wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:11 PM GMT

PDB ID	:	1SJ2
Title	:	Crystal structure of Mycobacterium tuberculosis catalase-peroxidase
Authors	:	Bertrand, T.; Eady, N.A.J.; Jones, J.N.; Bodiguel, J.; Jesmin; Nagy, J.M.; Raven, E.L.; Jamart-Gregoire, B.; Brown, K.A.
Deposited on	:	2004-03-02
Resolution	:	2.41 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

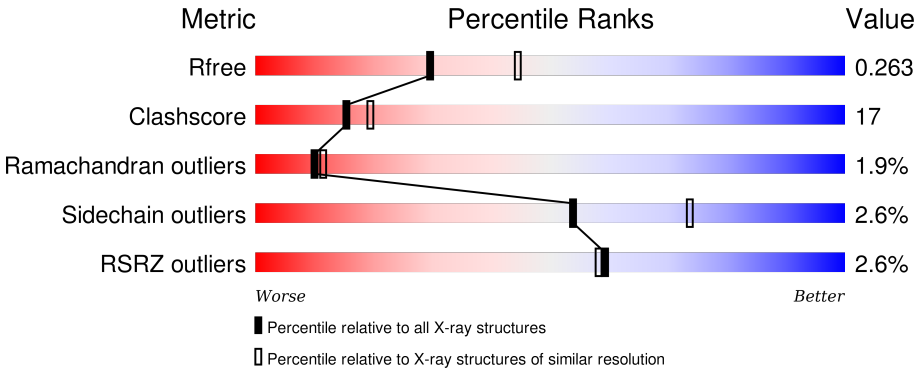
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.41 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}	91344	3386 (2.44-2.40)
Clashscore	102246	3897 (2.44-2.40)
Ramachandran outliers	100387	3837 (2.44-2.40)
Sidechain outliers	100360	3838 (2.44-2.40)
RSRZ outliers	91569	3396 (2.44-2.40)

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	743	2% 68% 26%
1	B	743	3% 64% 30%

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	1194	-	X	-	-
3	GOL	A	1195	-	X	-	X
3	GOL	A	1196	-	X	-	X
3	GOL	B	2194	-	X	-	-
3	GOL	B	2195	-	X	-	-
3	GOL	B	2196	-	X	-	X

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 11883 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.

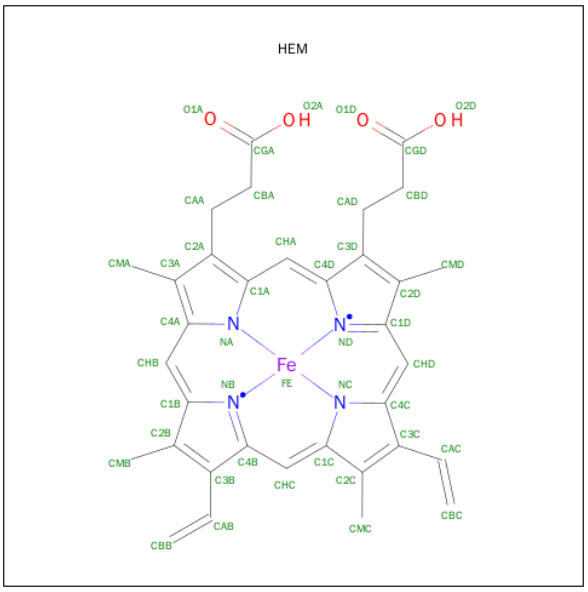
- Molecule 1 is a protein called Peroxidase/catalase T.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	717	Total	C	N	O	S	0	0	0
			5529	3515	953	1042	19			
1	B	717	Total	C	N	O	S	0	0	0
			5529	3515	953	1042	19			

There are 8 discrepancies between the modelled and reference sequences:

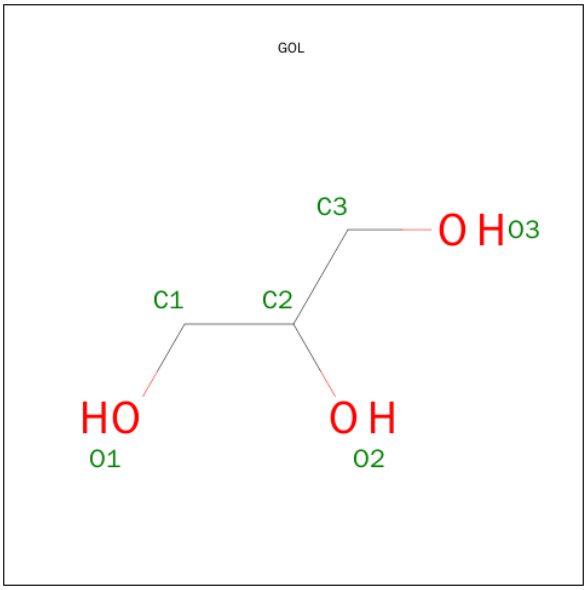
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	MET	-	INITIATING METHIONINE	UNP Q08129
A	-1	GLU	-	CLONING ARTIFACT	UNP Q08129
A	0	PHE	-	CLONING ARTIFACT	UNP Q08129
A	1	VAL	-	CLONING ARTIFACT	UNP Q08129
B	-2	MET	-	INITIATING METHIONINE	UNP Q08129
B	-1	GLU	-	CLONING ARTIFACT	UNP Q08129
B	0	PHE	-	CLONING ARTIFACT	UNP Q08129
B	1	VAL	-	CLONING ARTIFACT	UNP Q08129

- Molecule 2 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula: $C_{34}H_{32}FeN_4O_4$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
2	B	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		

- 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	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		

Continued on next page...

Continued from previous page...

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

- Molecule 4 is water.

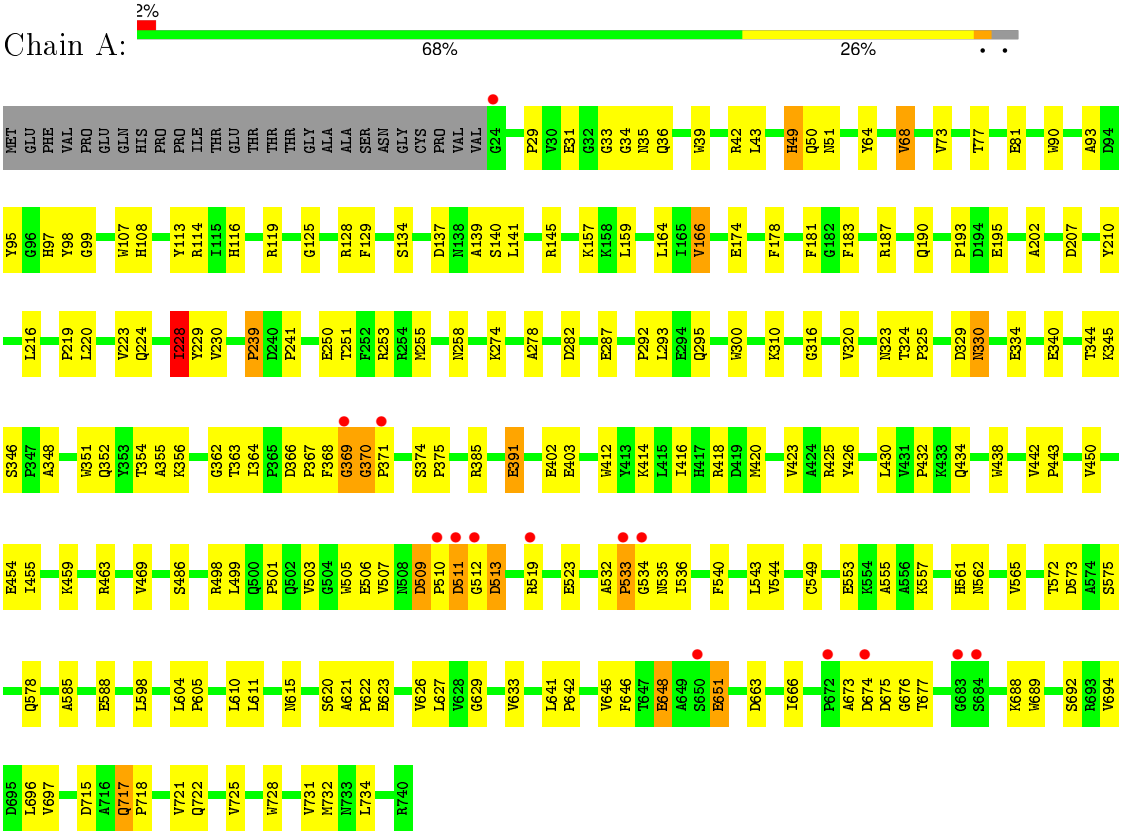
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	381	Total O 381 381	0	0
4	B	322	Total O 322 322	0	0

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 errors 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.

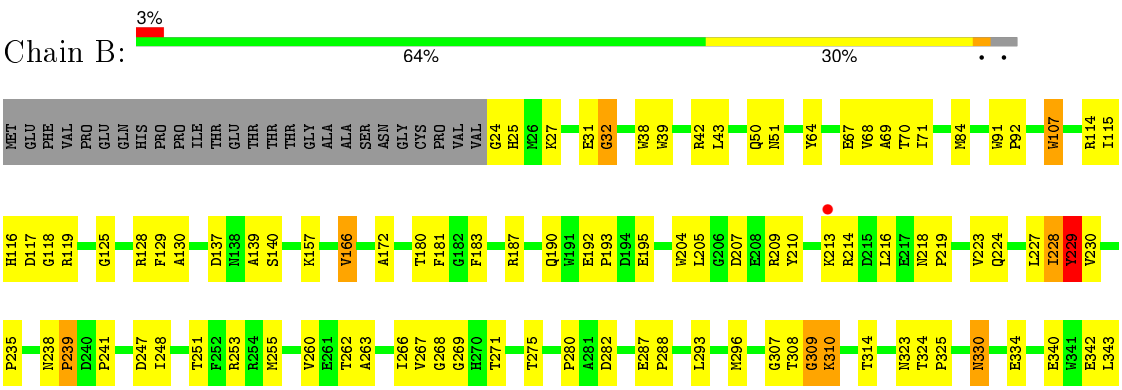
- Molecule 1: Peroxidase/catalase T

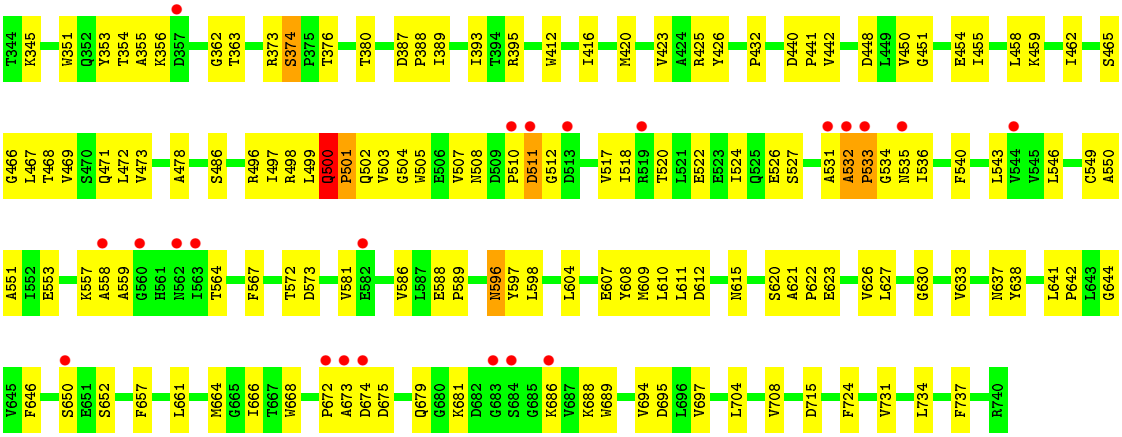
Chain A:



- Molecule 1: Peroxidase/catalase T

Chain B:




T344 K345 W351 Q352 Y353 T354 A355 K356 D357 G362 T363 R373 S374 P375 T376 T380 D387 P388 I389 I393 T394 R395 W412 I416 M420 V423 A424 R425 Y426 P432 D440 P441 V442 D448 L449 V450 G451 E454 I455 L458 K459 I462 S465
G466 L467 T468 V469 S470 Q471 L472 V473 A478 S486 R496 I497 R498 L499 Q500 P501 Q502 V503 G504 W505 E506 V507 N508 D509 P510 D511 G512 D513 V517 I518 R519 T520 L521 E522 E523 I524 Q525 E526 S527 A531 A532 P533 G534 N535 I536 F540 L543 V544 V545 L546 C549 A550
A551 I552 E553 K557 A558 A559 G560 H561 N562 I563 T564 F567 T572 D573 V681 E582 V686 L587 E588 P589 N596 Y597 L598 L604 E607 Y608 M609 L610 L611 D612 N615 S620 A621 P622 E623 V626 L627 G630 V633 N637 Y638 L641 P642 L643 G644
V645 F646 S650 E651 S652 F657 L661 H664 G665 I666 T667 W668 P672 A673 D674 D675 Q679 G680 K681 D682 G683 S684 G685 K686 V687 K688 W689 V694 D695 L696 V697 L704 V708 D715 F724 V731 L734 F737 R740

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 42 21 2	Depositor
Cell constants a, b, c, α, β, γ	150.33Å 150.33Å 154.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	23.71 – 2.41 23.71 – 2.41	Depositor EDS
% Data completeness (in resolution range)	90.3 (23.71-2.41) 90.3 (23.71-2.41)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.16	Depositor
$< I/\sigma(I) >$ [1]	3.70 (at 2.41Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.211 , 0.268 0.204 , 0.263	Depositor DCC
R_{free} test set	6260 reflections (11.27%)	DCC
Wilson B-factor (Å^2)	27.6	Xtriage
Anisotropy	1.346	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 45.5	EDS
Estimated twinning fraction	0.027 for -h,l,k 0.007 for -l,-k,-h	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 61822 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11883	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.72% 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, 0.375 respectively for untwinned datasets, and 0.333, 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, HEM

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	0.42	2/5680 (0.0%)	0.61	2/7732 (0.0%)
1	B	0.46	2/5680 (0.0%)	0.74	3/7732 (0.0%)
All	All	0.44	4/11360 (0.0%)	0.68	5/15464 (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	A	0	1
1	B	0	1
All	All	0	2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	228	ILE	C-N	-18.48	0.91	1.34
1	B	229	TYR	C-N	-12.69	1.04	1.34
1	A	228	ILE	C-N	-12.30	1.05	1.34
1	A	229	TYR	C-N	-11.96	1.06	1.34

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	228	ILE	O-C-N	-38.32	61.39	122.70
1	A	228	ILE	O-C-N	-6.67	112.02	122.70
1	B	228	ILE	CA-C-N	5.48	129.25	117.20
1	A	229	TYR	C-N-CA	5.13	134.53	121.70
1	B	500	GLN	N-CA-C	5.01	124.53	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	228	ILE	Mainchain
1	B	228	ILE	Mainchain

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	5529	0	5353	178	0
1	B	5529	0	5353	209	0
2	A	43	0	30	1	0
2	B	43	0	30	4	0
3	A	18	0	12	1	0
3	B	18	0	12	0	0
4	A	381	0	0	18	0
4	B	322	0	0	30	0
All	All	11883	0	10790	381	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 17.

The worst 5 of 381 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:51:ASN:HD21	1:A:190:GLN:HB3	1.28	0.98
1:A:366:ASP:HB3	1:A:370:GLY:HA2	1.51	0.93
1:B:51:ASN:HD21	1:B:190:GLN:HB3	1.38	0.88
1:A:253:ARG:HB3	1:A:253:ARG:HH11	1.36	0.88
1:A:450:VAL:HG13	1:A:454:GLU:HB2	1.56	0.88

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	715/743 (96%)	663 (93%)	39 (6%)	13 (2%)	11 12
1	B	715/743 (96%)	656 (92%)	45 (6%)	14 (2%)	9 10
All	All	1430/1486 (96%)	1319 (92%)	84 (6%)	27 (2%)	10 11

5 of 27 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	35	ASN
1	A	511	ASP
1	A	533	PRO
1	B	32	GLY
1	B	229	TYR

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	565/587 (96%)	549 (97%)	16 (3%)	51 71
1	B	565/587 (96%)	552 (98%)	13 (2%)	58 77
All	All	1130/1174 (96%)	1101 (97%)	29 (3%)	54 74

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	648	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	107	TRP
1	B	596	ASN
1	A	651	GLU
1	B	114	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	717	GLN
1	B	51	ASN
1	B	615	ASN
1	B	49	HIS
1	B	50	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	$\#\|Z\|>2$
3	GOL	A	1194	-	5,5,5	4.79	5 (100%)	5,5,5	5.68	3 (60%)
3	GOL	A	1195	-	5,5,5	4.81	5 (100%)	5,5,5	5.71	3 (60%)
3	GOL	A	1196	-	5,5,5	4.75	5 (100%)	5,5,5	5.72	3 (60%)
2	HEM	A	1500	1	30,50,50	2.16	7 (23%)	24,82,82	2.34	7 (29%)
2	HEM	B	1500	1	30,50,50	2.26	7 (23%)	24,82,82	2.44	8 (33%)
3	GOL	B	2194	-	5,5,5	4.86	5 (100%)	5,5,5	5.64	3 (60%)
3	GOL	B	2195	-	5,5,5	4.77	5 (100%)	5,5,5	5.67	3 (60%)
3	GOL	B	2196	-	5,5,5	4.66	5 (100%)	5,5,5	5.70	3 (60%)

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	1194	-	-	0/4/4/4	0/0/0/0
3	GOL	A	1195	-	-	0/4/4/4	0/0/0/0
3	GOL	A	1196	-	-	0/4/4/4	0/0/0/0
2	HEM	A	1500	1	-	0/10/54/54	0/0/8/8
2	HEM	B	1500	1	-	0/10/54/54	0/0/8/8
3	GOL	B	2194	-	-	0/4/4/4	0/0/0/0
3	GOL	B	2195	-	-	0/4/4/4	0/0/0/0
3	GOL	B	2196	-	-	0/4/4/4	0/0/0/0

The worst 5 of 44 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	2194	GOL	C3-C2	-8.57	1.19	1.52
3	A	1194	GOL	C3-C2	-8.22	1.20	1.52
3	A	1195	GOL	C3-C2	-8.16	1.21	1.52
3	B	2195	GOL	C3-C2	-8.13	1.21	1.52
3	A	1196	GOL	C3-C2	-8.08	1.21	1.52

The worst 5 of 33 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	1500	HEM	C3B-CAB-CBB	2.02	127.56	124.46
2	B	1500	HEM	CBD-CAD-C3D	2.08	119.61	113.55
2	A	1500	HEM	C3C-CAC-CBC	2.21	127.85	124.46
2	B	1500	HEM	C2D-C3D-C4D	2.80	106.25	101.50

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1500	HEM	C2D-C3D-C4D	2.96	106.51	101.50

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1195	GOL	1	0
2	A	1500	HEM	1	0
2	B	1500	HEM	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

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^{th} 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	717/743 (96%)	-0.21	14 (1%)	68	67	21, 34, 50, 72	0
1	B	717/743 (96%)	-0.07	23 (3%)	51	50	24, 39, 63, 89	0
All	All	1434/1486 (96%)	-0.14	37 (2%)	59	58	21, 36, 59, 89	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	532	ALA	5.3
1	B	673	ALA	4.4
1	A	511	ASP	4.2
1	B	533	PRO	4.1
1	B	535	ASN	4.1

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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, 95^{th} 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
3	GOL	B	2196	6/6	0.78	0.28	4.95	49,52,53,53	0
3	GOL	A	1196	6/6	0.89	0.20	4.19	48,50,52,52	0
3	GOL	A	1195	6/6	0.83	0.20	2.65	38,39,42,42	0
3	GOL	B	2194	6/6	0.94	0.14	1.76	27,31,34,38	0
3	GOL	A	1194	6/6	0.95	0.15	1.75	31,33,35,35	0
2	HEM	A	1500	43/43	0.97	0.17	0.90	24,29,32,34	0
2	HEM	B	1500	43/43	0.96	0.15	0.51	25,32,34,39	0
3	GOL	B	2195	6/6	0.94	0.13	-0.62	41,43,43,44	0

6.5 Other polymers (i)

There are no such residues in this entry.