W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 31, 2017 – 02:03 PM EDT

PDB ID : 3MB8
Title : Crystal structure of purine nucleoside phosphorylase from toxoplasma gondii in complex with immucillin-H
Authors : Ho, M.; Almo, S.C.; Schramm, V.L.
Deposited on : unknown
Resolution : 1.90 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030345

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 1.90 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	Percentile Ranks	Value
Rfree		0.239
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		1.0%
RSRZ outliers		2.4%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
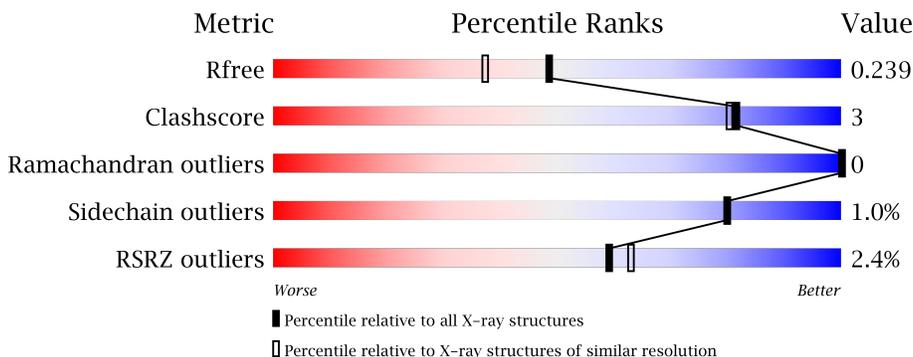
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	5047 (1.90-1.90)
Clashscore	112137	5731 (1.90-1.90)
Ramachandran outliers	110173	5669 (1.90-1.90)
Sidechain outliers	110143	5670 (1.90-1.90)
RSRZ outliers	101464	5100 (1.90-1.90)

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	279	3% / 83% / 6% / 11%
1	B	279	2% / 83% / 5% · / 11%

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

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	GOL	B	283	-	-	-	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4178 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 Purine nucleoside phosphorylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	248	Total 1881	C 1183	N 318	O 362	S 18	0	1	0
1	B	248	Total 1885	C 1185	N 319	O 363	S 18	0	1	0

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	EXPRESSION TAG	UNP Q2HXR2
A	2	PRO	-	EXPRESSION TAG	UNP Q2HXR2
A	3	LEU	-	EXPRESSION TAG	UNP Q2HXR2
A	236	VAL	ILE	SEE REMARK 999	UNP Q2HXR2
A	251	LYS	-	EXPRESSION TAG	UNP Q2HXR2
A	252	GLY	-	EXPRESSION TAG	UNP Q2HXR2
A	253	GLU	-	EXPRESSION TAG	UNP Q2HXR2
A	254	PHE	-	EXPRESSION TAG	UNP Q2HXR2
A	255	GLU	-	EXPRESSION TAG	UNP Q2HXR2
A	256	ALA	-	EXPRESSION TAG	UNP Q2HXR2
A	257	TYR	-	EXPRESSION TAG	UNP Q2HXR2
A	258	VAL	-	EXPRESSION TAG	UNP Q2HXR2
A	259	GLU	-	EXPRESSION TAG	UNP Q2HXR2
A	260	GLN	-	EXPRESSION TAG	UNP Q2HXR2
A	261	LYS	-	EXPRESSION TAG	UNP Q2HXR2
A	262	LEU	-	EXPRESSION TAG	UNP Q2HXR2
A	263	ILE	-	EXPRESSION TAG	UNP Q2HXR2
A	264	SER	-	EXPRESSION TAG	UNP Q2HXR2
A	265	GLU	-	EXPRESSION TAG	UNP Q2HXR2
A	266	GLU	-	EXPRESSION TAG	UNP Q2HXR2
A	267	ASP	-	EXPRESSION TAG	UNP Q2HXR2
A	268	LEU	-	EXPRESSION TAG	UNP Q2HXR2
A	269	ASN	-	EXPRESSION TAG	UNP Q2HXR2
A	270	SER	-	EXPRESSION TAG	UNP Q2HXR2
A	271	ALA	-	EXPRESSION TAG	UNP Q2HXR2

Continued on next page...

Continued from previous page...

| Chain	Residue	Modelled	Actual	Comment	Reference
A	272	VAL	-	EXPRESSION TAG	UNP Q2HXR2
A	273	ASP	-	EXPRESSION TAG	UNP Q2HXR2
A	274	HIS	-	EXPRESSION TAG	UNP Q2HXR2
A	275	HIS	-	EXPRESSION TAG	UNP Q2HXR2
A	276	HIS	-	EXPRESSION TAG	UNP Q2HXR2
A	277	HIS	-	EXPRESSION TAG	UNP Q2HXR2
A	278	HIS	-	EXPRESSION TAG	UNP Q2HXR2
A	279	HIS	-	EXPRESSION TAG	UNP Q2HXR2
B	1	MET	-	EXPRESSION TAG	UNP Q2HXR2
B	2	PRO	-	EXPRESSION TAG	UNP Q2HXR2
B	3	LEU	-	EXPRESSION TAG	UNP Q2HXR2
B	236	VAL	ILE	SEE REMARK 999	UNP Q2HXR2
B	251	LYS	-	EXPRESSION TAG	UNP Q2HXR2
B	252	GLY	-	EXPRESSION TAG	UNP Q2HXR2
B	253	GLU	-	EXPRESSION TAG	UNP Q2HXR2
B	254	PHE	-	EXPRESSION TAG	UNP Q2HXR2
B	255	GLU	-	EXPRESSION TAG	UNP Q2HXR2
B	256	ALA	-	EXPRESSION TAG	UNP Q2HXR2
B	257	TYR	-	EXPRESSION TAG	UNP Q2HXR2
B	258	VAL	-	EXPRESSION TAG	UNP Q2HXR2
B	259	GLU	-	EXPRESSION TAG	UNP Q2HXR2
B	260	GLN	-	EXPRESSION TAG	UNP Q2HXR2
B	261	LYS	-	EXPRESSION TAG	UNP Q2HXR2
B	262	LEU	-	EXPRESSION TAG	UNP Q2HXR2
B	263	ILE	-	EXPRESSION TAG	UNP Q2HXR2
B	264	SER	-	EXPRESSION TAG	UNP Q2HXR2
B	265	GLU	-	EXPRESSION TAG	UNP Q2HXR2
B	266	GLU	-	EXPRESSION TAG	UNP Q2HXR2
B	267	ASP	-	EXPRESSION TAG	UNP Q2HXR2
B	268	LEU	-	EXPRESSION TAG	UNP Q2HXR2
B	269	ASN	-	EXPRESSION TAG	UNP Q2HXR2
B	270	SER	-	EXPRESSION TAG	UNP Q2HXR2
B	271	ALA	-	EXPRESSION TAG	UNP Q2HXR2
B	272	VAL	-	EXPRESSION TAG	UNP Q2HXR2
B	273	ASP	-	EXPRESSION TAG	UNP Q2HXR2
B	274	HIS	-	EXPRESSION TAG	UNP Q2HXR2
B	275	HIS	-	EXPRESSION TAG	UNP Q2HXR2
B	276	HIS	-	EXPRESSION TAG	UNP Q2HXR2
B	277	HIS	-	EXPRESSION TAG	UNP Q2HXR2
B	278	HIS	-	EXPRESSION TAG	UNP Q2HXR2
B	279	HIS	-	EXPRESSION TAG	UNP Q2HXR2

- Molecule 2 is 1,4-DIDEOXY-4-AZA-1-(S)-(9-DEAZAHYPOXANTHIN-9-YL)-D-RIBITOL

(three-letter code: IMH) (formula: $C_{11}H_{14}N_4O_4$).
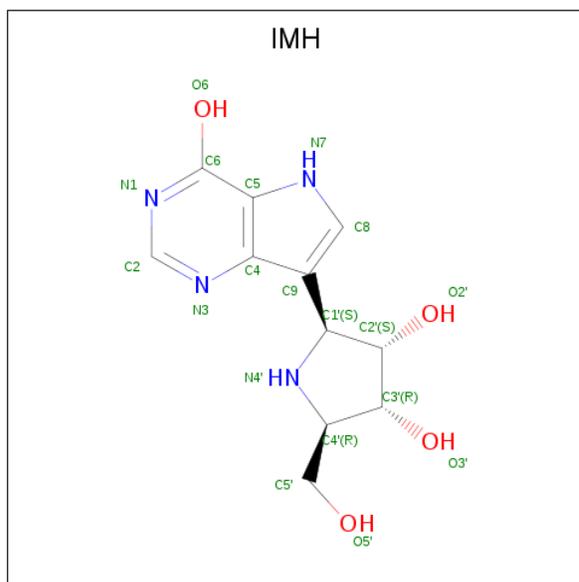
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			19	11	4	4		
2	B	1	Total	C	N	O	0	0
			19	11	4	4		

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).
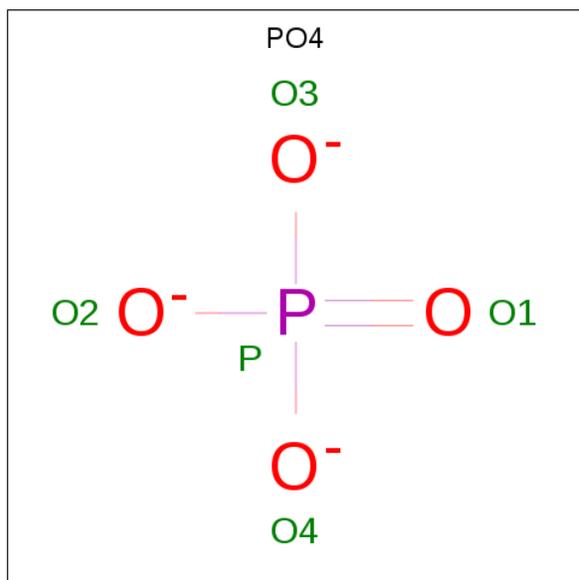
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	P	0	0
			5	4	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total 5	O 4	P 1	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).
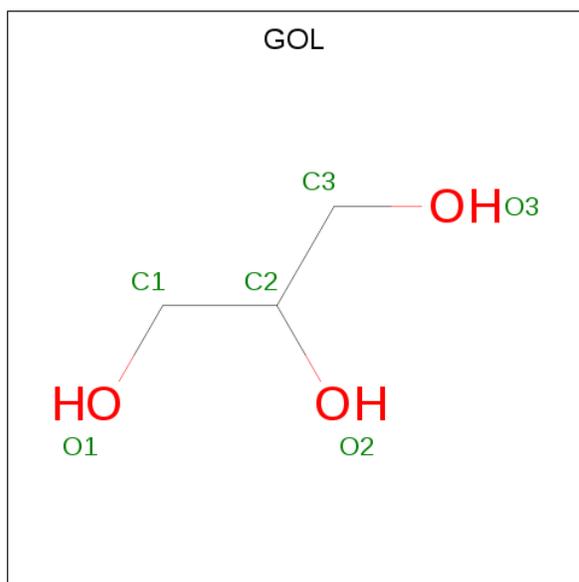
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 6	C 3	O 3	0	0
4	A	1	Total 6	C 3	O 3	0	0
4	B	1	Total 6	C 3	O 3	0	0
4	B	1	Total 6	C 3	O 3	0	0

- Molecule 5 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	171	Total 171	O 171	0	0
5	B	169	Total 169	O 169	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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: Purine nucleoside phosphorylase
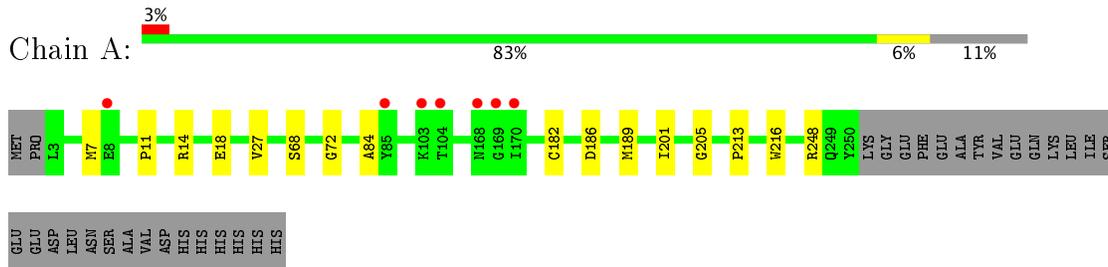
- Molecule 1: Purine nucleoside phosphorylase
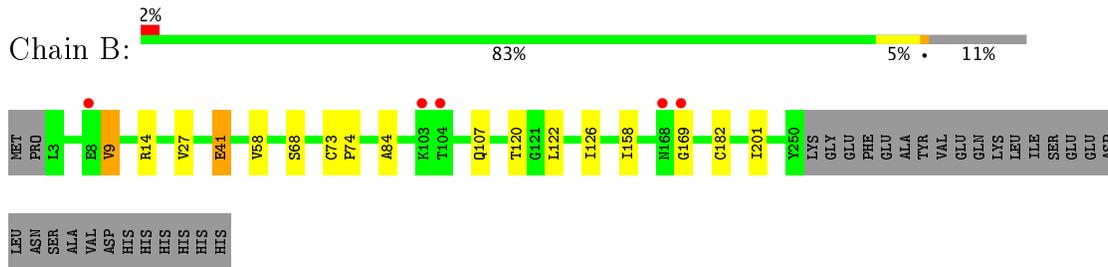
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 6	Depositor
Cell constants a, b, c, α, β, γ	159.62Å 159.62Å 53.61Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 1.90 19.75 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.1 (20.00-1.90) 99.3 (19.75-1.90)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	3.08 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.221 , 0.243 0.218 , 0.239	Depositor DCC
R_{free} test set	3112 reflections (5.34%)	DCC
Wilson B-factor (Å^2)	18.6	Xtriage
Anisotropy	0.074	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 52.8	EDS
L-test for twinning[2]	$<|L|> = 0.48, <L^2> = 0.31$	Xtriage
Estimated twinning fraction	0.410 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	4178	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.83% 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: GOL, PO4, IMH

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	1/1910 (0.1%)	0.61	0/2584
1	B	0.55	0/1914	0.62	1/2590 (0.0%)
All	All	0.56	1/3824 (0.0%)	0.61	1/5174 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	182	CYS	CB-SG	-7.78	1.69	1.82

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	169	GLY	N-CA-C	5.10	125.85	113.10

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	1881	0	1899	13	0
1	B	1885	0	1904	9	0
2	A	19	0	13	1	0
2	B	19	0	13	1	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	5	0	0	0	0
3	B	5	0	0	0	0
4	A	12	0	16	3	0
4	B	12	0	16	0	0
5	A	171	0	0	4	0
5	B	169	0	0	1	0
All	All	4178	0	3861	23	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 3.

The worst 5 of 23 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:7:MET:CE	5:A:415:HOH:O	2.46	0.62
1:B:9:VAL:HG13	1:B:14:ARG:HA	1.84	0.59
1:A:7:MET:HE2	5:A:415:HOH:O	2.08	0.54
1:A:7:MET:HE1	5:A:415:HOH:O	2.06	0.53
1:A:14:ARG:O	4:A:282:GOL:H31	2.09	0.52

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	247/279 (88%)	244 (99%)	3 (1%)	0	100	100
1	B	247/279 (88%)	244 (99%)	3 (1%)	0	100	100
All	All	494/558 (88%)	488 (99%)	6 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	203/231 (88%)	201 (99%)	2 (1%)	80	80
1	B	204/231 (88%)	202 (99%)	2 (1%)	80	80
All	All	407/462 (88%)	403 (99%)	4 (1%)	80	80

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	186	ASP
1	A	248	ARG
1	B	9	VAL
1	B	41	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	5	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

8 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	#$|Z| > 2$
2	IMH	A	280	-	19,21,21	1.40	2 (10%)	15,31,31	2.40	5 (33%)
3	PO4	A	281	-	4,4,4	0.84	0	6,6,6	0.67	0
4	GOL	A	282	-	5,5,5	0.32	0	5,5,5	0.32	0
4	GOL	A	283	-	5,5,5	0.44	0	5,5,5	0.61	0
2	IMH	B	280	-	19,21,21	1.34	2 (10%)	15,31,31	2.29	5 (33%)
3	PO4	B	281	-	4,4,4	0.86	0	6,6,6	0.69	0
4	GOL	B	282	-	5,5,5	0.31	0	5,5,5	0.30	0
4	GOL	B	283	-	5,5,5	0.48	0	5,5,5	0.55	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
2	IMH	A	280	-	-	0/2/22/22	0/3/3/3
3	PO4	A	281	-	-	0/0/0/0	0/0/0/0
4	GOL	A	282	-	-	0/4/4/4	0/0/0/0
4	GOL	A	283	-	-	0/4/4/4	0/0/0/0
2	IMH	B	280	-	-	0/2/22/22	0/3/3/3
3	PO4	B	281	-	-	0/0/0/0	0/0/0/0
4	GOL	B	282	-	-	0/4/4/4	0/0/0/0
4	GOL	B	283	-	-	0/4/4/4	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	280	IMH	C9-C1'	-4.32	1.49	1.53
2	B	280	IMH	C9-C1'	-3.83	1.49	1.53
2	A	280	IMH	C3'-C4'	-2.10	1.51	1.53
2	B	280	IMH	C3'-C4'	-2.04	1.51	1.53

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	280	IMH	N3-C2-N1	-6.79	122.95	128.86
2	B	280	IMH	N3-C2-N1	-6.00	123.63	128.86
2	B	280	IMH	C5'-C4'-C3'	-3.13	108.06	113.67
2	B	280	IMH	C9-C4-C5	-2.99	102.21	116.31
2	A	280	IMH	C9-C4-C5	-2.98	102.26	116.31

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	280	IMH	1	0
4	A	282	GOL	1	0
4	A	283	GOL	2	0
2	B	280	IMH	1	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, 95th 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	248/279 (88%)	-0.14	7 (2%)	53	57	11, 16, 28, 37	0
1	B	248/279 (88%)	-0.14	5 (2%)	65	69	11, 16, 29, 37	0
All	All	496/558 (88%)	-0.14	12 (2%)	59	63	11, 16, 29, 37	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	169	GLY	7.2
1	B	169	GLY	5.5
1	A	170	ILE	3.2
1	B	168	ASN	3.1
1	A	8	GLU	3.0

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

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

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
4	GOL	B	283	6/6	0.86	0.14	4.39	28,32,33,35	0
4	GOL	B	282	6/6	0.71	0.20	1.91	46,47,48,48	0
4	GOL	A	283	6/6	0.92	0.11	1.80	24,29,30,32	0
4	GOL	A	282	6/6	0.69	0.17	1.63	47,47,47,47	0
3	PO4	A	281	5/5	0.99	0.08	0.27	13,13,14,15	0
2	IMH	A	280	19/19	0.97	0.09	0.06	8,11,13,13	0
2	IMH	B	280	19/19	0.97	0.08	-0.03	9,11,14,14	0
3	PO4	B	281	5/5	0.99	0.07	-0.58	11,12,15,15	0

6.5 Other polymers ⓘ

There are no such residues in this entry.