



wwPDB X-ray Structure Validation Summary Report ⓘ

May 13, 2020 – 02:13 am BST

PDB ID : 6GMZ
Title : Staphylopine dehydrogenase in complex with Histidine tag and citrate
Authors : Hajjar, C.; Arnoux, P.
Deposited on : 2018-05-29
Resolution : 2.22 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 ⓘ 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

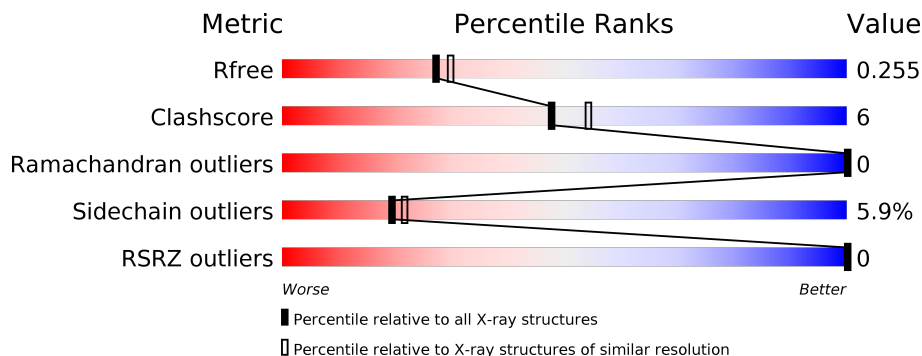
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.22 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	5912 (2.24-2.20)
Clashscore	141614	6646 (2.24-2.20)
Ramachandran outliers	138981	6543 (2.24-2.20)
Sidechain outliers	138945	6544 (2.24-2.20)
RSRZ outliers	127900	5797 (2.24-2.20)

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 ≥ 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 $\leq 5\%$. 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	465	
2	B	4	

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	NH4	A	501	-	-	X	-

2 Entry composition i

There are 6 unique types of molecules in this entry. The entry contains 3749 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 Staphylopin dehydrogenase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	428	3476	2231	566	656	4	19	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	434	LYS	-	expression tag	UNP A0A1Q4GXD5
A	435	GLY	-	expression tag	UNP A0A1Q4GXD5
A	436	GLU	-	expression tag	UNP A0A1Q4GXD5
A	437	LEU	-	expression tag	UNP A0A1Q4GXD5
A	438	ASN	-	expression tag	UNP A0A1Q4GXD5
A	439	SER	-	expression tag	UNP A0A1Q4GXD5
A	440	LYS	-	expression tag	UNP A0A1Q4GXD5
A	441	LEU	-	expression tag	UNP A0A1Q4GXD5
A	442	GLU	-	expression tag	UNP A0A1Q4GXD5
A	443	GLY	-	expression tag	UNP A0A1Q4GXD5
A	444	LYS	-	expression tag	UNP A0A1Q4GXD5
A	445	PRO	-	expression tag	UNP A0A1Q4GXD5
A	446	ILE	-	expression tag	UNP A0A1Q4GXD5
A	447	PRO	-	expression tag	UNP A0A1Q4GXD5
A	448	ASN	-	expression tag	UNP A0A1Q4GXD5
A	449	PRO	-	expression tag	UNP A0A1Q4GXD5
A	450	LEU	-	expression tag	UNP A0A1Q4GXD5
A	451	LEU	-	expression tag	UNP A0A1Q4GXD5
A	452	GLY	-	expression tag	UNP A0A1Q4GXD5
A	453	LEU	-	expression tag	UNP A0A1Q4GXD5
A	454	ASP	-	expression tag	UNP A0A1Q4GXD5
A	455	SER	-	expression tag	UNP A0A1Q4GXD5
A	456	THR	-	expression tag	UNP A0A1Q4GXD5
A	457	ARG	-	expression tag	UNP A0A1Q4GXD5
A	458	THR	-	expression tag	UNP A0A1Q4GXD5
A	459	GLY	-	expression tag	UNP A0A1Q4GXD5
A	460	HIS	-	expression tag	UNP A0A1Q4GXD5

Continued on next page...

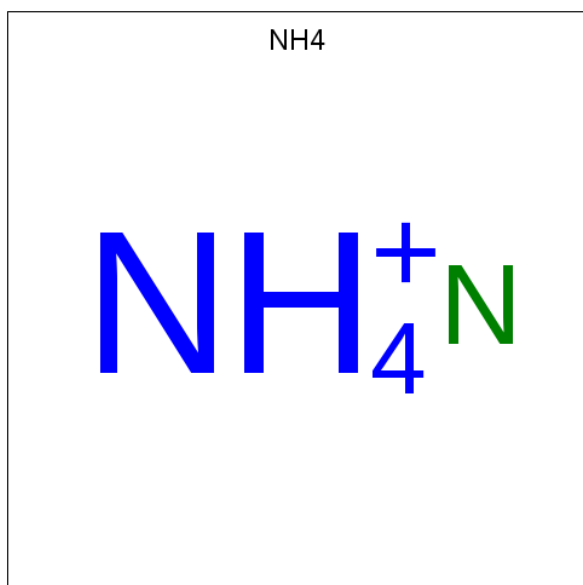
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	461	HIS	-	expression tag	UNP A0A1Q4GXD5
A	462	HIS	-	expression tag	UNP A0A1Q4GXD5
A	463	HIS	-	expression tag	UNP A0A1Q4GXD5
A	464	HIS	-	expression tag	UNP A0A1Q4GXD5
A	465	HIS	-	expression tag	UNP A0A1Q4GXD5

- Molecule 2 is a protein called ALA-HIS-HIS-ALA.

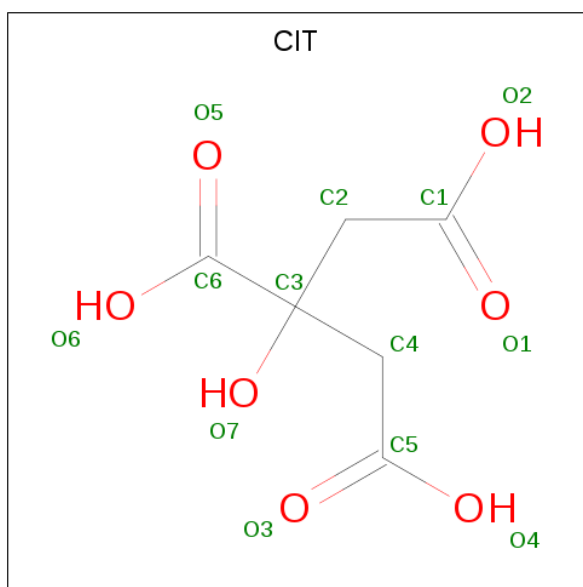
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	4	30	18	8	4	0	0	0

- Molecule 3 is AMMONIUM ION (three-letter code: NH4) (formula: H₄N).



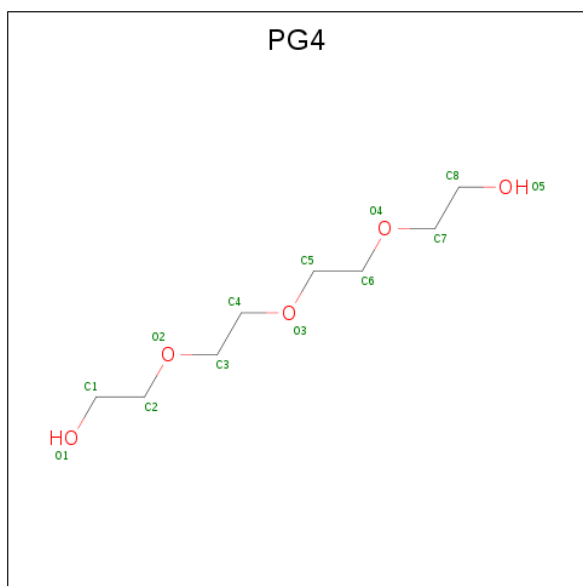
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	N		
3	A	1	1	1	0	0

- Molecule 4 is CITRIC ACID (three-letter code: CIT) (formula: C₆H₈O₇).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			13	6	7		

- Molecule 5 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			13	8	5		

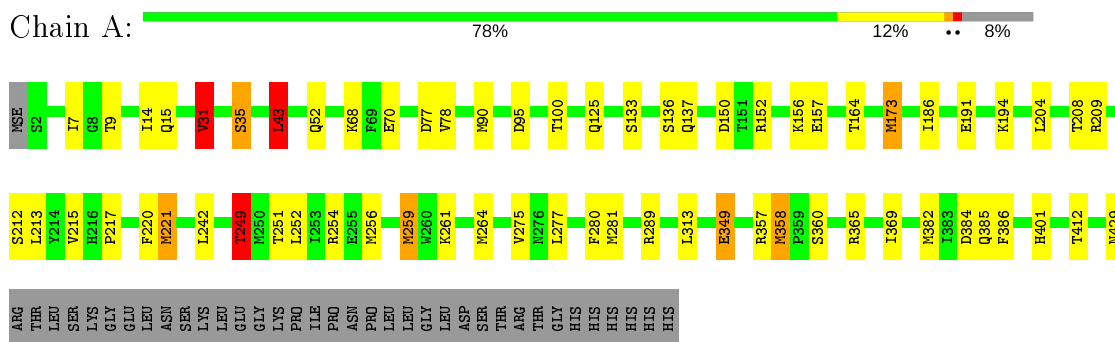
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	212	Total 212	O 212	0	0
6	B	4	Total 4	O 4	0	0

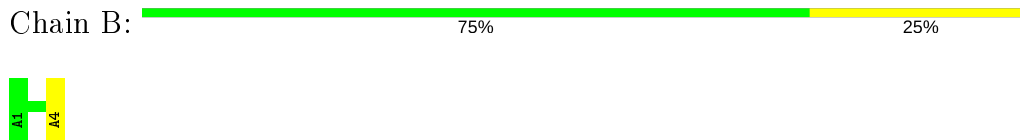
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: Staphylopin dehydrogenase



- Molecule 2: ALA-HIS-HIS-ALA



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	186.44Å 48.79Å 60.16Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.20 – 2.22 47.20 – 2.22	Depositor EDS
% Data completeness (in resolution range)	99.3 (47.20-2.22) 99.4 (47.20-2.22)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.70 (at 2.22Å)	Xtrriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.181 , 0.249 0.194 , 0.255	Depositor DCC
R_{free} test set	1438 reflections (5.14%)	wwPDB-VP
Wilson B-factor (Å ²)	46.7	Xtrriage
Anisotropy	0.041	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 31.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	3749	wwPDB-VP
Average B, all atoms (Å ²)	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.68% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: NH4, PG4, CIT

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	0.79	1/3532 (0.0%)	0.91	11/4743 (0.2%)
2	B	0.79	0/31	0.59	0/41
All	All	0.79	1/3563 (0.0%)	0.90	11/4784 (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	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	249	THR	CB-CG2	-5.41	1.34	1.52

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	289	ARG	NE-CZ-NH1	8.06	124.33	120.30
1	A	90	MSE	CA-CB-CG	-6.34	102.52	113.30
1	A	221	MSE	CG-SE-CE	5.71	111.46	98.90
1	A	357	ARG	NE-CZ-NH1	5.65	123.13	120.30
1	A	31	VAL	CB-CA-C	-5.60	100.75	111.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	412	THR	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	3476	0	3468	40	0
2	B	30	0	26	1	0
3	A	1	0	0	2	0
4	A	13	0	5	1	0
5	A	13	0	18	1	0
6	A	212	0	0	9	0
6	B	4	0	0	1	0
All	All	3749	0	3517	43	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 43 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:14:ILE:HG13	6:A:683:HOH:O	1.44	1.15
1:A:256:MSE:HE1	1:A:281:MSE:HE1	1.67	0.77
1:A:217:PRO:HB3	1:A:221:MSE:HE2	1.65	0.76
1:A:221:MSE:HE3	1:A:386:PHE:HB3	1.71	0.72
1:A:217:PRO:HB3	1:A:221:MSE:CE	2.21	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	426/465 (92%)	408 (96%)	18 (4%)	0	100	100
2	B	2/4 (50%)	2 (100%)	0	0	100	100
All	All	428/469 (91%)	410 (96%)	18 (4%)	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	391/404 (97%)	368 (94%)	23 (6%)	19	22
2	B	2/2 (100%)	2 (100%)	0	100	100
All	All	393/406 (97%)	370 (94%)	23 (6%)	19	22

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	157	GLU
1	A	194	LYS
1	A	360	SER
1	A	191	GLU
1	A	209	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	177	HIS
1	A	429	ASN
1	A	205	HIS
1	A	52	GLN
1	A	409	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](#)

Of 3 ligands modelled in this entry, 1 is modelled with single atom - leaving 2 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
5	PG4	A	503	-	12,12,12	0.73	0	11,11,11	0.53	0
4	CIT	A	502	-	3,12,12	0.70	0	3,17,17	1.80	1 (33%)

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
5	PG4	A	503	-	-	5/10/10/10	-
4	CIT	A	502	-	-	3/6/16/16	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
4	A	502	CIT	C3-C4-C5	3.04	119.86	114.98

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	502	CIT	C1-C2-C3-C4
4	A	502	CIT	C1-C2-C3-C6
5	A	503	PG4	O3-C5-C6-O4
5	A	503	PG4	O2-C3-C4-O3
5	A	503	PG4	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	503	PG4	1	0
4	A	502	CIT	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	409/465 (87%)	-0.41	0 100 100	29, 46, 74, 97	0
2	B	4/4 (100%)	0.44	0 100 100	55, 64, 76, 81	0
All	All	413/469 (88%)	-0.40	0 100 100	29, 46, 75, 97	0

There are no RSRZ outliers to report.

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(Å ²)	Q<0.9
5	PG4	A	503	13/13	0.88	0.17	63,67,71,71	0
4	CIT	A	502	13/13	0.90	0.15	47,79,92,94	0
3	NH4	A	501	1/1	0.96	0.22	34,34,34,34	0

6.5 Other polymers

There are no such residues in this entry.