



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:46 PM GMT

PDB ID : 4G3H
Title : Crystal structure of helicobacter pylori arginase
Authors : Zhang, J.; Zhang, X.; Li, D.; Hu, Y.; Zou, Q.; Wang, D.
Deposited on : 2012-07-13
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
A user guide is available at
<http://wwpdb.org/validation/2016/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 : 1.7 (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

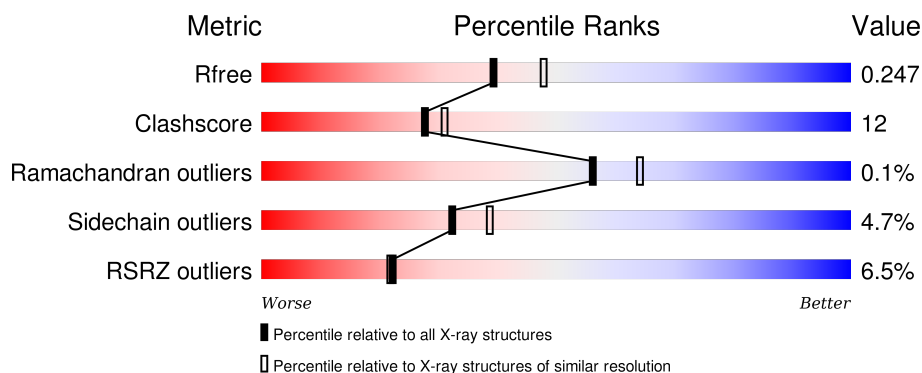
1 Overall quality at a glance

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}	91344	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (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 ≥ 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 $\leq 5\%$. 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	330	<div> <div>6%</div> <div> <div></div> <div>65%</div> <div>22%</div> <div>•</div> <div>9%</div> </div> </div>
1	B	330	<div> <div>8%</div> <div> <div></div> <div>65%</div> <div>24%</div> <div>•</div> <div>8%</div> </div> </div>
1	C	330	<div> <div>6%</div> <div> <div></div> <div>70%</div> <div>21%</div> <div>•</div> <div>6%</div> </div> </div>
1	D	330	<div> <div>5%</div> <div> <div></div> <div>66%</div> <div>22%</div> <div>•</div> <div>9%</div> </div> </div>

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
2	MN	C	501	-	-	-	X

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 10439 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 Arginase (RocF).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	301	Total	C	N	O	S	0	0	0
			2419	1539	411	454	15			
1	B	305	Total	C	N	O	S	0	0	0
			2447	1551	417	465	14			
1	C	310	Total	C	N	O	S	6	0	0
			2491	1581	424	471	15			
1	D	300	Total	C	N	O	S	0	0	0
			2399	1524	407	453	15			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	323	LEU	-	EXPRESSION TAG	UNP O25949
A	324	GLU	-	EXPRESSION TAG	UNP O25949
A	325	HIS	-	EXPRESSION TAG	UNP O25949
A	326	HIS	-	EXPRESSION TAG	UNP O25949
A	327	HIS	-	EXPRESSION TAG	UNP O25949
A	328	HIS	-	EXPRESSION TAG	UNP O25949
A	329	HIS	-	EXPRESSION TAG	UNP O25949
A	330	HIS	-	EXPRESSION TAG	UNP O25949
B	323	LEU	-	EXPRESSION TAG	UNP O25949
B	324	GLU	-	EXPRESSION TAG	UNP O25949
B	325	HIS	-	EXPRESSION TAG	UNP O25949
B	326	HIS	-	EXPRESSION TAG	UNP O25949
B	327	HIS	-	EXPRESSION TAG	UNP O25949
B	328	HIS	-	EXPRESSION TAG	UNP O25949
B	329	HIS	-	EXPRESSION TAG	UNP O25949
B	330	HIS	-	EXPRESSION TAG	UNP O25949
C	323	LEU	-	EXPRESSION TAG	UNP O25949
C	324	GLU	-	EXPRESSION TAG	UNP O25949
C	325	HIS	-	EXPRESSION TAG	UNP O25949
C	326	HIS	-	EXPRESSION TAG	UNP O25949
C	327	HIS	-	EXPRESSION TAG	UNP O25949

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	328	HIS	-	EXPRESSION TAG	UNP O25949
C	329	HIS	-	EXPRESSION TAG	UNP O25949
C	330	HIS	-	EXPRESSION TAG	UNP O25949
D	323	LEU	-	EXPRESSION TAG	UNP O25949
D	324	GLU	-	EXPRESSION TAG	UNP O25949
D	325	HIS	-	EXPRESSION TAG	UNP O25949
D	326	HIS	-	EXPRESSION TAG	UNP O25949
D	327	HIS	-	EXPRESSION TAG	UNP O25949
D	328	HIS	-	EXPRESSION TAG	UNP O25949
D	329	HIS	-	EXPRESSION TAG	UNP O25949
D	330	HIS	-	EXPRESSION TAG	UNP O25949

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	2	Total 2	Mn 2	0	0
2	A	2	Total 2	Mn 2	0	0
2	D	2	Total 2	Mn 2	0	0
2	C	2	Total 2	Mn 2	0	0

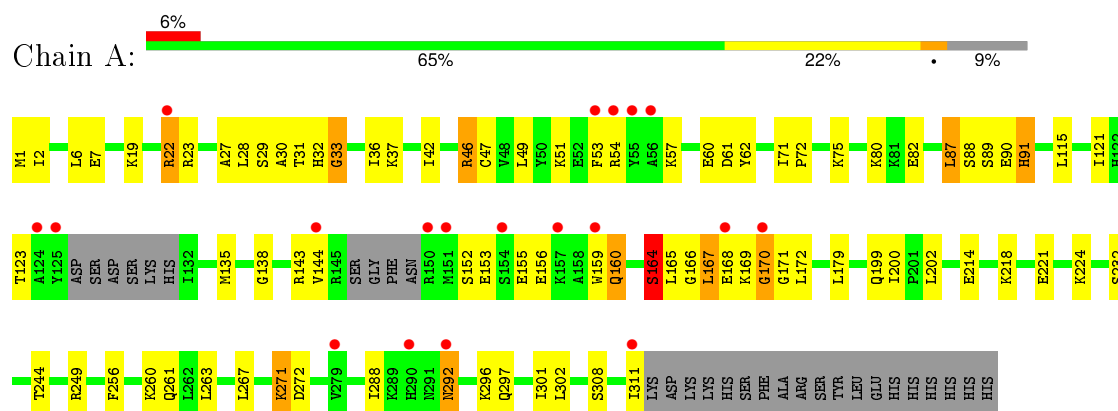
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	174	Total 174	O 174	0	0
3	B	153	Total 153	O 153	0	0
3	C	168	Total 168	O 168	0	0
3	D	180	Total 180	O 180	0	0

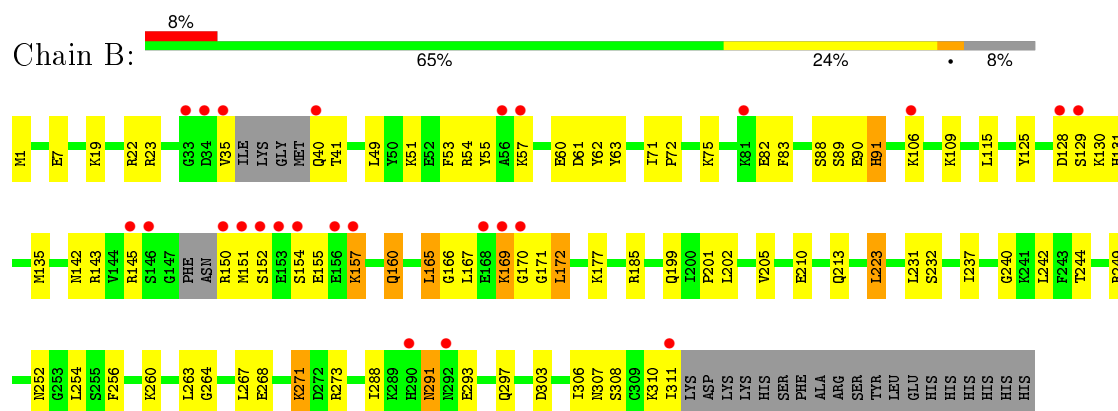
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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 ($\text{RSRZ} > 2$). 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.

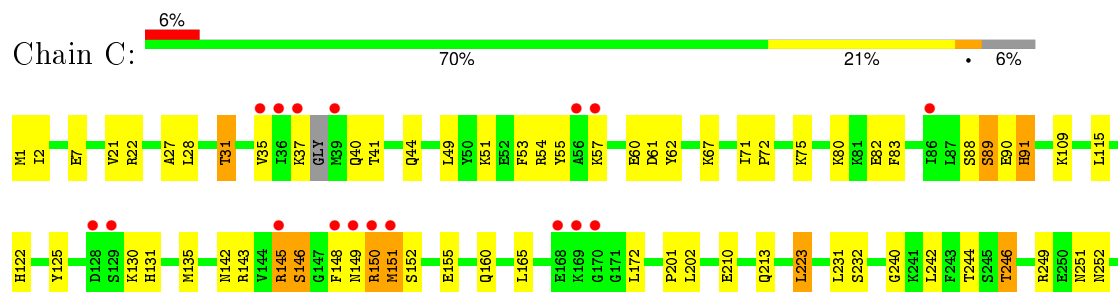
• Molecule 1: Arginase (RocF)

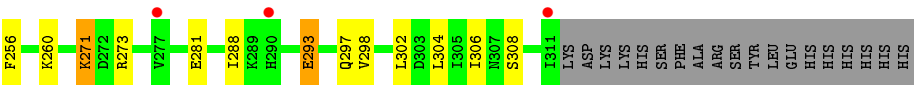


• Molecule 1: Arginase (RocF)

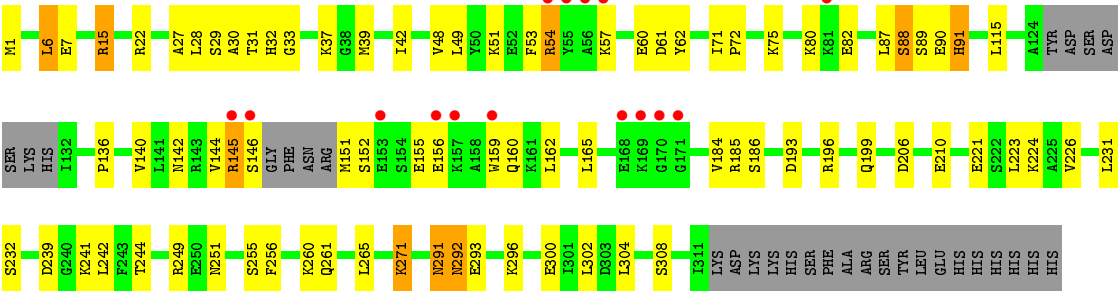


• Molecule 1: Arginase (RocF)





● Molecule 1: Arginase (RocF)



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	94.69Å 102.24Å 148.61Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.90 – 2.20 34.92 – 2.20	Depositor EDS
% Data completeness (in resolution range)	(Not available) (34.90-2.20) 99.9 (34.92-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.62 (at 2.20Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.218 , 0.230 0.208 , 0.247	Depositor DCC
R_{free} test set	7448 reflections (10.10%)	DCC
Wilson B-factor (Å ²)	23.1	Xtriage
Anisotropy	1.069	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 54.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.55$, $\langle L^2 \rangle = 0.39$	Xtriage
Outliers	7 of 73768 reflections (0.009%)	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	10439	wwPDB-VP
Average B, all atoms (Å ²)	36.0	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 52.93 % 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 4.5344e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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: MN

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	1.22	35/2453 (1.4%)	0.78	10/3289 (0.3%)
1	B	0.82	14/2482 (0.6%)	0.67	5/3329 (0.2%)
1	C	0.67	8/2528 (0.3%)	0.63	1/3391 (0.0%)
1	D	1.26	31/2432 (1.3%)	0.75	5/3261 (0.2%)
All	All	1.02	88/9895 (0.9%)	0.71	21/13270 (0.2%)

The worst 5 of 88 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	89	SER	CB-OG	-16.67	1.20	1.42
1	D	29	SER	CB-OG	-16.50	1.20	1.42
1	D	184	VAL	CB-CG2	-15.99	1.19	1.52
1	B	89	SER	CB-OG	-15.55	1.22	1.42
1	D	89	SER	C-O	-15.13	0.94	1.23

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	54	ARG	NE-CZ-NH2	-10.39	115.11	120.30
1	D	54	ARG	NE-CZ-NH1	9.65	125.13	120.30
1	B	171	GLY	N-CA-C	8.63	134.67	113.10
1	B	128	ASP	N-CA-C	-7.55	90.62	111.00
1	B	165	LEU	CB-CG-CD1	-7.07	98.97	111.00

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	2419	0	2472	57	0
1	B	2447	0	2482	66	0
1	C	2491	0	2530	62	0
1	D	2399	0	2446	56	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
3	A	174	0	0	4	0
3	B	153	0	0	7	0
3	C	168	0	0	8	0
3	D	180	0	0	4	0
All	All	10439	0	9930	233	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 12.

The worst 5 of 233 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:123:THR:H	1:A:135:MET:HE2	1.26	1.00
1:C:244:THR:HA	1:C:249:ARG:HD3	1.40	0.98
1:B:244:THR:HA	1:B:249:ARG:HD3	1.44	0.97
1:A:244:THR:HA	1:A:249:ARG:HD3	1.53	0.91
1:B:1:MET:HA	1:B:82:GLU:HG2	1.55	0.88

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	295/330 (89%)	281 (95%)	14 (5%)	0	100	100
1	B	299/330 (91%)	287 (96%)	12 (4%)	0	100	100
1	C	306/330 (93%)	285 (93%)	20 (6%)	1 (0%)	46	50
1	D	294/330 (89%)	280 (95%)	14 (5%)	0	100	100
All	All	1194/1320 (90%)	1133 (95%)	60 (5%)	1 (0%)	56	64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	150	ARG

5.3.2 Protein sidechains ⓘ

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	269/296 (91%)	259 (96%)	10 (4%)	41	50
1	B	273/296 (92%)	260 (95%)	13 (5%)	31	37
1	C	278/296 (94%)	261 (94%)	17 (6%)	23	26
1	D	267/296 (90%)	256 (96%)	11 (4%)	37	45
All	All	1087/1184 (92%)	1036 (95%)	51 (5%)	32	39

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	31	THR
1	C	145	ARG
1	D	271	LYS
1	C	91	HIS
1	C	146	SER

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	B	217	GLN
1	B	291	ASN
1	D	251	ASN
1	B	251	ASN
1	B	252	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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 [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	301/330 (91%)	0.07	19 (6%) 23 23	15, 28, 69, 127	0
1	B	305/330 (92%)	0.25	25 (8%) 14 14	15, 31, 83, 118	1 (0%)
1	C	310/330 (93%)	0.18	20 (6%) 22 22	14, 28, 80, 130	1 (0%)
1	D	300/330 (90%)	-0.12	15 (5%) 32 32	14, 27, 70, 117	0
All	All	1216/1320 (92%)	0.10	79 (6%) 22 22	14, 28, 78, 130	2 (0%)

The worst 5 of 79 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	148	PHE	8.5
1	A	125	TYR	5.8
1	D	145	ARG	5.8
1	D	55	TYR	5.7
1	A	55	TYR	5.0

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. 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, 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	LLDF	B-factors(\AA^2)	Q<0.9
2	MN	C	501	1/1	0.99	0.18	2.97	25,25,25,25	0
2	MN	B	501	1/1	0.99	0.16	1.30	26,26,26,26	0
2	MN	A	501	1/1	1.00	0.13	0.80	28,28,28,28	0
2	MN	D	501	1/1	0.99	0.10	-0.75	24,24,24,24	0
2	MN	B	500	1/1	0.97	0.07	-2.22	40,40,40,40	0
2	MN	C	500	1/1	0.98	0.08	-2.57	37,37,37,37	0
2	MN	D	500	1/1	0.97	0.06	-2.78	34,34,34,34	0
2	MN	A	500	1/1	0.98	0.05	-2.82	32,32,32,32	0

6.5 Other polymers [i](#)

There are no such residues in this entry.