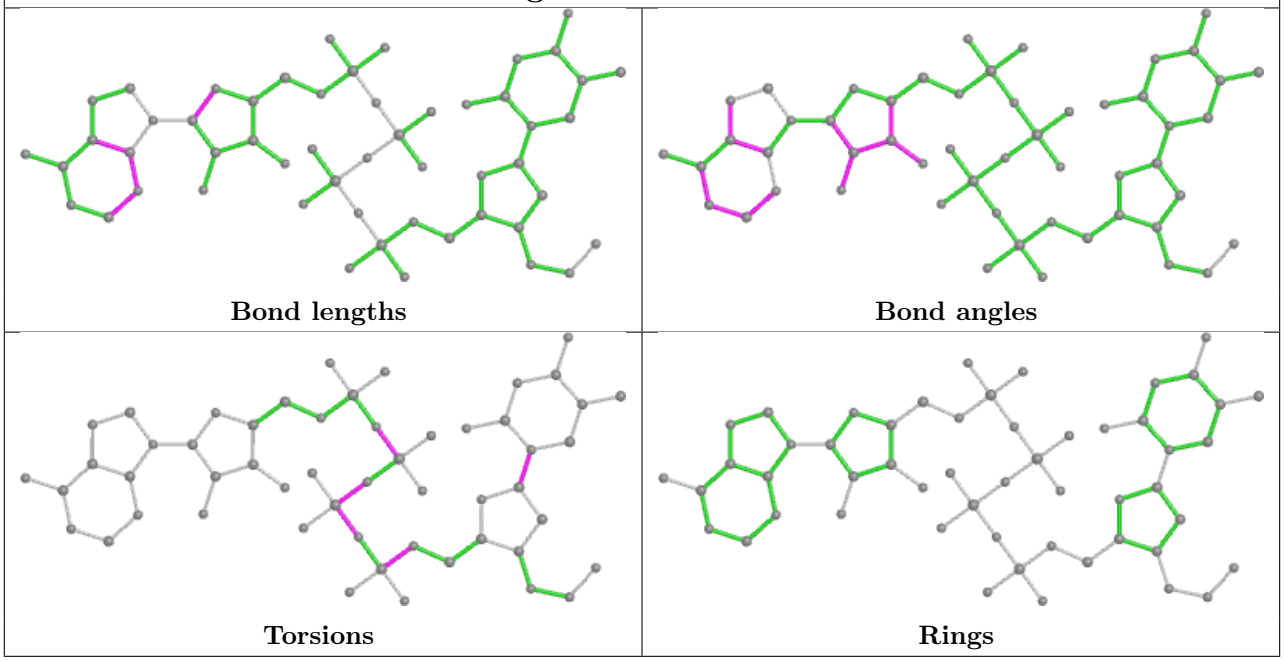


Torsions

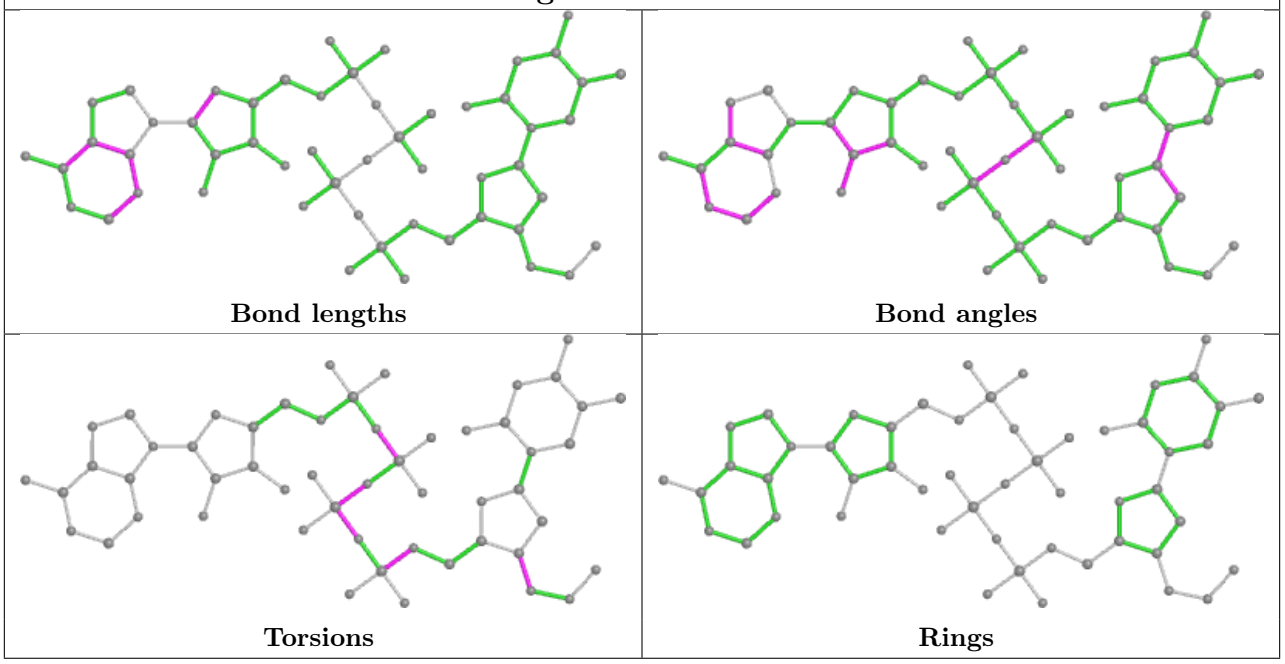


Rings

Ligand ZP4 M 823



Ligand ZP4 I 823



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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

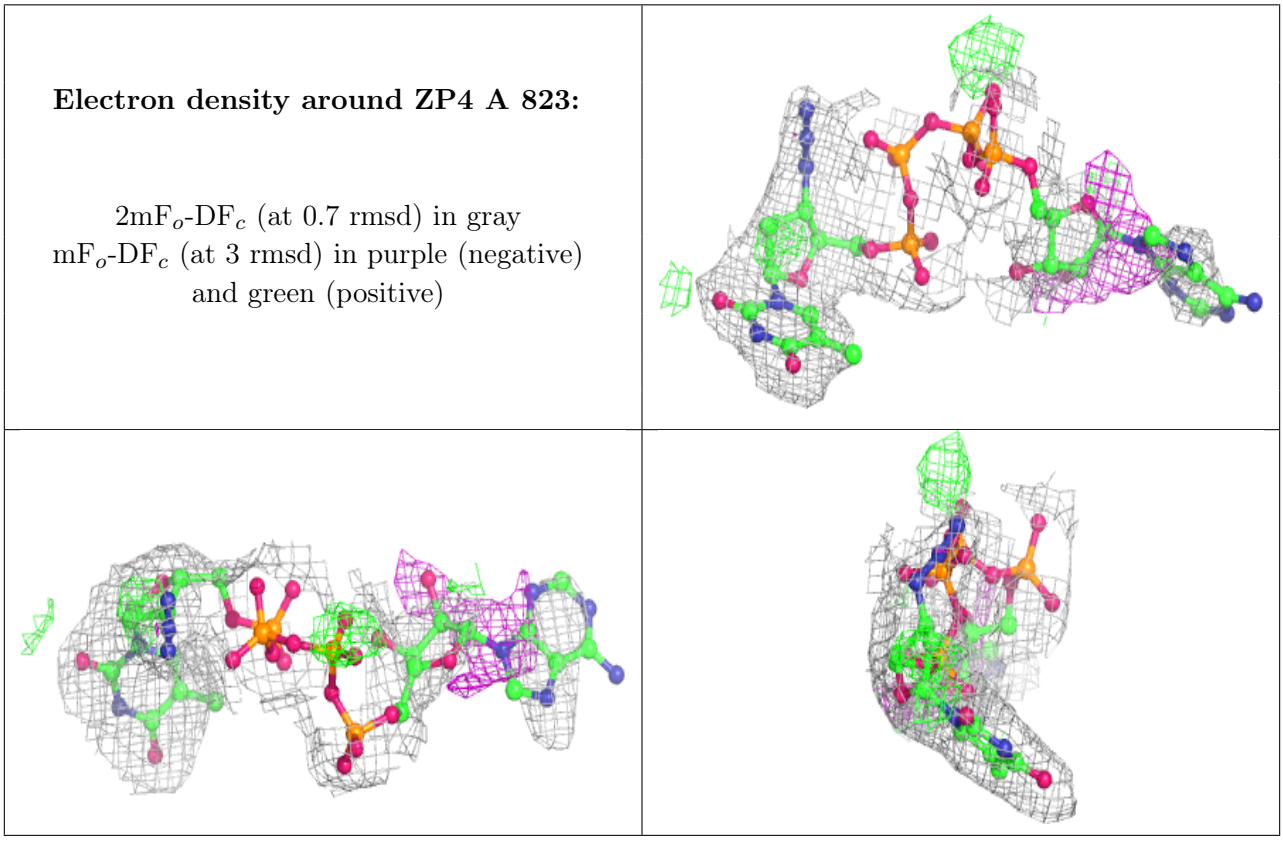
6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

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.

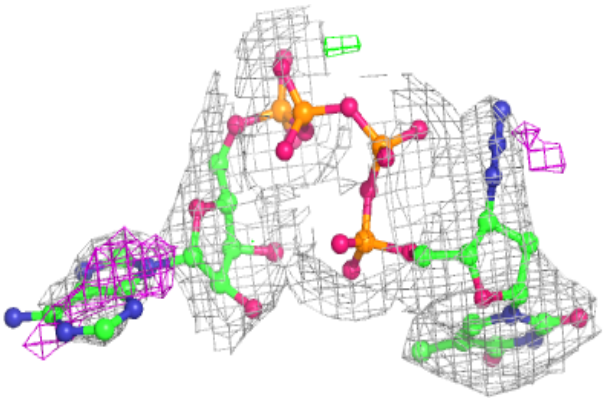
Electron density around ZP4 A 823:

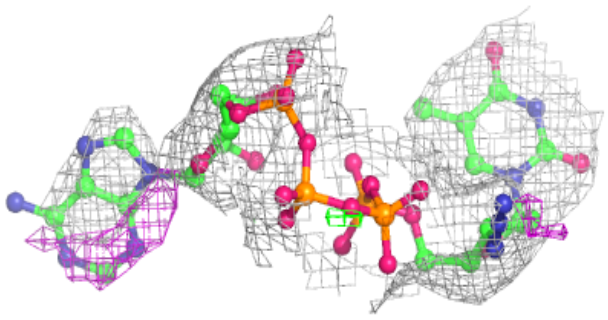
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

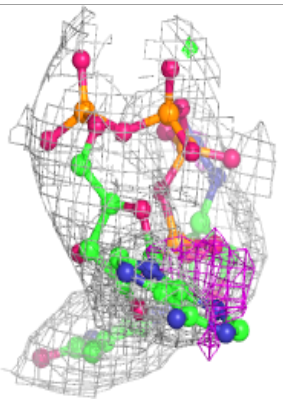


Electron density around ZP4 E 823:

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

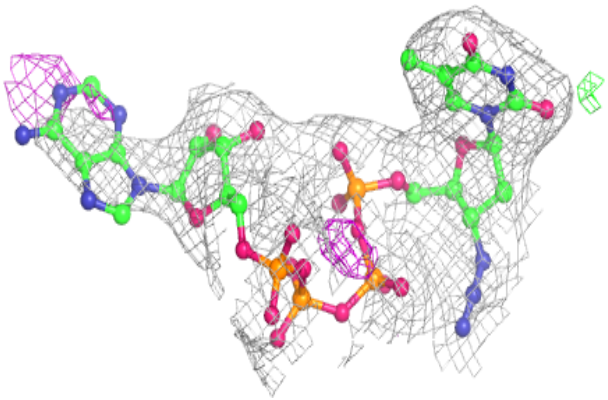


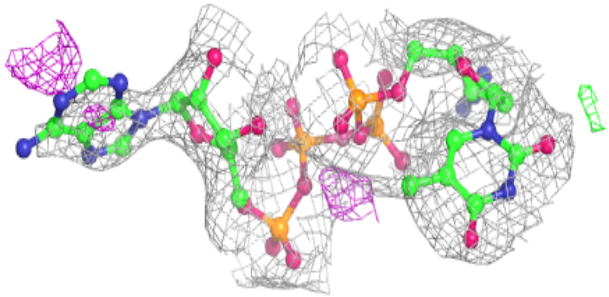


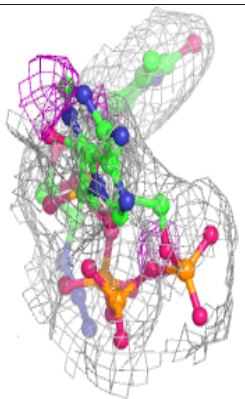


Electron density around ZP4 I 823:

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

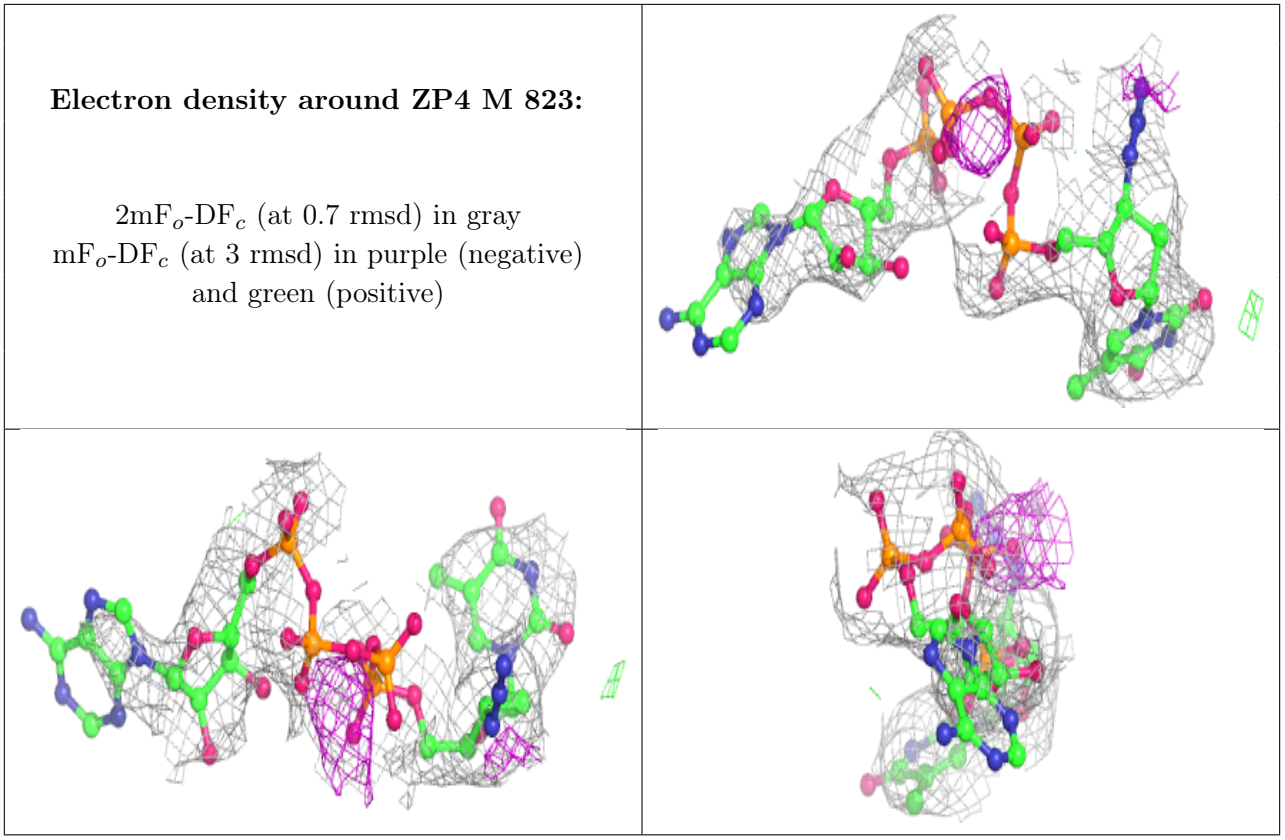






Electron density around ZP4 M 823:

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



6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.