



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 20, 2020 – 09:05 PM BST

PDB ID : 3KRC  
Title : Mint heterotetrameric geranyl pyrophosphate synthase in complex with IPP  
Authors : Chang, T.-H.; Hsieh, F.-L.; Ko, T.-P.; Wang, A.H.-J.  
Deposited on : 2009-11-18  
Resolution : 2.30 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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 : 2.13.1  
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) : 2.13.1

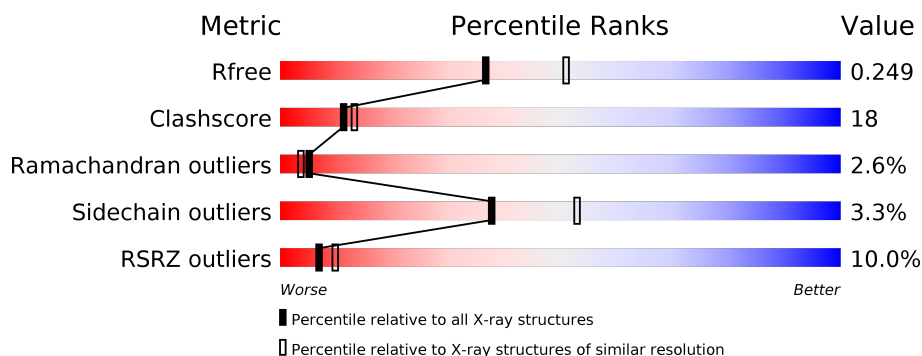
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	295	<div> <div>6%</div> <div> <div></div> <div>72%</div> <div>21%</div> <div>• 5%</div> </div> </div>
1	D	295	<div> <div>13%</div> <div> <div></div> <div>62%</div> <div>29%</div> <div>• 6%</div> </div> </div>
2	B	274	<div> <div>9%</div> <div> <div></div> <div>68%</div> <div>24%</div> <div>• 5%</div> </div> </div>
2	C	274	<div> <div>11%</div> <div> <div></div> <div>67%</div> <div>28%</div> <div>• •</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8913 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 Geranyl diphosphate synthase large subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	279	Total	C	N	O	S	0	0	0
			2120	1338	372	392	18			
1	D	277	Total	C	N	O	S	0	0	0
			2102	1326	368	390	18			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	EXPRESSION TAG	UNP Q9SBR3
D	1	MET	-	EXPRESSION TAG	UNP Q9SBR3

- Molecule 2 is a protein called Geranyl diphosphate synthase small subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	259	Total	C	N	O	S	0	0	0
			1954	1233	339	368	14			
2	C	271	Total	C	N	O	S	0	0	0
			2056	1297	361	384	14			

There are 18 discrepancies between the modelled and reference sequences:

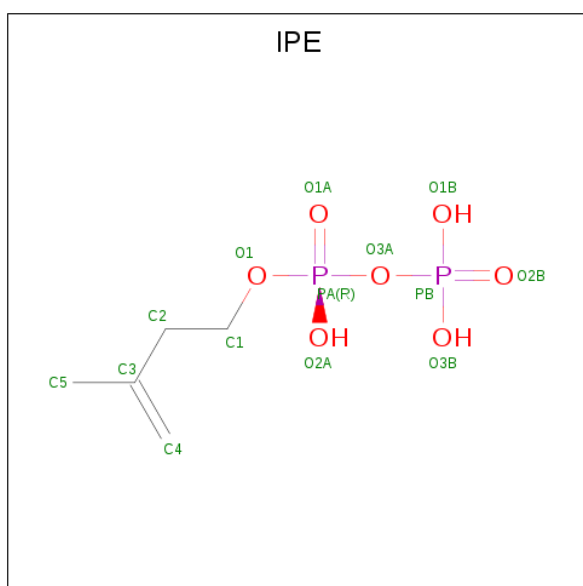
Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	EXPRESSION TAG	UNP Q9SBR4
B	267	HIS	-	EXPRESSION TAG	UNP Q9SBR4
B	268	HIS	-	EXPRESSION TAG	UNP Q9SBR4
B	269	HIS	-	EXPRESSION TAG	UNP Q9SBR4
B	270	HIS	-	EXPRESSION TAG	UNP Q9SBR4
B	271	HIS	-	EXPRESSION TAG	UNP Q9SBR4
B	272	HIS	-	EXPRESSION TAG	UNP Q9SBR4
B	273	HIS	-	EXPRESSION TAG	UNP Q9SBR4
B	274	HIS	-	EXPRESSION TAG	UNP Q9SBR4
C	1	MET	-	EXPRESSION TAG	UNP Q9SBR4

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	267	HIS	-	EXPRESSION TAG	UNP Q9SBR4
C	268	HIS	-	EXPRESSION TAG	UNP Q9SBR4
C	269	HIS	-	EXPRESSION TAG	UNP Q9SBR4
C	270	HIS	-	EXPRESSION TAG	UNP Q9SBR4
C	271	HIS	-	EXPRESSION TAG	UNP Q9SBR4
C	272	HIS	-	EXPRESSION TAG	UNP Q9SBR4
C	273	HIS	-	EXPRESSION TAG	UNP Q9SBR4
C	274	HIS	-	EXPRESSION TAG	UNP Q9SBR4

- Molecule 3 is 3-METHYLBUT-3-ENYL TRIHYDROGEN DIPHOSPHATE (three-letter code: IPE) (formula:  $C_5H_{12}O_7P_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	O	P	0	0
			14	5	7	2		
3	D	1	Total	C	O	P	0	0
			14	5	7	2		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	C	1	Total	C	O	0	0
			4	2	2		

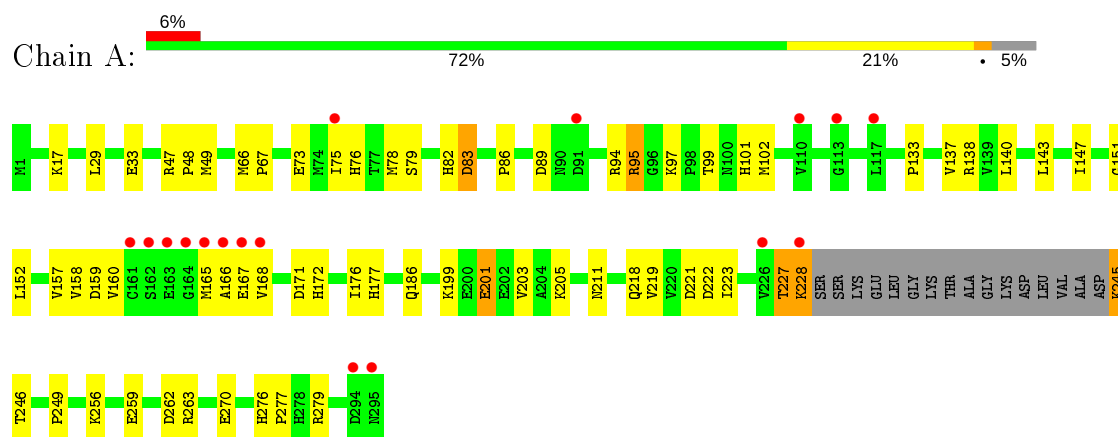
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	163	Total	O	0	0
			163	163		
5	B	144	Total	O	0	0
			144	144		
5	C	153	Total	O	0	0
			153	153		
5	D	189	Total	O	0	0
			189	189		

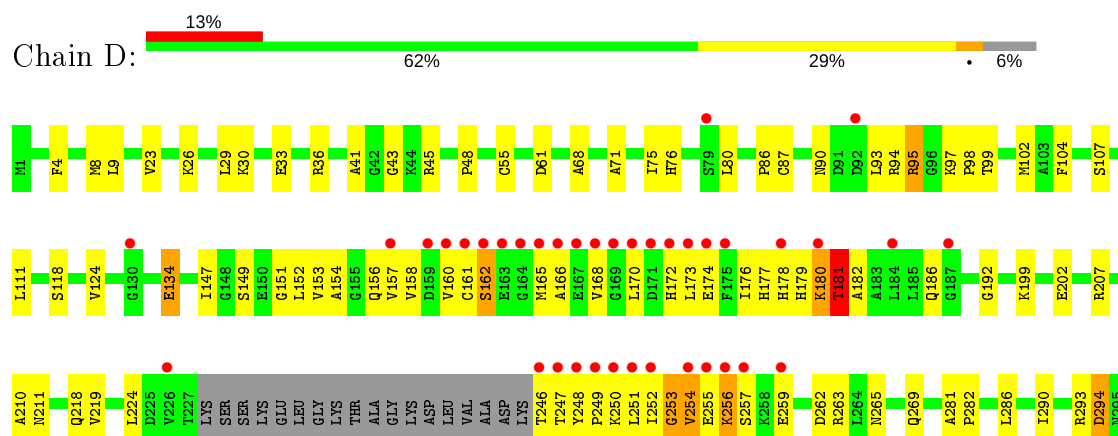
### 3 Residue-property plots [i](#)

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.

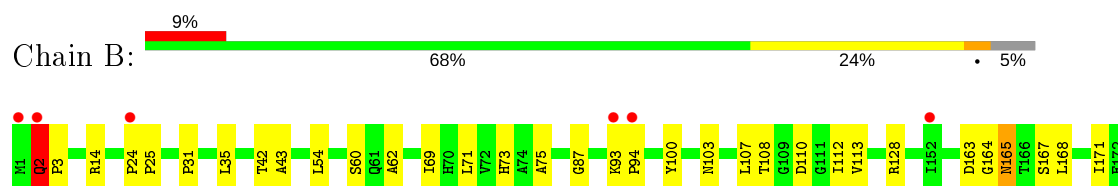
- Molecule 1: Geranyl diphosphate synthase large subunit

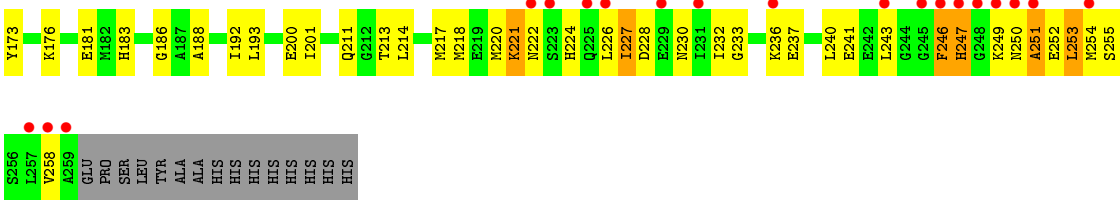


- Molecule 1: Geranyl diphosphate synthase large subunit

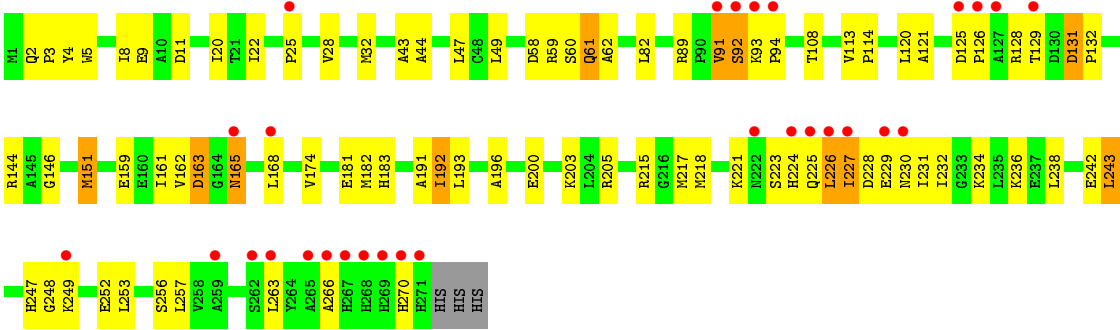


- Molecule 2: Geranyl diphosphate synthase small subunit





● Molecule 2: Geranyl diphosphate synthase small subunit



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	54.12Å 108.97Å 182.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.30 29.76 – 2.28	Depositor EDS
% Data completeness (in resolution range)	88.2 (30.00-2.30) 87.1 (29.76-2.28)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.53 (at 2.29Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.192 , 0.257 0.185 , 0.249	Depositor DCC
$R_{free}$ test set	2188 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	34.2	Xtriage
Anisotropy	0.759	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 56.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	8913	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.16% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: EDO, IPE

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	0.99	0/2155	0.93	1/2903 (0.0%)
1	D	0.88	0/2137	0.89	2/2881 (0.1%)
2	B	0.87	0/1993	0.85	0/2695
2	C	0.89	0/2102	0.85	1/2845 (0.0%)
All	All	0.91	0/8387	0.88	4/11324 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	49	MET	CG-SD-CE	5.58	109.13	100.20
2	C	144	ARG	NE-CZ-NH2	-5.47	117.56	120.30
1	D	43	GLY	N-CA-C	-5.35	99.74	113.10
1	D	61	ASP	CB-CG-OD2	-5.08	113.73	118.30

There are no chirality outliers.

There are no planarity outliers.

### 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	2120	0	2146	53	0
1	D	2102	0	2120	99	0
2	B	1954	0	1948	75	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	2056	0	2031	71	0
3	A	14	0	9	2	0
3	D	14	0	9	1	0
4	C	4	0	6	0	0
5	A	163	0	0	6	1
5	B	144	0	0	4	0
5	C	153	0	0	5	0
5	D	189	0	0	8	0
All	All	8913	0	8269	294	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 18.

The worst 5 of 294 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:C:168:LEU:HD22	2:C:231:ILE:HD12	1.43	0.98
2:B:249:LYS:HG3	2:B:253:LEU:HB3	1.43	0.96
1:A:17:LYS:HE2	5:A:355:HOH:O	1.64	0.96
1:D:177:HIS:CB	1:D:218:GLN:HG2	1.97	0.94
2:B:249:LYS:HZ3	2:B:253:LEU:HA	1.32	0.94

All (1) 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)
2:B:246:PHE:O	5:A:509:HOH:O[3_545]	2.15	0.05

## 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	275/295 (93%)	261 (95%)	12 (4%)	2 (1%)	22	26
1	D	273/295 (92%)	251 (92%)	14 (5%)	8 (3%)	4	3
2	B	257/274 (94%)	239 (93%)	12 (5%)	6 (2%)	6	5
2	C	269/274 (98%)	242 (90%)	15 (6%)	12 (4%)	2	1
All	All	1074/1138 (94%)	993 (92%)	53 (5%)	28 (3%)	5	4

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	227	THR
2	B	165	ASN
2	B	227	ILE
2	B	247	HIS
2	C	226	LEU

### 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	222/234 (95%)	214 (96%)	8 (4%)	35	49
1	D	220/234 (94%)	215 (98%)	5 (2%)	50	67
2	B	201/214 (94%)	194 (96%)	7 (4%)	36	50
2	C	211/214 (99%)	203 (96%)	8 (4%)	33	47
All	All	854/896 (95%)	826 (97%)	28 (3%)	38	53

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	220	MET
2	C	61	GLN
1	D	181	THR
2	B	246	PHE
2	B	253	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	206	ASN
2	B	211	GLN
1	D	172	HIS
2	B	183	HIS
2	B	202	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](#)

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

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
4	EDO	C	903	-	3,3,3	1.20	0	2,2,2	0.21	0
3	IPE	D	901	-	11,13,13	3.58	2 (18%)	15,19,19	1.98	5 (33%)
3	IPE	A	902	-	11,13,13	3.16	3 (27%)	15,19,19	1.93	6 (40%)

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	EDO	C	903	-	-	1/1/1/1	-
3	IPE	D	901	-	-	4/13/13/13	-
3	IPE	A	902	-	-	4/13/13/13	-

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	901	IPE	C4-C3	8.67	1.57	1.33
3	A	902	IPE	C4-C3	7.76	1.55	1.33
3	D	901	IPE	C2-C3	-7.65	1.37	1.51
3	A	902	IPE	C2-C3	-5.78	1.40	1.51
3	A	902	IPE	PA-O1A	2.67	1.60	1.50

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	901	IPE	O1B-PB-O3A	4.10	118.38	104.64
3	A	902	IPE	O1B-PB-O3A	4.04	118.17	104.64
3	D	901	IPE	O3B-PB-O3A	3.26	115.56	104.64
3	D	901	IPE	O3B-PB-O1B	-2.67	97.43	107.64
3	A	902	IPE	PA-O3A-PB	-2.60	123.91	132.83

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	D	901	IPE	O1-C1-C2-C3
3	D	901	IPE	C1-O1-PA-O3A
3	A	902	IPE	C1-O1-PA-O1A
3	A	902	IPE	C1-O1-PA-O3A
3	D	901	IPE	C1-O1-PA-O1A

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	901	IPE	1	0
3	A	902	IPE	2	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	279/295 (94%)	-0.03	17 (6%) 21 27	27, 41, 77, 108	0
1	D	277/295 (93%)	0.61	38 (13%) 3 4	27, 44, 120, 130	0
2	B	259/274 (94%)	0.28	25 (9%) 7 10	30, 47, 120, 136	0
2	C	271/274 (98%)	0.39	29 (10%) 6 8	27, 47, 102, 119	0
All	All	1086/1138 (95%)	0.31	109 (10%) 7 10	27, 44, 113, 136	0

The worst 5 of 109 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	166	ALA	11.5
1	D	167	GLU	8.4
1	D	254	VAL	8.4
1	D	165	MET	8.3
1	D	251	LEU	8.2

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

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

### 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
3	IPE	A	902	14/14	0.91	0.15	90,92,94,94	0
3	IPE	D	901	14/14	0.97	0.17	55,60,73,73	0
4	EDO	C	903	4/4	0.97	0.18	30,30,32,32	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.