wwPDB X-ray Structure Validation Summary Report (i)

Jul 21, 2022 – 08:47 pm BST

PDB ID	:	7ZUD
Title	:	Crystal structure of HIV-1 capsid IP6-CPSF6 complex
Authors	:	Nicastro, G.; Taylor, I.A.
Deposited on	:	2022-05-12
Resolution	:	2.93 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:

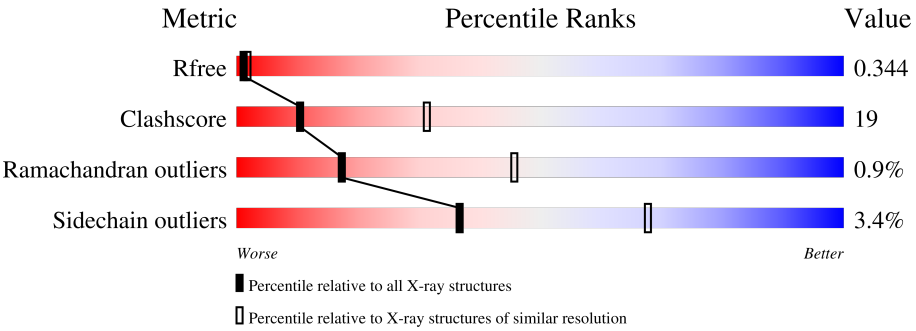
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.29
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 2.93 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	2969 (2.98-2.90)
Clashscore	141614	3218 (2.98-2.90)
Ramachandran outliers	138981	3122 (2.98-2.90)
Sidechain outliers	138945	3124 (2.98-2.90)

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	233	55% 34% • 10%
2	M	13	69% 31%

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
3	IHP	A	502[A]	-	-	X	-
3	IHP	A	502[B]	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1849 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 Capsid protein p24.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	210	Total	C	N	O	S	0	2	0
			1601	1006	279	302	14			

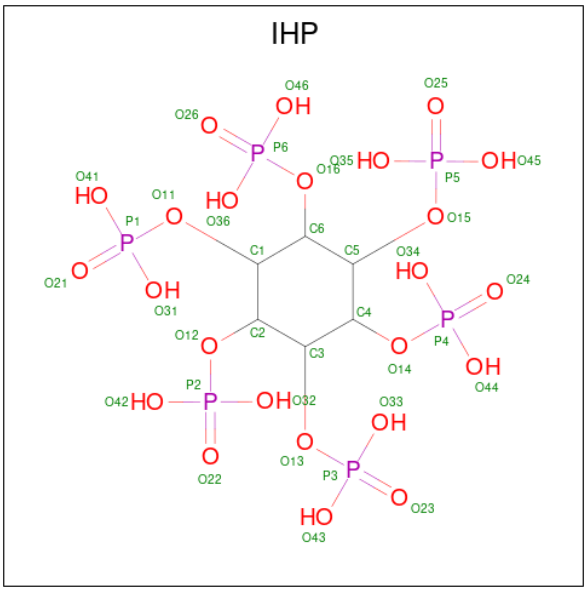
There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	14	CYS	ALA	engineered mutation	UNP P12497
A	45	CYS	GLU	engineered mutation	UNP P12497
A	184	ALA	TRP	engineered mutation	UNP P12497
A	185	ALA	MET	engineered mutation	UNP P12497
A	232	LEU	-	expression tag	UNP P12497
A	233	GLU	-	expression tag	UNP P12497

- Molecule 2 is a protein called Cleavage and polyadenylation specificity factor subunit 6.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	M	13	Total	C	N	O	0	0	0
			98	68	15	15			

- Molecule 3 is INOSITOL HEXAKISPHOSPHATE (three-letter code: IHP) (formula: $C_6H_{18}O_{24}P_6$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	O	P	0	1
			72	12	48	12		
3	A	1	Total	C	O	P	0	1
			72	12	48	12		

- Molecule 4 is water.

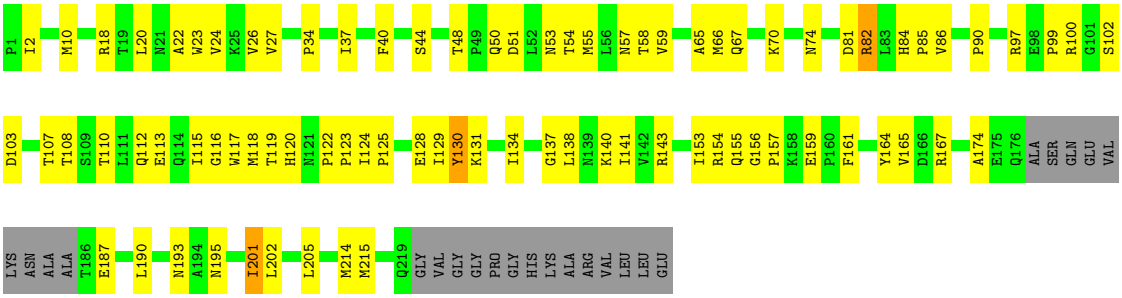
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	6	Total	O	0	0
			6	6		

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.

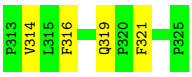
• Molecule 1: Capsid protein p24

Chain A:



• Molecule 2: Cleavage and polyadenylation specificity factor subunit 6

Chain M:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 6	Depositor
Cell constants a, b, c, α, β, γ	90.74Å 90.74Å 56.59Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	78.58 – 2.93 78.58 – 2.93	Depositor EDS
% Data completeness (in resolution range)	98.0 (78.58-2.93) 98.2 (78.58-2.93)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.24 (at 2.91Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.278 , 0.336 0.279 , 0.344	Depositor DCC
R_{free} test set	512 reflections (8.94%)	wwPDB-VP
Wilson B-factor (Å^2)	62.7	Xtriage
Anisotropy	0.096	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.39$, $< L^2 > = 0.22$	Xtriage
Estimated twinning fraction	0.388 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	1849	wwPDB-VP
Average B, all atoms (Å^2)	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.70% 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: IHP

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	0.28	0/1641	0.50	0/2232
2	M	0.29	0/104	0.41	0/143
All	All	0.28	0/1745	0.50	0/2375

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1601	0	1545	58	27
2	M	98	0	95	8	0
3	A	144	0	13	0	27
4	A	6	0	0	0	0
All	All	1849	0	1653	62	27

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 19.

The worst 5 of 62 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:A:70:LYS:O	1:A:74:ASN:ND2	2.06	0.88
1:A:84:HIS:O	1:A:100:ARG:NH1	2.09	0.86
1:A:2:ILE:HD11	1:A:115:ILE:HG22	1.61	0.82
1:A:40:PHE:O	1:A:44:SER:OG	2.03	0.77
1:A:153:ILE:O	1:A:193:ASN:ND2	2.21	0.73

The worst 5 of 27 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:18:ARG:NH1	3:A:502[B]:IHP:O26[3_675]	1.28	0.92
1:A:18:ARG:NH1	3:A:502[B]:IHP:O23[2_765]	1.30	0.90
1:A:18:ARG:NH1	3:A:502[A]:IHP:O23[3_675]	1.30	0.90
1:A:18:ARG:NE	3:A:502[A]:IHP:O21[6_655]	1.31	0.89
1:A:18:ARG:NH1	3:A:502[B]:IHP:O12[6_655]	1.31	0.89

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	208/233 (89%)	200 (96%)	6 (3%)	2 (1%)	15 43
2	M	11/13 (85%)	11 (100%)	0	0	100 100
All	All	219/246 (89%)	211 (96%)	6 (3%)	2 (1%)	17 46

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	187	GLU
1	A	90	PRO

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	168/195 (86%)	162 (96%)	6 (4%)	35 66
2	M	11/11 (100%)	11 (100%)	0	100 100
All	All	179/206 (87%)	173 (97%)	6 (3%)	37 68

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

Mol	Chain	Res	Type
1	A	85	PRO
1	A	130	TYR
1	A	201	ILE
1	A	67	GLN
1	A	10	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	139	ASN
2	M	319	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

4 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	$\#\|Z\| > 2$
3	IHP	A	501[B]	-	36,36,36	1.49	6 (16%)	54,60,60	2.49	13 (24%)
3	IHP	A	502[A]	-	36,36,36	1.60	6 (16%)	54,60,60	2.44	13 (24%)
3	IHP	A	501[A]	-	36,36,36	1.49	6 (16%)	54,60,60	2.47	13 (24%)
3	IHP	A	502[B]	-	36,36,36	1.49	6 (16%)	54,60,60	2.54	13 (24%)

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
3	IHP	A	501[B]	-	-	13/30/54/54	0/1/1/1
3	IHP	A	502[A]	-	-	9/30/54/54	0/1/1/1
3	IHP	A	501[A]	-	-	9/30/54/54	0/1/1/1
3	IHP	A	502[B]	-	-	10/30/54/54	0/1/1/1

The worst 5 of 24 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	502[A]	IHP	P5-O15	3.95	1.66	1.59
3	A	502[A]	IHP	P2-O12	3.79	1.66	1.59
3	A	502[A]	IHP	P1-O11	3.46	1.65	1.59
3	A	502[B]	IHP	P4-O14	3.39	1.65	1.59
3	A	502[A]	IHP	P6-O16	3.31	1.65	1.59

The worst 5 of 52 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	502[B]	IHP	O11-C1-C2	7.97	127.48	108.69

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	501[B]	IHP	O11-C1-C2	7.43	126.21	108.69
3	A	501[A]	IHP	O11-C1-C2	7.41	126.15	108.69
3	A	501[B]	IHP	C6-C5-C4	7.36	126.52	110.41
3	A	501[B]	IHP	C3-C2-C1	7.19	126.15	110.41

There are no chirality outliers.

5 of 41 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	501[A]	IHP	C2-C1-O11-P1
3	A	501[A]	IHP	C3-C4-O14-P4
3	A	501[A]	IHP	C5-C4-O14-P4
3	A	501[A]	IHP	C4-C5-O15-P5
3	A	501[A]	IHP	C1-C6-O16-P6

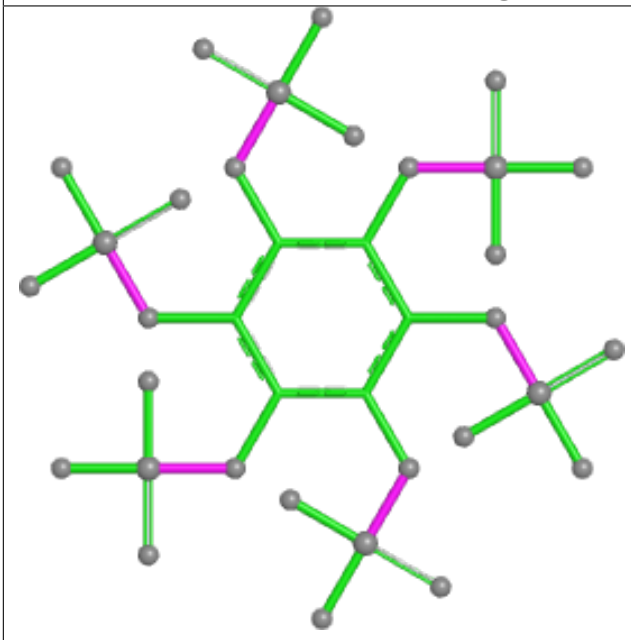
There are no ring outliers.

2 monomers are involved in 27 short contacts:

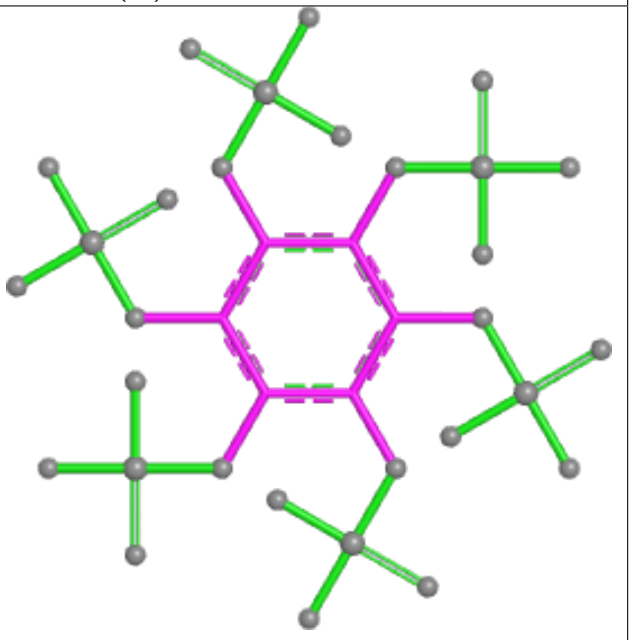
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	502[A]	IHP	0	12
3	A	502[B]	IHP	0	15

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.

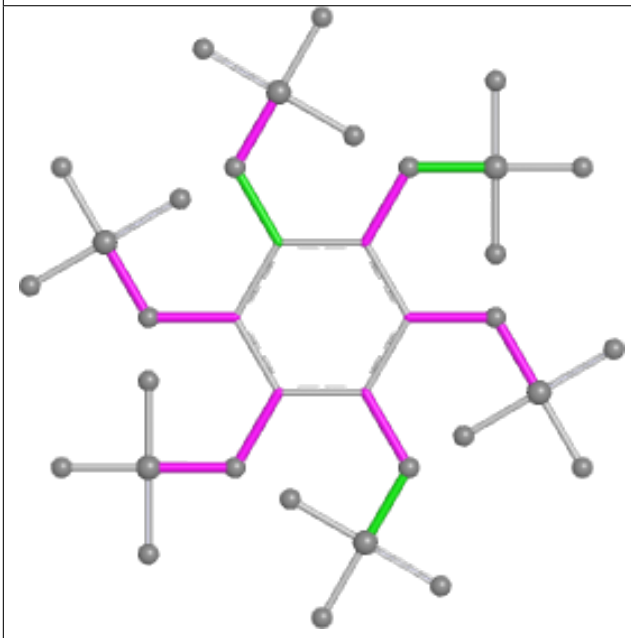
Ligand IHP A 501 (B)



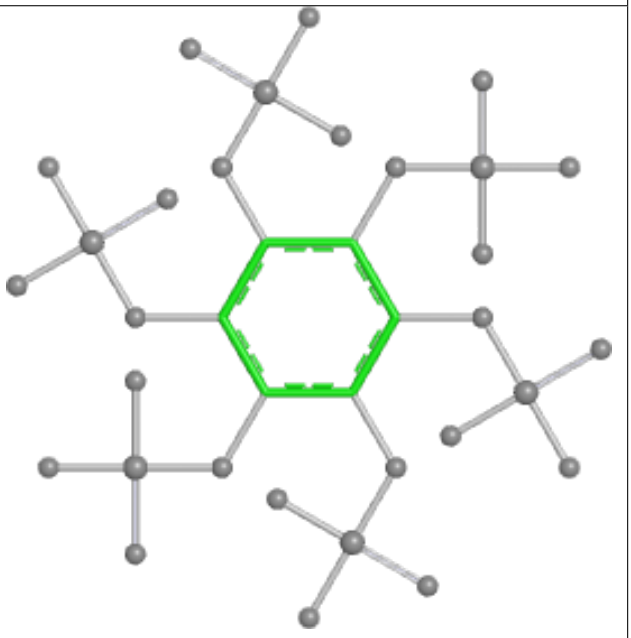
Bond lengths



Bond angles

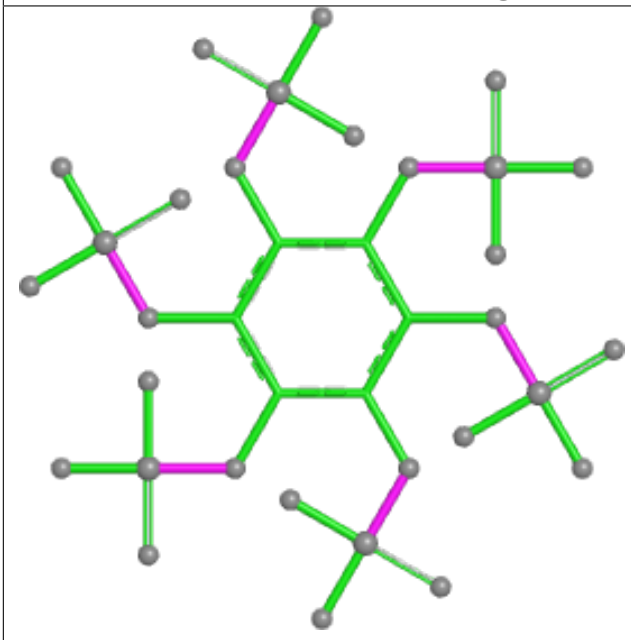


Torsions

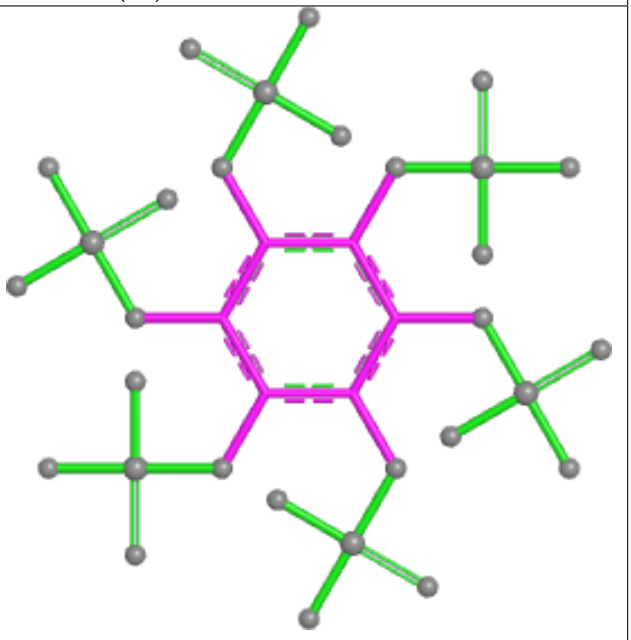


Rings

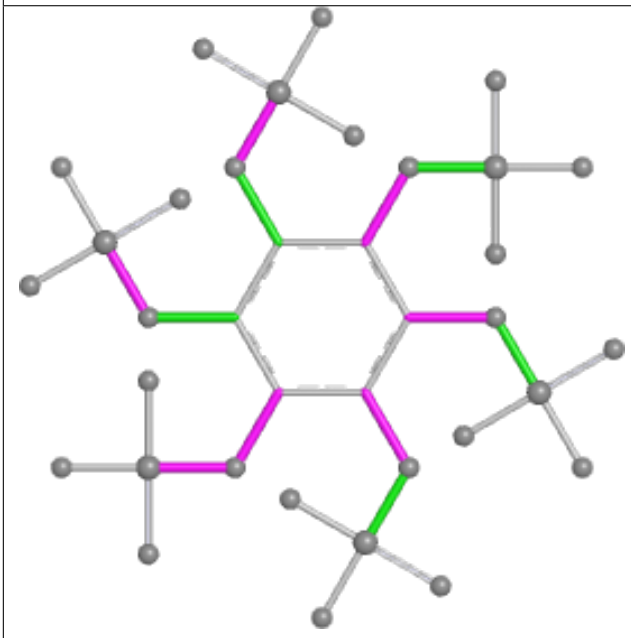
Ligand IHP A 502 (A)



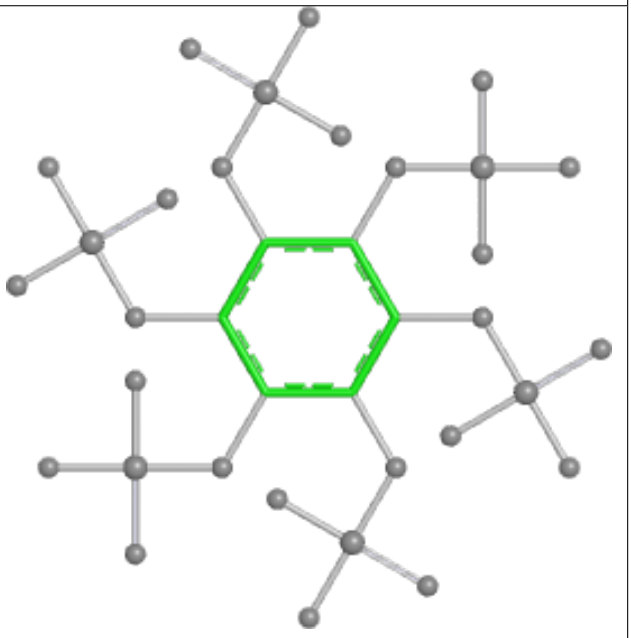
Bond lengths



Bond angles

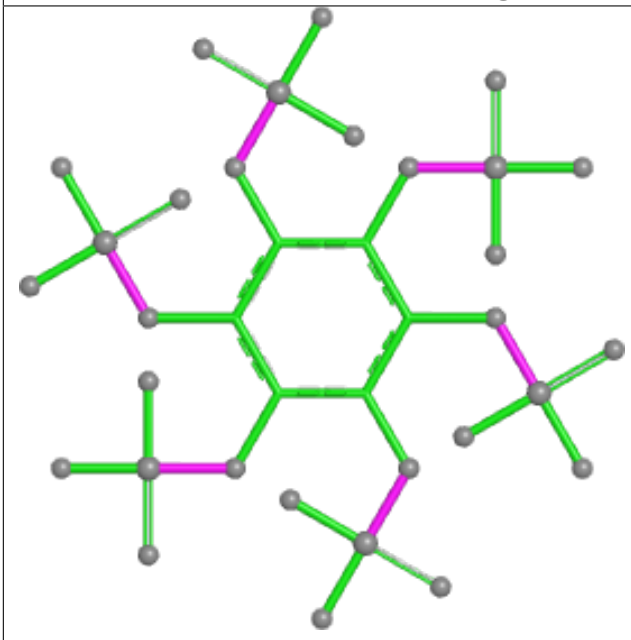


Torsions

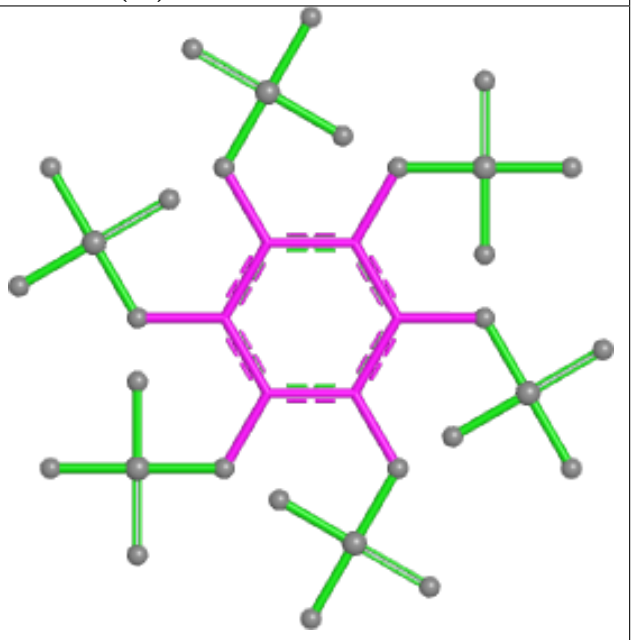


Rings

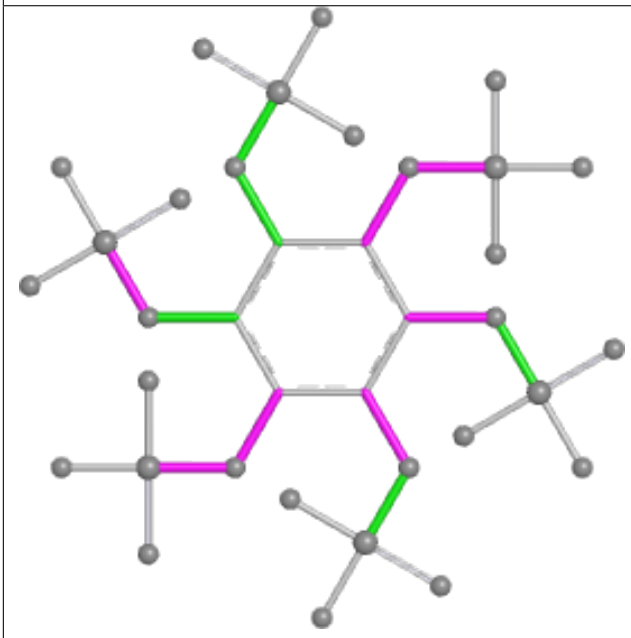
Ligand IHP A 501 (A)



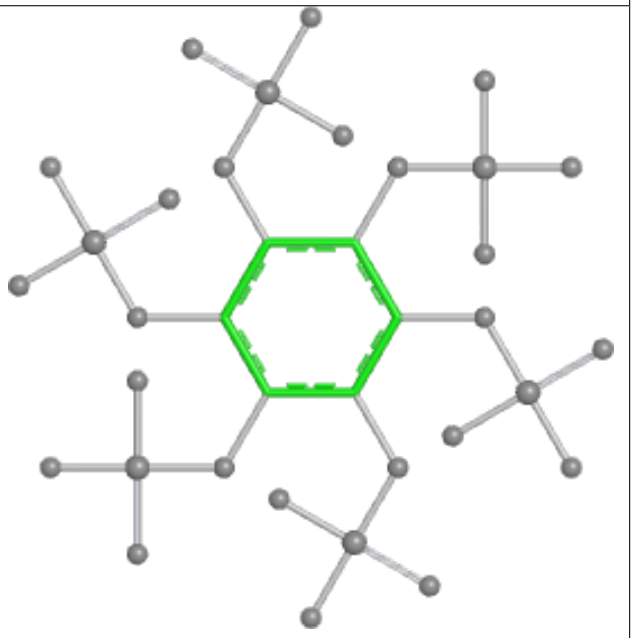
Bond lengths



Bond angles

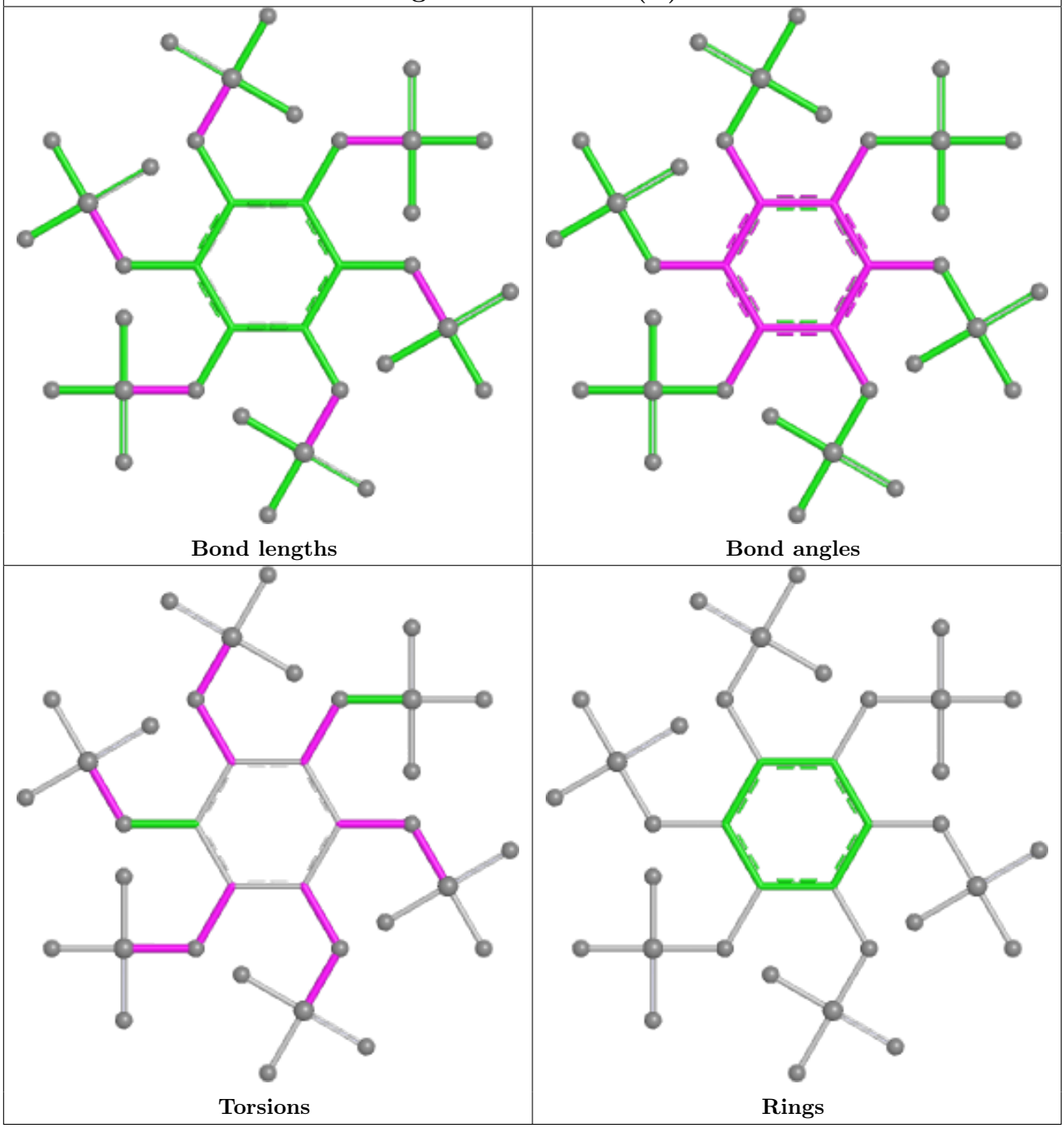


Torsions



Rings

Ligand IHP A 502 (B)



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.

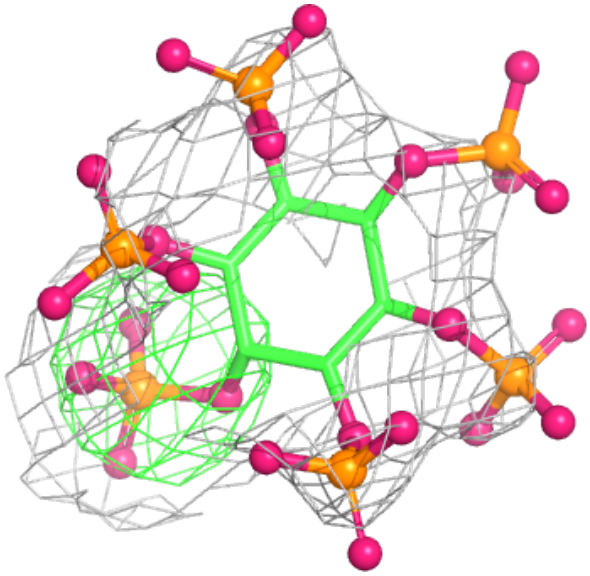
6.4 Ligands

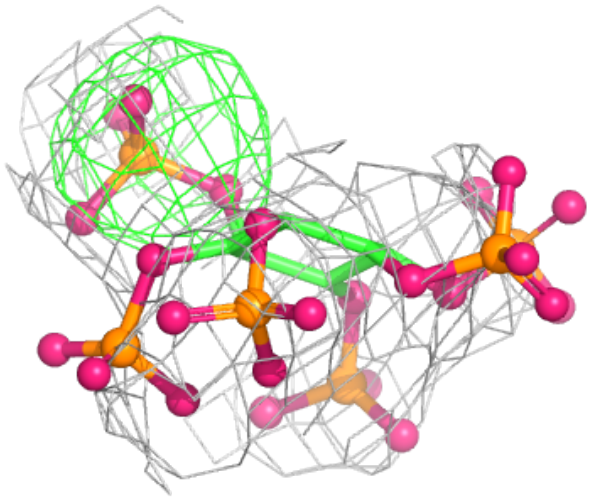
Unable to reproduce the depositors R factor - this section is therefore empty.

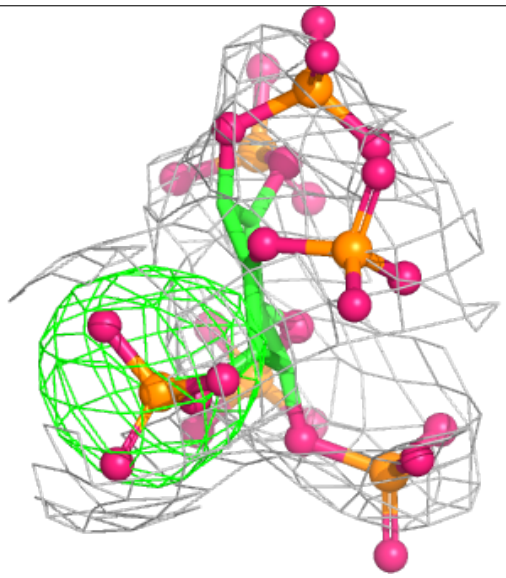
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around IHP A 501 (A):

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

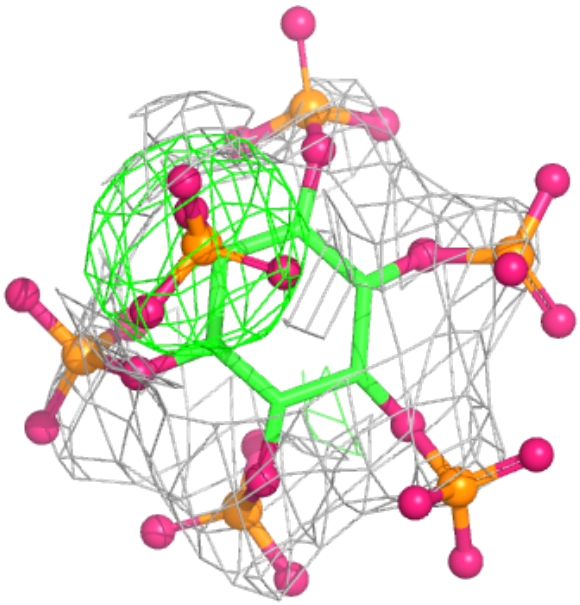


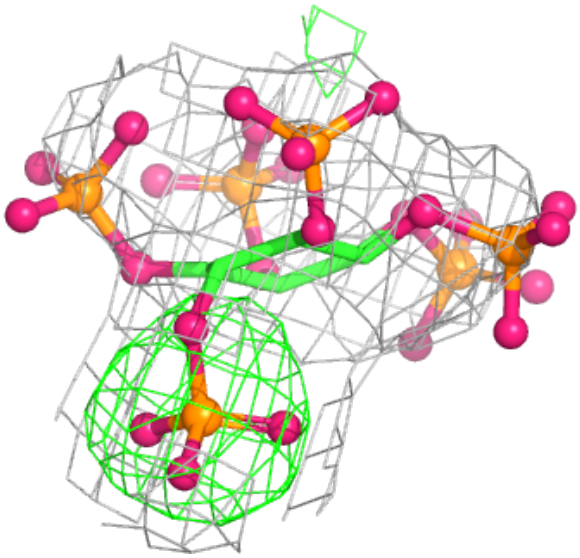


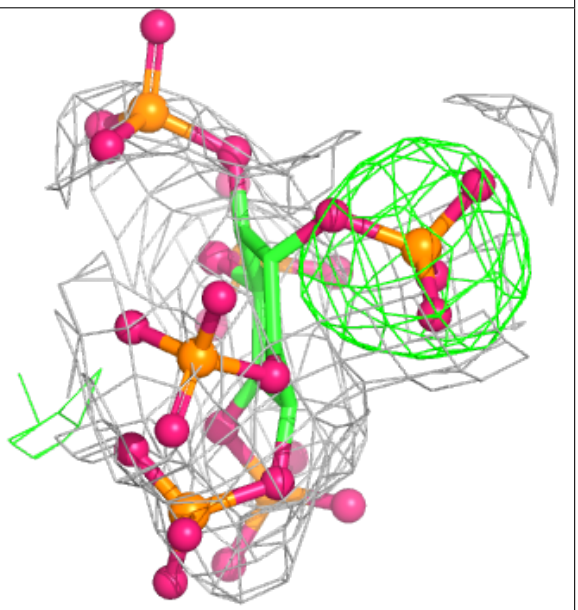


Electron density around IHP A 501 (B):

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

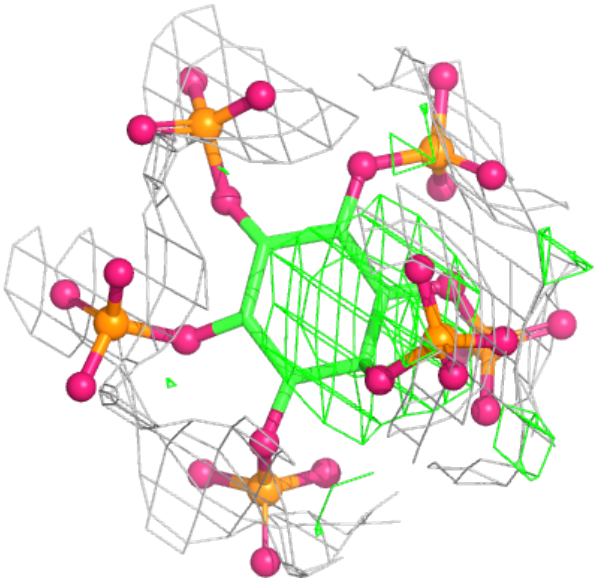


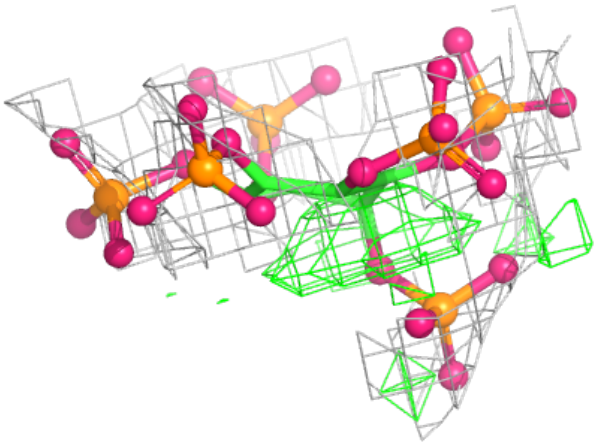


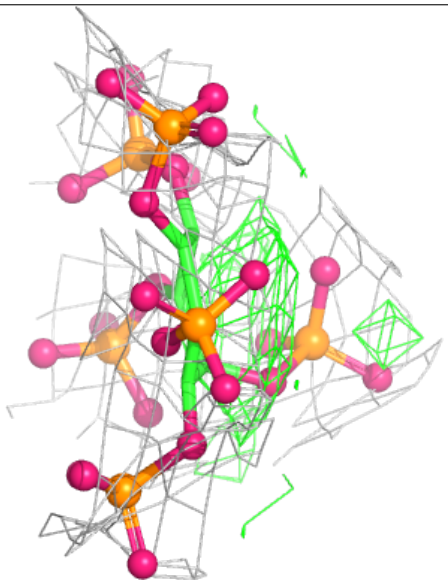


Electron density around IHP A 502 (A):

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

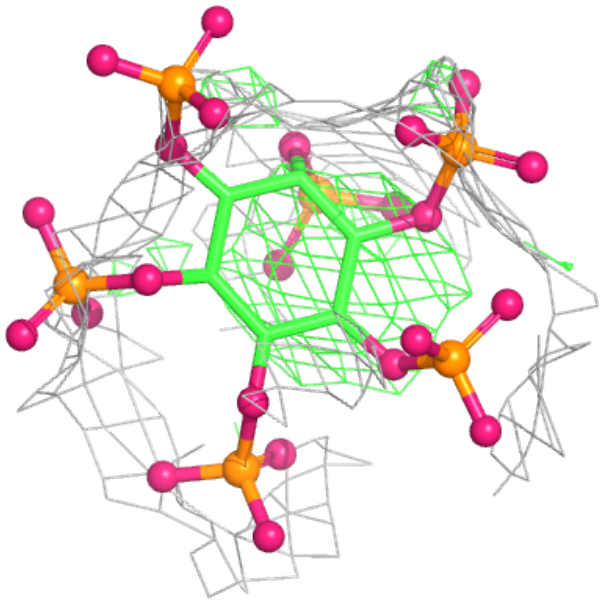


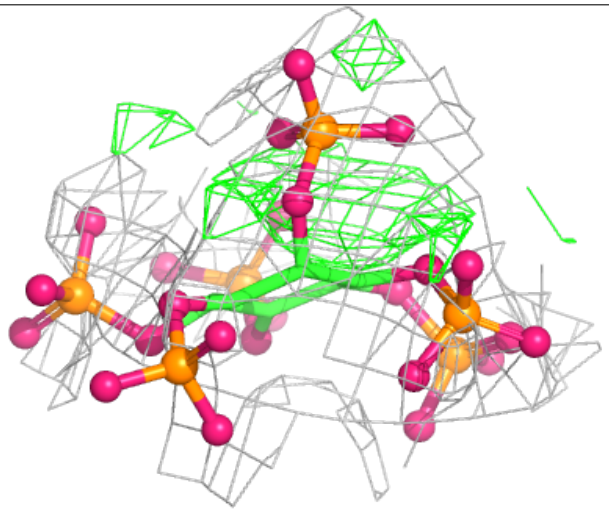


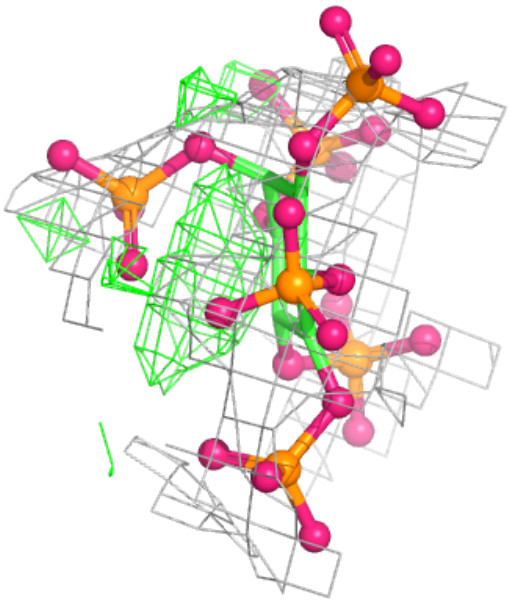


Electron density around IHP A 502 (B):

$2mF_o$-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

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