Full wwPDB X-ray Structure Validation Report (i)

Aug 21, 2023 – 10:23 PM EDT

PDB ID	:	2PEE
Title	:	Crystal Structure of a Thermophilic Serpin, Tengpin, in the Native State
Authors	:	Zhang, Q.W.; Buckle, A.M.; Whisstock, J.C.
Deposited on	:	2007-04-02
Resolution	:	2.70 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (1)) were used in the production of this report:

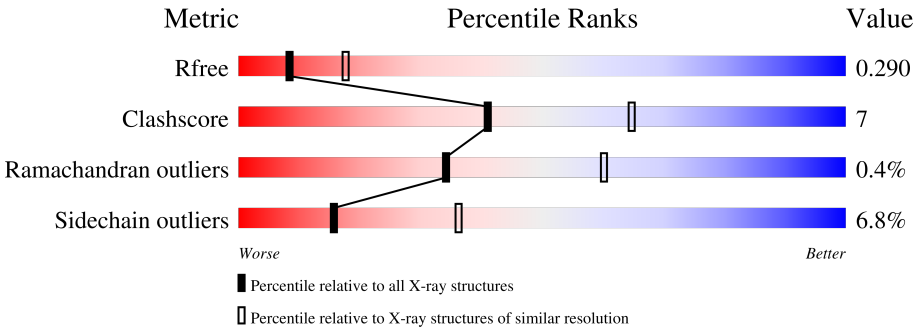
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)

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	387	85% 14% .
1	B	387	80% 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
3	GOL	A	425	-	X	-	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5950 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.

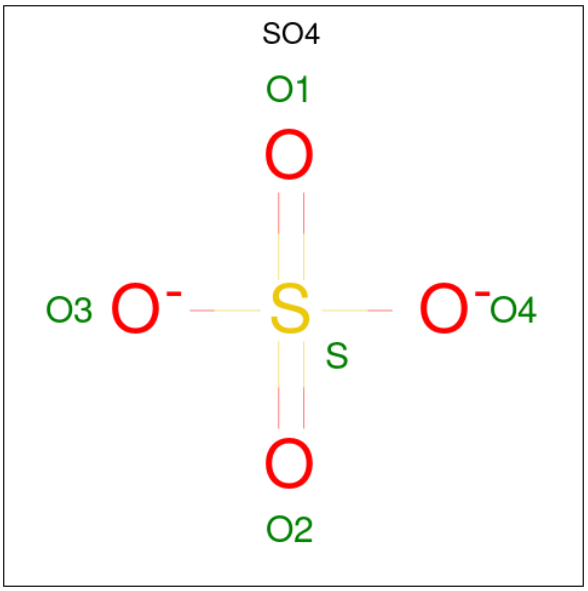
- Molecule 1 is a protein called Serine protease inhibitor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	387	Total	C	N	O	S	0	0	0
			3007	1938	469	589	11			
1	B	387	Total	C	N	O	S	0	0	0
			2888	1859	448	572	9			

There are 2 discrepancies between the modelled and reference sequences:

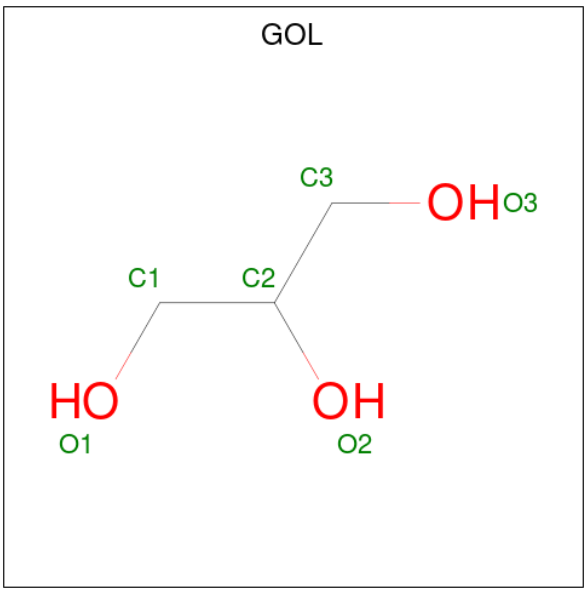
Chain	Residue	Modelled	Actual	Comment	Reference
A	261	ILE	VAL	engineered mutation	UNP Q8R9P5
B	261	ILE	VAL	engineered mutation	UNP Q8R9P5

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	O	S		0	0
			5	4	1			

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 4 is water.

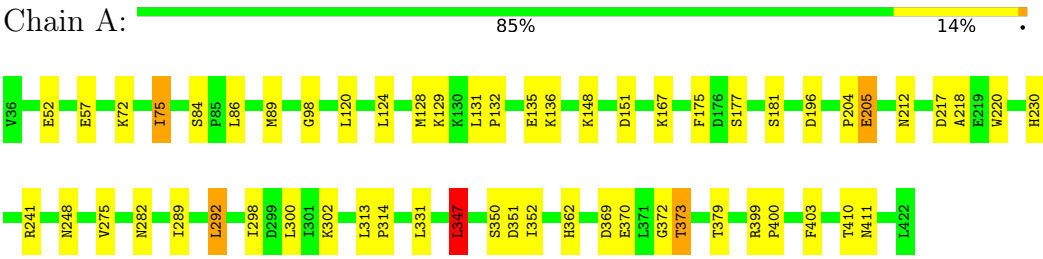
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	25	Total O 25 25	0	0
4	B	13	Total O 13 13	0	0

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.

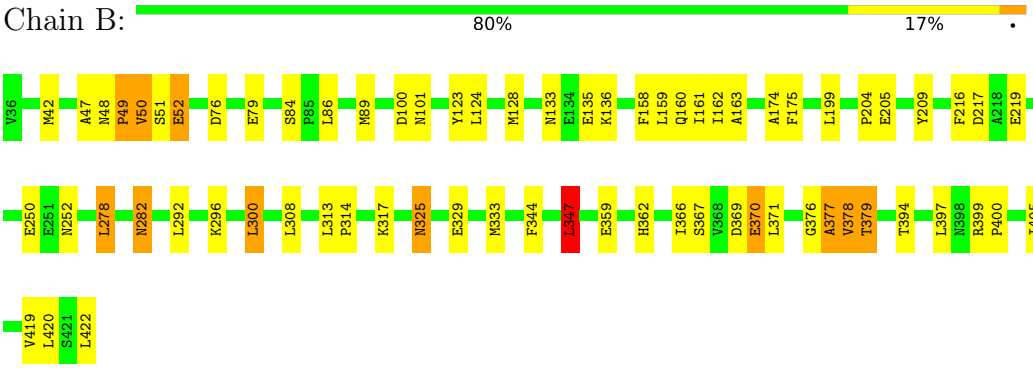
- Molecule 1: Serine protease inhibitor

Chain A:



- Molecule 1: Serine protease inhibitor

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, α, β, γ	217.94Å 217.94Å 217.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	154.00 – 2.70 77.05 – 2.70	Depositor EDS
% Data completeness (in resolution range)	97.8 (154.00-2.70) 97.8 (77.05-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	7.70	Depositor
$< I/\sigma(I) >$ [1]	2.23 (at 2.69Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.195 , 0.235 0.264 , 0.290	Depositor DCC
R_{free} test set	2321 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	66.4	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 42.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.027 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	5950	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.01% 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, 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	0.54	0/3065	0.67	2/4159 (0.0%)
1	B	0.53	1/2945 (0.0%)	0.65	1/4021 (0.0%)
All	All	0.54	1/6010 (0.0%)	0.66	3/8180 (0.0%)

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	B	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	159	LEU	C-O	6.06	1.34	1.23

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	300	LEU	CA-CB-CG	7.42	132.37	115.30
1	A	347	LEU	CA-CB-CG	6.28	129.75	115.30
1	B	347	LEU	CA-CB-CG	5.39	127.69	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	377	ALA	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	B	50	VAL	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	3007	0	2922	32	0
1	B	2888	0	2634	50	0
2	A	5	0	0	0	0
3	A	12	0	16	0	0
4	A	25	0	0	0	0
4	B	13	0	0	0	0
All	All	5950	0	5572	81	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 7.

All (81) 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:204:PRO:O	1:A:205:GLU:HB2	1.72	0.89
1:B:51:SER:O	1:B:52:GLU:C	2.29	0.71
1:B:250:GLU:HG2	1:B:308:LEU:HD11	1.73	0.70
1:B:47:ALA:C	1:B:48:ASN:OD1	2.30	0.70
1:B:371:LEU:CD1	1:B:376:GLY:HA3	2.24	0.68
1:A:84:SER:H	1:A:362:HIS:HE1	1.41	0.67
1:A:52:GLU:OE2	1:A:129:LYS:HE2	1.94	0.67
1:B:399:ARG:HB2	1:B:400:PRO:HD2	1.77	0.66
1:A:220:TRP:NE1	1:A:372:GLY:HA2	2.12	0.64
1:B:369:ASP:OD1	1:B:370:GLU:HG2	2.00	0.62
1:B:371:LEU:HG	1:B:376:GLY:HA3	1.81	0.62
1:B:84:SER:H	1:B:362:HIS:CE1	2.18	0.61
1:B:371:LEU:HD11	1:B:376:GLY:HA3	1.83	0.60
1:B:399:ARG:HB2	1:B:400:PRO:CD	2.32	0.59
1:B:378:VAL:CG2	1:B:379:THR:N	2.64	0.59
1:B:329:GLU:HA	1:B:333:MET:O	2.03	0.59

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:42:MET:CE	1:B:163:ALA:HB2	2.32	0.58
1:B:377:ALA:O	1:B:378:VAL:HG12	2.03	0.57
1:A:220:TRP:CD1	1:A:372:GLY:HA2	2.39	0.57
1:B:366:ILE:HD11	1:B:405:ILE:HD11	1.87	0.56
1:B:344:PHE:HB3	1:B:347:LEU:HD22	1.87	0.56
1:A:98:GLY:HA3	1:A:347:LEU:HB2	1.87	0.56
1:B:371:LEU:CG	1:B:376:GLY:HA3	2.35	0.56
1:A:84:SER:H	1:A:362:HIS:CE1	2.24	0.55
1:B:216:PHE:HB2	1:B:415:PHE:HE1	1.72	0.55
1:B:84:SER:H	1:B:362:HIS:HE1	1.54	0.55
1:B:136:LYS:HE3	1:B:219:GLU:OE2	2.07	0.54
1:B:282:ASN:H	1:B:282:ASN:HD22	1.56	0.54
1:B:209:TYR:CE2	1:B:359:GLU:HB2	2.44	0.53
1:A:135:GLU:O	1:A:136:LYS:HB2	2.08	0.53
1:A:217:ASP:HA	1:A:373:THR:HG21	1.89	0.53
1:B:136:LYS:HE3	1:B:219:GLU:CD	2.29	0.52
1:B:42:MET:HE3	1:B:163:ALA:HB2	1.91	0.52
1:A:399:ARG:HB2	1:A:400:PRO:CD	2.39	0.52
1:A:233:ILE:HG21	1:A:241:ARG:HB3	1.93	0.51
1:B:160:GLN:O	1:B:161:ILE:C	2.48	0.50
1:B:209:TYR:HE2	1:B:359:GLU:HB2	1.75	0.50
1:B:48:ASN:O	1:B:49:PRO:O	2.30	0.50
1:B:400:PRO:HA	1:B:419:VAL:O	2.12	0.49
1:B:158:PHE:O	1:B:162:ILE:HG12	2.12	0.49
1:A:89:MET:CE	1:A:124:LEU:HD21	2.43	0.48
1:A:86:LEU:HD21	1:A:128:MET:CE	2.44	0.48
1:B:174:ALA:O	1:B:175:PHE:HB2	2.14	0.48
1:B:86:LEU:HD21	1:B:128:MET:CE	2.44	0.47
1:A:151:ASP:OD1	1:A:351:ASP:N	2.44	0.47
1:A:230:HIS:HD2	1:A:248:ASN:HD22	1.62	0.47
1:A:86:LEU:HD11	1:A:128:MET:HE3	1.97	0.46
1:B:366:ILE:HG12	1:B:367:SER:N	2.30	0.46
1:A:212:ASN:OD1	1:A:362:HIS:HD2	1.98	0.46
1:B:296:LYS:O	1:B:300:LEU:HB2	2.15	0.46
1:B:236:ASP:HB2	1:B:420:LEU:O	2.16	0.46
1:B:86:LEU:HD21	1:B:128:MET:HE3	1.96	0.46
1:A:75:ILE:H	1:A:75:ILE:HG13	1.64	0.46
1:A:220:TRP:CE2	1:A:372:GLY:HA2	2.51	0.45
1:B:366:ILE:HG12	1:B:367:SER:H	1.81	0.45
1:A:148:LYS:HG3	1:A:175:PHE:HB2	1.98	0.45
1:B:48:ASN:OD1	1:B:48:ASN:N	2.45	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:205:GLU:HG3	1:B:238:TYR:CG	2.52	0.45
1:B:86:LEU:HD11	1:B:128:MET:HE3	1.99	0.45
1:A:313:LEU:HD12	1:A:314:PRO:HD2	1.99	0.44
1:B:370:GLU:HG2	1:B:370:GLU:H	1.61	0.44
1:A:218:ALA:O	1:A:373:THR:HB	2.17	0.44
1:B:325:ASN:HD21	1:B:359:GLU:HA	1.82	0.44
1:B:100:ASP:OD1	1:B:101:ASN:N	2.49	0.43
1:A:298:ILE:HG22	1:A:302:LYS:HE2	2.00	0.43
1:B:278:LEU:HG	1:B:397:LEU:HD23	2.00	0.43
1:B:378:VAL:HG23	1:B:379:THR:N	2.33	0.43
1:A:151:ASP:OD2	1:A:352:ILE:N	2.51	0.43
1:A:275:VAL:O	1:A:403:PHE:HA	2.19	0.43
1:B:377:ALA:C	1:B:378:VAL:HG12	2.37	0.42
1:B:51:SER:O	1:B:52:GLU:O	2.38	0.42
1:A:86:LEU:HD21	1:A:128:MET:HE3	2.02	0.42
1:A:289:ILE:O	1:A:292:LEU:HB2	2.19	0.42
1:A:233:ILE:CG2	1:A:241:ARG:HB3	2.50	0.41
1:B:325:ASN:OD1	1:B:325:ASN:N	2.53	0.41
1:B:313:LEU:HD12	1:B:314:PRO:HD2	2.03	0.41
1:B:204:PRO:O	1:B:205:GLU:CB	2.68	0.41
1:B:89:MET:HE3	1:B:124:LEU:HD21	2.02	0.40
1:A:131:LEU:HA	1:A:132:PRO:HD3	1.93	0.40
1:A:331:LEU:HD12	1:A:331:LEU:HA	1.92	0.40
1:A:410:THR:O	1:A:411:ASN:HB2	2.21	0.40

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	385/387 (100%)	362 (94%)	23 (6%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	385/387 (100%)	358 (93%)	24 (6%)	3 (1%)	19 43
All	All	770/774 (100%)	720 (94%)	47 (6%)	3 (0%)	34 60

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	50	VAL
1	B	52	GLU
1	B	49	PRO

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	320/342 (94%)	302 (94%)	18 (6%)	21 45
1	B	280/342 (82%)	257 (92%)	23 (8%)	11 26
All	All	600/684 (88%)	559 (93%)	41 (7%)	16 36

All (41) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	57	GLU
1	A	72	LYS
1	A	75	ILE
1	A	120	LEU
1	A	167	LYS
1	A	177	SER
1	A	181	SER
1	A	196	ASP
1	A	205	GLU
1	A	237	VAL
1	A	282	ASN
1	A	292	LEU
1	A	347	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	350	SER
1	A	369	ASP
1	A	370	GLU
1	A	373	THR
1	A	379	THR
1	B	76	ASP
1	B	79	GLU
1	B	123	TYR
1	B	133	ASN
1	B	135	GLU
1	B	199	LEU
1	B	217	ASP
1	B	235	THR
1	B	237	VAL
1	B	252	ASN
1	B	278	LEU
1	B	282	ASN
1	B	292	LEU
1	B	300	LEU
1	B	317	LYS
1	B	325	ASN
1	B	347	LEU
1	B	370	GLU
1	B	378	VAL
1	B	379	THR
1	B	394	THR
1	B	407	ASP
1	B	422	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (14) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	200	ASN
1	A	212	ASN
1	A	230	HIS
1	A	282	ASN
1	A	303	ASN
1	A	362	HIS
1	B	133	ASN
1	B	212	ASN
1	B	240	ASN
1	B	252	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	282	ASN
1	B	303	ASN
1	B	362	HIS
1	B	391	ASN

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	$\#\|Z\| > 2$
3	GOL	A	425	-	5,5,5	0.47	0	5,5,5	4.83	3 (60%)
2	SO4	A	423	-	4,4,4	0.19	0	6,6,6	0.30	0
3	GOL	A	424	-	5,5,5	0.46	0	5,5,5	0.57	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	425	-	-	3/4/4/4	-
3	GOL	A	424	-	-	2/4/4/4	-

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	425	GOL	O2-C2-C1	-6.73	79.50	109.12
3	A	425	GOL	O2-C2-C3	-6.41	80.89	109.12
3	A	425	GOL	C3-C2-C1	5.47	132.99	111.70

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	425	GOL	C1-C2-C3-O3
3	A	425	GOL	O1-C1-C2-O2
3	A	424	GOL	O1-C1-C2-C3
3	A	425	GOL	O2-C2-C3-O3
3	A	424	GOL	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

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.