



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 14, 2020 – 08:54 pm BST

PDB ID : 3C5I  
Title : Crystal structure of Plasmodium knowlesi choline kinase, PKH\_134520  
Authors : Wernimont, A.K.; Hills, T.; Lew, J.; Wasney, G.; Senesterra, G.; Kozieradzki, I.; Cossar, D.; Vedadi, M.; Schapira, M.; Bochkarev, A.; Arrowsmith, C.H.; Bountra, C.; Weigelt, J.; Edwards, A.M.; Hui, R.; Artz, J.D.; Xiao, T.; Structural Genomics Consortium (SGC)  
Deposited on : 2008-01-31  
Resolution : 2.20 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 : 2.11  
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

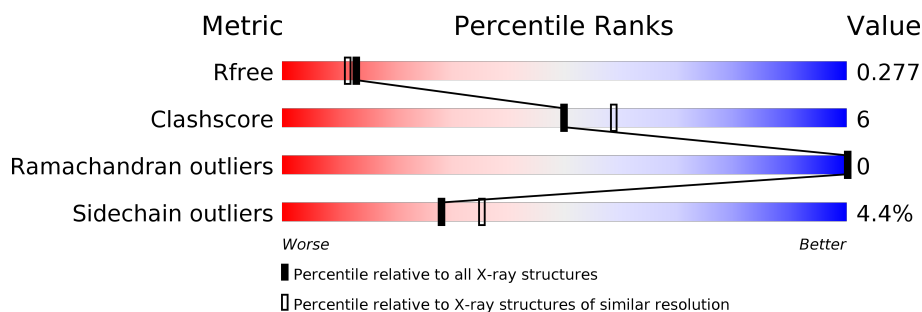
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	369	77% 15% 7%
1	B	369	76% 15% 7%
1	C	369	83% 13% .
1	D	369	83% 12% .
2	E	6	33% 50% 17%

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
6	GOL	D	374	-	-	X	-
6	GOL	D	375	-	-	X	-

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 12122 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 Choline kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	343	Total	C	N	O	S	0	5	0
			2836	1856	451	519	10			
1	B	345	Total	C	N	O	S	0	4	0
			2863	1872	456	526	9			
1	C	353	Total	C	N	O	S	0	5	0
			2953	1928	472	543	10			
1	D	355	Total	C	N	O	S	0	3	0
			2933	1919	466	538	10			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	PDB 3C5I
B	1	GLY	-	EXPRESSION TAG	PDB 3C5I
C	1	GLY	-	EXPRESSION TAG	PDB 3C5I
D	1	GLY	-	EXPRESSION TAG	PDB 3C5I

- Molecule 2 is a protein called Cleaved fragment of N-terminal expression tag.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	5	Total	C	N	O	0	0	0
			48	33	7	8			

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Mg	0	0
			1	1		
3	A	1	Total	Mg	0	0
			1	1		
3	D	1	Total	Mg	0	0
			1	1		

*Continued on next page...*

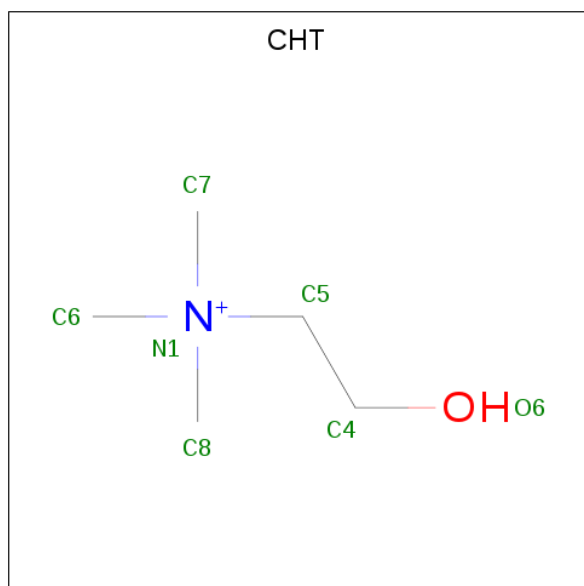
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	1	Total	Mg	0	0
			1	1		

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Ca	0	0
			1	1		
4	A	1	Total	Ca	0	0
			1	1		
4	D	1	Total	Ca	0	0
			1	1		
4	C	1	Total	Ca	0	0
			1	1		

- Molecule 5 is CHOLINE ION (three-letter code: CHT) (formula: C<sub>5</sub>H<sub>14</sub>NO).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			7	5	1	1		
5	B	1	Total	C	N	O	0	0
			7	5	1	1		

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			6	3	3		
6	A	1	Total	C	O	0	0
			6	3	3		
6	B	1	Total	C	O	0	0
			6	3	3		
6	B	1	Total	C	O	0	0
			6	3	3		
6	C	1	Total	C	O	0	0
			6	3	3		
6	C	1	Total	C	O	0	0
			6	3	3		
6	C	1	Total	C	O	0	0
			6	3	3		
6	D	1	Total	C	O	0	0
			6	3	3		
6	D	1	Total	C	O	0	0
			6	3	3		
6	D	1	Total	C	O	0	0
			6	3	3		

- Molecule 7 is UNKNOWN LIGAND (three-letter code: UNL) (formula: ).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	B	1	Total	U	0	0
			1	1		
7	D	1	Total	U	0	0
			1	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	C	1	Total	U	0	0
			1	1		

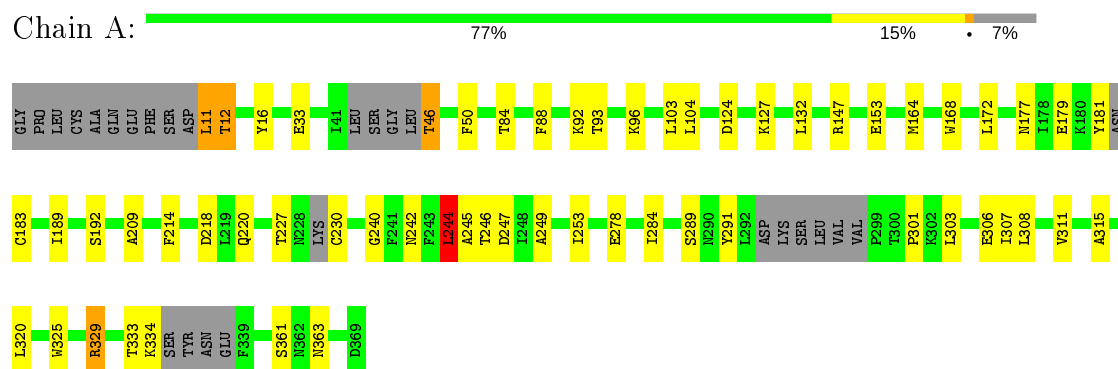
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	105	Total	O	0	0
			105	105		
8	B	83	Total	O	0	0
			83	83		
8	C	122	Total	O	0	0
			122	122		
8	D	93	Total	O	0	0
			93	93		
8	E	1	Total	O	0	0
			1	1		

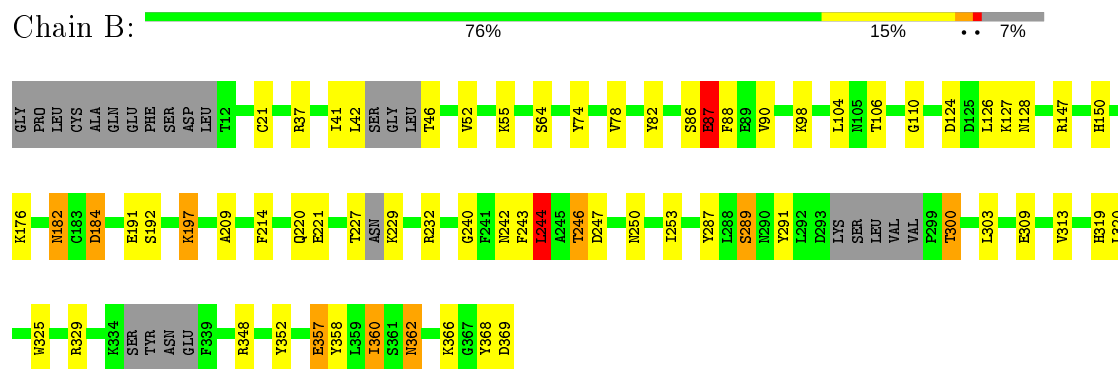
### 3 Residue-property plots [i](#)

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. 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.

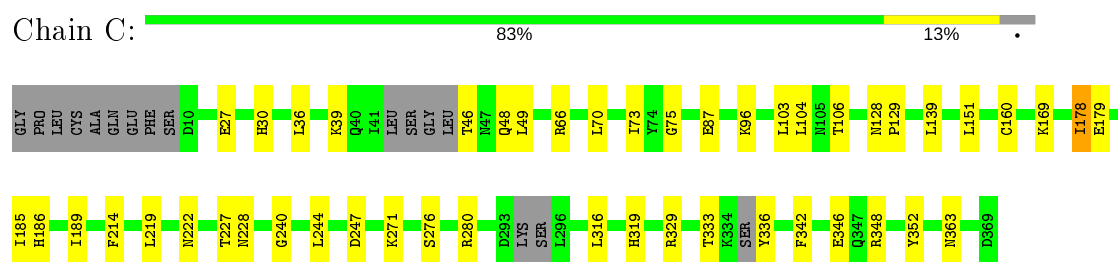
- Molecule 1: Choline kinase



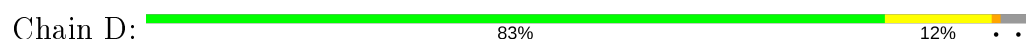
- Molecule 1: Choline kinase



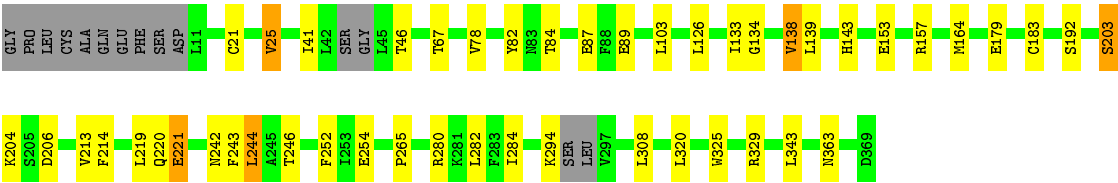
- Molecule 1: Choline kinase



- Molecule 1: Choline kinase







- Molecule 2: Cleaved fragment of N-terminal expression tag



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	112.98Å 71.99Å 116.39Å 90.00° 97.72° 90.00°	Depositor
Resolution (Å)	24.79 – 2.20 24.79 – 2.20	Depositor EDS
% Data completeness (in resolution range)	100.0 (24.79-2.20) 100.0 (24.79-2.20)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	0.05	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	36.68 (at 2.19Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.212 , 0.260 0.245 , 0.277	Depositor DCC
$R_{free}$ test set	4735 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	41.5	Xtriage
Anisotropy	0.180	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 32.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.000 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	12122	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 33.83 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 7.5163e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: UNL, GOL, MG, CHT, CA

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	1.14	17/2919 (0.6%)	0.73	3/3947 (0.1%)
1	B	0.94	15/2944 (0.5%)	0.75	4/3985 (0.1%)
1	C	0.63	0/3040	0.67	0/4113
1	D	1.10	10/3015 (0.3%)	0.73	3/4088 (0.1%)
2	E	0.63	0/49	0.45	0/65
All	All	0.97	42/11967 (0.4%)	0.72	10/16198 (0.1%)

The worst 5 of 42 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	153	GLU	CD-OE1	31.28	1.60	1.25
1	A	46	THR	N-CA	27.55	2.01	1.46
1	A	278	GLU	CD-OE1	26.41	1.54	1.25
1	D	294	LYS	C-O	17.23	1.56	1.23
1	D	203	SER	CB-OG	16.85	1.64	1.42

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	46	THR	N-CA-C	-10.13	83.66	111.00
1	D	153	GLU	OE1-CD-OE2	8.28	133.24	123.30
1	D	244	LEU	CA-CB-CG	7.99	133.68	115.30
1	A	244	LEU	CA-CB-CG	7.82	133.29	115.30
1	D	329	ARG	NE-CZ-NH2	-7.08	116.76	120.30

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	2836	0	2714	34	0
1	B	2863	0	2727	39	0
1	C	2953	0	2855	22	0
1	D	2933	0	2798	36	0
2	E	48	0	42	2	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
5	A	7	0	14	4	0
5	B	7	0	14	3	0
6	A	12	0	16	1	0
6	B	12	0	16	0	0
6	C	18	0	24	5	0
6	D	18	0	23	16	0
7	B	1	0	0	0	0
7	C	1	0	0	0	0
7	D	1	0	0	0	0
8	A	105	0	0	1	0
8	B	83	0	0	3	0
8	C	122	0	0	2	0
8	D	93	0	0	5	0
8	E	1	0	0	0	0
All	All	12122	0	11243	134	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 134 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:B:360:ILE:CB	1:B:360:ILE:CG2	1.78	1.54
1:D:203:SER:CB	1:D:203:SER:OG	1.64	1.44
1:B:197:LYS:CE	1:B:197:LYS:NZ	1.80	1.43
1:B:289[B]:SER:OG	1:B:289[B]:SER:CB	1.66	1.43
1:A:46:THR:CA	1:A:46:THR:N	2.01	1.24

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	336/369 (91%)	325 (97%)	11 (3%)	0	100	100
1	B	339/369 (92%)	329 (97%)	10 (3%)	0	100	100
1	C	350/369 (95%)	340 (97%)	10 (3%)	0	100	100
1	D	352/369 (95%)	339 (96%)	13 (4%)	0	100	100
2	E	3/6 (50%)	3 (100%)	0	0	100	100
All	All	1380/1482 (93%)	1336 (97%)	44 (3%)	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	298/343 (87%)	288 (97%)	10 (3%)	37	47

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	298/343 (87%)	283 (95%)	15 (5%)	24	30
1	C	316/343 (92%)	299 (95%)	17 (5%)	22	26
1	D	307/343 (90%)	293 (95%)	14 (5%)	27	34
2	E	5/6 (83%)	5 (100%)	0	100	100
All	All	1224/1378 (89%)	1168 (95%)	56 (5%)	28	34

5 of 56 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	36	LEU
1	C	104	LEU
1	D	138	VAL
1	C	39[A]	LYS
1	C	87	GLU

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	C	319	HIS
1	C	363	ASN
1	D	207	ASN
1	C	222	ASN
1	D	171	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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 23 ligands modelled in this entry, 3 are unknown and 8 are monoatomic - leaving 12 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	GOL	B	1003	-	5,5,5	0.54	0	5,5,5	0.35	0
6	GOL	D	375	-	5,5,5	0.66	0	5,5,5	0.89	0
6	GOL	C	373	-	5,5,5	0.63	0	5,5,5	0.89	0
6	GOL	A	1002	-	5,5,5	0.38	0	5,5,5	0.68	0
6	GOL	C	374	-	5,5,5	0.73	0	5,5,5	0.96	0
5	CHT	B	1001	-	6,6,6	0.80	0	8,8,8	0.98	0
6	GOL	D	374	-	5,5,5	0.39	0	5,5,5	0.59	0
6	GOL	D	373	-	5,5,5	0.48	0	5,5,5	0.41	0
6	GOL	A	1003	-	5,5,5	0.48	0	5,5,5	0.87	0
6	GOL	C	375	-	5,5,5	0.43	0	5,5,5	0.44	0
5	CHT	A	1001	-	6,6,6	0.81	0	8,8,8	0.49	0
6	GOL	B	1004	-	5,5,5	0.41	0	5,5,5	0.72	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
6	GOL	B	1003	-	-	2/4/4/4	-
6	GOL	D	375	-	-	4/4/4/4	-
6	GOL	C	373	-	-	3/4/4/4	-
6	GOL	A	1002	-	-	4/4/4/4	-
6	GOL	C	374	-	-	1/4/4/4	-
5	CHT	B	1001	-	-	3/4/4/4	-
6	GOL	D	374	-	-	3/4/4/4	-
6	GOL	D	373	-	-	0/4/4/4	-
6	GOL	A	1003	-	-	4/4/4/4	-
6	GOL	C	375	-	-	2/4/4/4	-

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	CHT	A	1001	-	-	3/4/4/4	-
6	GOL	B	1004	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 31 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	C	373	GOL	C1-C2-C3-O3
6	A	1002	GOL	O1-C1-C2-C3
6	A	1002	GOL	C1-C2-C3-O3
6	B	1004	GOL	O1-C1-C2-O2
6	B	1004	GOL	O1-C1-C2-C3

There are no ring outliers.

8 monomers are involved in 29 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	D	375	GOL	10	0
6	C	373	GOL	2	0
6	C	374	GOL	2	0
5	B	1001	CHT	3	0
6	D	374	GOL	6	0
6	A	1003	GOL	1	0
6	C	375	GOL	1	0
5	A	1001	CHT	4	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 ⓘ

### 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.

### 6.5 Other polymers ⓘ

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