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

wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 12:52 pm BST

PDB ID : 1K75
Title : The L-histidinol dehydrogenase (hisD) structure implicates domain swapping and gene duplication.
Authors : Barbosa, J.A.R.G.; Sivaraman, J.; Li, Y.; Larocque, R.; Matte, A.; Schrag, J.; Cygler, M.; Montreal-Kingston Bacterial Structural Genomics Initiative (BSGI)
Deposited on : 2001-10-18
Resolution : 1.75 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
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)
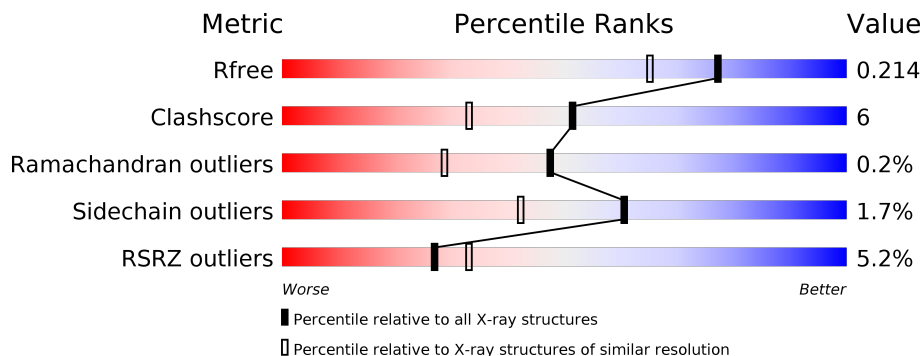Validation Pipeline (wwPDB-VP) : 2.11

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 1.75 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		0.214
Clashscore		6
Ramachandran outliers		0.2%
Sidechain outliers		1.7%
RSRZ outliers		5.2%

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}	130704	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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 respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	434	7% / 87% / 11% ••
1	B	434	3% / 91% / 7% ••

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7382 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 L-histidinol dehydrogenase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	431	Total 3221	C 2004	N 562	O 642	S 7	Se 6	47	4	0
1	B	431	Total 3211	C 2000	N 559	O 639	S 7	Se 6	22	1	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	15	GLU	VAL	CONFLICT	UNP P06988
A	22	MSE	MET	MODIFIED RESIDUE	UNP P06988
A	88	MSE	MET	MODIFIED RESIDUE	UNP P06988
A	144	MSE	MET	MODIFIED RESIDUE	UNP P06988
A	150	SER	ARG	CONFLICT	UNP P06988
A	232	MSE	MET	MODIFIED RESIDUE	UNP P06988
A	277	MSE	MET	MODIFIED RESIDUE	UNP P06988
A	313	LEU	SER	CONFLICT	UNP P06988
A	390	MSE	MET	MODIFIED RESIDUE	UNP P06988
A	403	LEU	VAL	CONFLICT	UNP P06988
B	15	GLU	VAL	CONFLICT	UNP P06988
B	22	MSE	MET	MODIFIED RESIDUE	UNP P06988
B	88	MSE	MET	MODIFIED RESIDUE	UNP P06988
B	144	MSE	MET	MODIFIED RESIDUE	UNP P06988
B	150	SER	ARG	CONFLICT	UNP P06988
B	232	MSE	MET	MODIFIED RESIDUE	UNP P06988
B	277	MSE	MET	MODIFIED RESIDUE	UNP P06988
B	313	LEU	SER	CONFLICT	UNP P06988
B	390	MSE	MET	MODIFIED RESIDUE	UNP P06988
B	403	LEU	VAL	CONFLICT	UNP P06988

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
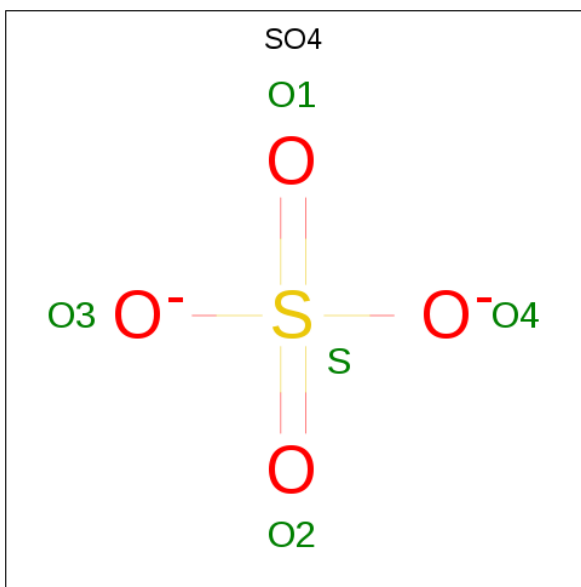
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 5	O 4	S 1	0	0
2	A	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0

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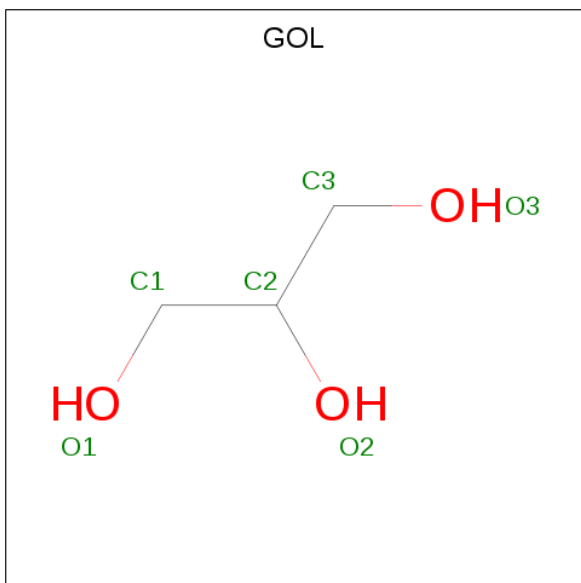
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 6	C 3	O 3	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	443	Total 443	O 443	0	0
4	B	481	Total 481	O 481	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: L-histidinol dehydrogenase
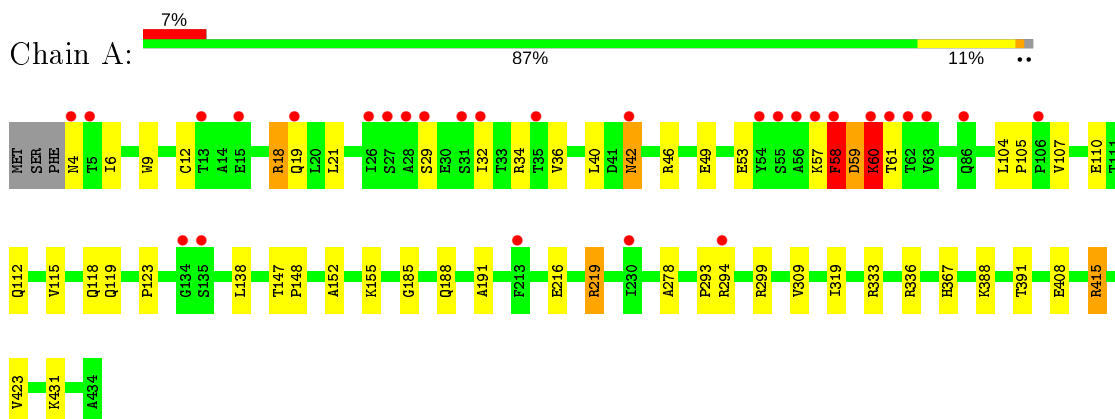
- Molecule 1: L-histidinol dehydrogenase
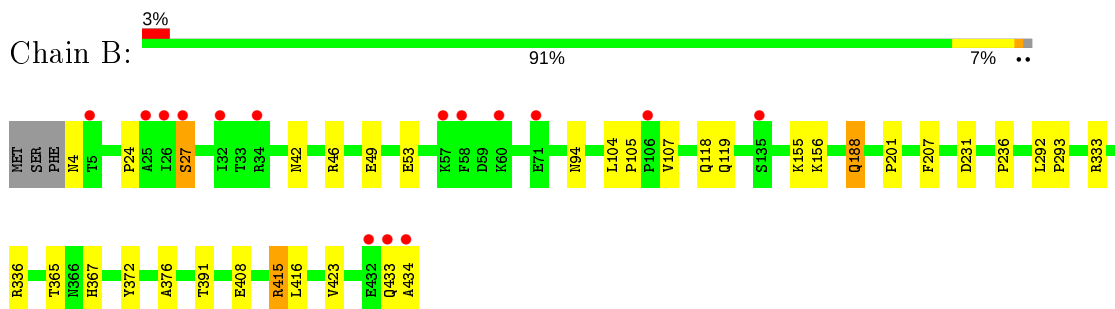
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	54.37Å 107.52Å 157.20Å	Depositor
a, b, c, α, β, γ	90.00° 90.00° 90.00°	
Resolution (Å)	40.00 – 1.75	Depositor
	38.23 – 1.75	EDS
% Data completeness	96.9 (40.00-1.75)	Depositor
(in resolution range)	94.5 (38.23-1.75)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$<I/\sigma(I)>$ [1]	2.26 (at 1.75Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.190 , 0.225	Depositor
	0.181 , 0.214	DCC
R_{free} test set	2290 reflections (2.51%)	wwPDB-VP
Wilson B-factor (Å^2)	17.4	Xtriage
Anisotropy	0.667	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 49.6	EDS
L-test for twinning[2]	$<|L|> = 0.49$, $<L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7382	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% 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, SO4

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	1.23	7/3282 (0.2%)	1.02	12/4454 (0.3%)
1	B	1.57	2/3258 (0.1%)	0.97	2/4424 (0.0%)
All	All	1.41	9/6540 (0.1%)	1.00	14/8878 (0.2%)

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
1	A	0	2
1	B	0	2
All	All	0	4

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	27	SER	CB-OG	70.74	2.34	1.42
1	B	27	SER	C-N	-42.99	0.35	1.34
1	A	61	THR	CB-OG1	39.32	2.21	1.43
1	A	59	ASP	C-N	-29.59	0.66	1.34
1	A	34	ARG	CB-CG	-23.36	0.89	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	27	SER	O-C-N	-39.55	59.41	122.70
1	A	60	LYS	CB-CG-CD	-23.36	50.87	111.60
1	A	58	PHE	O-C-N	-22.34	86.95	122.70
1	A	59	ASP	O-C-N	-21.19	88.80	122.70

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	27	SER	CA-CB-OG	-13.67	74.28	111.20

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	58	PHE	Mainchain
1	A	59	ASP	Mainchain
1	B	27	SER	Mainchain,Peptide

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	3221	0	3237	48	0
1	B	3211	0	3228	36	0
2	A	10	0	0	0	0
2	B	10	0	0	1	0
3	A	6	0	8	1	0
4	A	443	0	0	3	0
4	B	481	0	0	6	0
All	All	7382	0	6473	72	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 6.

The worst 5 of 72 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:336:ARG:HH12	1:B:118:GLN:HE22	1.22	0.84
1:A:118:GLN:HE22	1:B:336:ARG:HH12	1.29	0.79
1:B:156:LYS:NZ	1:B:156:LYS:HB3	2.01	0.73
1:A:415:ARG:HH11	1:A:415:ARG:HB3	1.56	0.71
1:B:42:ASN:HD21	1:B:46:ARG:HH11	1.37	0.70

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	433/434 (100%)	421 (97%)	10 (2%)	2 (0%)	29	12
1	B	430/434 (99%)	421 (98%)	9 (2%)	0	100	100
All	All	863/868 (99%)	842 (98%)	19 (2%)	2 (0%)	47	29

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	60	LYS
1	A	57	LYS

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	348/341 (102%)	340 (98%)	8 (2%)	50	28
1	B	345/341 (101%)	341 (99%)	4 (1%)	71	56
All	All	693/682 (102%)	681 (98%)	12 (2%)	60	42

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	333	ARG
1	A	367	HIS
1	B	188	GLN
1	A	219	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	4	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	387	GLN
1	A	393	GLN
1	B	188	GLN
1	A	303	ASN
1	B	118	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)

5 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	SO4	A	2001	-	4,4,4	0.37	0	6,6,6	0.38	0
2	SO4	B	2002	-	4,4,4	0.29	0	6,6,6	0.29	0
2	SO4	B	2003	-	4,4,4	0.37	0	6,6,6	0.09	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
3	GOL	A	1001	-	5,5,5	0.51	0	5,5,5	0.69	0
2	SO4	A	2004	-	4,4,4	0.26	0	6,6,6	0.25	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
3	GOL	A	1001	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1001	GOL	O1-C1-C2-C3
3	A	1001	GOL	C1-C2-C3-O3
3	A	1001	GOL	O1-C1-C2-O2
3	A	1001	GOL	O2-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	2002	SO4	1	0
3	A	1001	GOL	1	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Continued from previous page...

Mol	Chain	Number of breaks

Mol	Chain	Number of breaks
1	A	2
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	58:PHE	C	59:ASP	N	1.15
1	A	59:ASP	C	60:LYS	N	0.66
1	B	27:SER	C	28:ALA	N	0.35

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	424/434 (97%)	0.20	29 (6%)	17	22	12, 19, 45, 68	9 (2%)
1	B	423/434 (97%)	-0.08	15 (3%)	44	50	11, 17, 34, 59	3 (0%)
All	All	847/868 (97%)	0.06	44 (5%)	27	33	11, 18, 40, 68	12 (1%)

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	28	ALA	9.8
1	B	27	SER	8.9
1	A	27	SER	8.7
1	B	434	ALA	8.3
1	B	26	ILE	8.0

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. 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	B-factors(Å²)	Q<0.9
3	GOL	A	1001	6/6	0.84	0.23	32,38,39,43	0
2	SO4	B	2003	5/5	0.96	0.15	38,38,40,40	0
2	SO4	B	2002	5/5	0.97	0.11	26,29,31,31	0
2	SO4	A	2004	5/5	0.98	0.15	33,33,34,35	0
2	SO4	A	2001	5/5	0.99	0.07	20,23,25,25	0

6.5 Other polymers (i)

There are no such residues in this entry.