



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 13, 2023 – 01:48 PM EDT

PDB ID : 4R85  
Title : Crystal structure of 5-methylcytosine deaminase from *Klebsiella pneumoniae* liganded with 5-methylcytosine  
Authors : Fedorov, A.A.; Fedorov, E.V.; Hitchcock, D.S.; Raushel, F.M.; Almo, S.C.  
Deposited on : 2014-08-29  
Resolution : 1.80 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35.1  
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

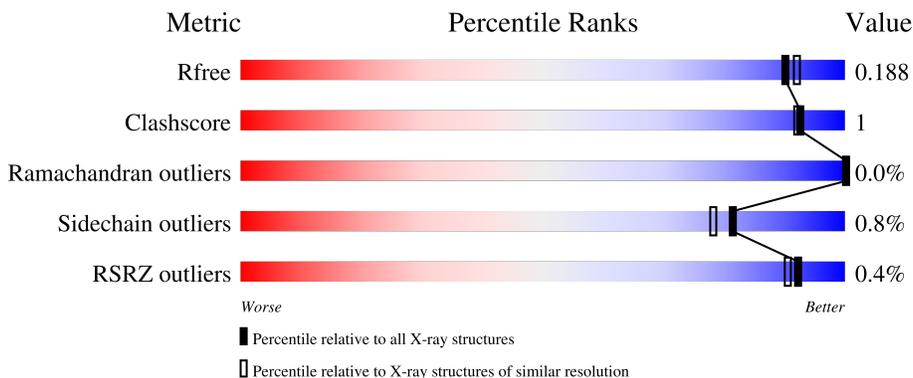
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	431	 91% 5% 5% 5%
1	B	431	 92% 5% 5% 5%
1	C	431	 90% 5% 5% 5%
1	D	431	 90% 5% 5% 5%
1	E	431	 94% 5% 5% 5%

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	431	 94%

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	17E	C	501	-	X	-	-
2	17E	F	502	-	X	-	-

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 21294 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 Cytosine deaminase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	411	Total 3217	C 2006	N 592	O 601	S 18	0	0	0
1	B	412	Total 3239	C 2018	N 596	O 607	S 18	0	2	0
1	C	412	Total 3248	C 2024	N 598	O 608	S 18	0	3	0
1	D	411	Total 3225	C 2010	N 593	O 604	S 18	0	1	0
1	E	411	Total 3217	C 2006	N 592	O 601	S 18	0	0	0
1	F	412	Total 3231	C 2014	N 594	O 605	S 18	0	1	0

There are 114 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	413	ASN	-	expression tag	UNP W8V4R8
A	414	SER	-	expression tag	UNP W8V4R8
A	415	SER	-	expression tag	UNP W8V4R8
A	416	SER	-	expression tag	UNP W8V4R8
A	417	VAL	-	expression tag	UNP W8V4R8
A	418	ASP	-	expression tag	UNP W8V4R8
A	419	LYS	-	expression tag	UNP W8V4R8
A	420	LEU	-	expression tag	UNP W8V4R8
A	421	ALA	-	expression tag	UNP W8V4R8
A	422	ALA	-	expression tag	UNP W8V4R8
A	423	ALA	-	expression tag	UNP W8V4R8
A	424	LEU	-	expression tag	UNP W8V4R8
A	425	GLU	-	expression tag	UNP W8V4R8
A	426	HIS	-	expression tag	UNP W8V4R8
A	427	HIS	-	expression tag	UNP W8V4R8
A	428	HIS	-	expression tag	UNP W8V4R8
A	429	HIS	-	expression tag	UNP W8V4R8

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	430	HIS	-	expression tag	UNP W8V4R8
A	431	HIS	-	expression tag	UNP W8V4R8
B	413	ASN	-	expression tag	UNP W8V4R8
B	414	SER	-	expression tag	UNP W8V4R8
B	415	SER	-	expression tag	UNP W8V4R8
B	416	SER	-	expression tag	UNP W8V4R8
B	417	VAL	-	expression tag	UNP W8V4R8
B	418	ASP	-	expression tag	UNP W8V4R8
B	419	LYS	-	expression tag	UNP W8V4R8
B	420	LEU	-	expression tag	UNP W8V4R8
B	421	ALA	-	expression tag	UNP W8V4R8
B	422	ALA	-	expression tag	UNP W8V4R8
B	423	ALA	-	expression tag	UNP W8V4R8
B	424	LEU	-	expression tag	UNP W8V4R8
B	425	GLU	-	expression tag	UNP W8V4R8
B	426	HIS	-	expression tag	UNP W8V4R8
B	427	HIS	-	expression tag	UNP W8V4R8
B	428	HIS	-	expression tag	UNP W8V4R8
B	429	HIS	-	expression tag	UNP W8V4R8
B	430	HIS	-	expression tag	UNP W8V4R8
B	431	HIS	-	expression tag	UNP W8V4R8
C	413	ASN	-	expression tag	UNP W8V4R8
C	414	SER	-	expression tag	UNP W8V4R8
C	415	SER	-	expression tag	UNP W8V4R8
C	416	SER	-	expression tag	UNP W8V4R8
C	417	VAL	-	expression tag	UNP W8V4R8
C	418	ASP	-	expression tag	UNP W8V4R8
C	419	LYS	-	expression tag	UNP W8V4R8
C	420	LEU	-	expression tag	UNP W8V4R8
C	421	ALA	-	expression tag	UNP W8V4R8
C	422	ALA	-	expression tag	UNP W8V4R8
C	423	ALA	-	expression tag	UNP W8V4R8
C	424	LEU	-	expression tag	UNP W8V4R8
C	425	GLU	-	expression tag	UNP W8V4R8
C	426	HIS	-	expression tag	UNP W8V4R8
C	427	HIS	-	expression tag	UNP W8V4R8
C	428	HIS	-	expression tag	UNP W8V4R8
C	429	HIS	-	expression tag	UNP W8V4R8
C	430	HIS	-	expression tag	UNP W8V4R8
C	431	HIS	-	expression tag	UNP W8V4R8
D	413	ASN	-	expression tag	UNP W8V4R8
D	414	SER	-	expression tag	UNP W8V4R8

*Continued on next page...*

*Continued from previous page...*

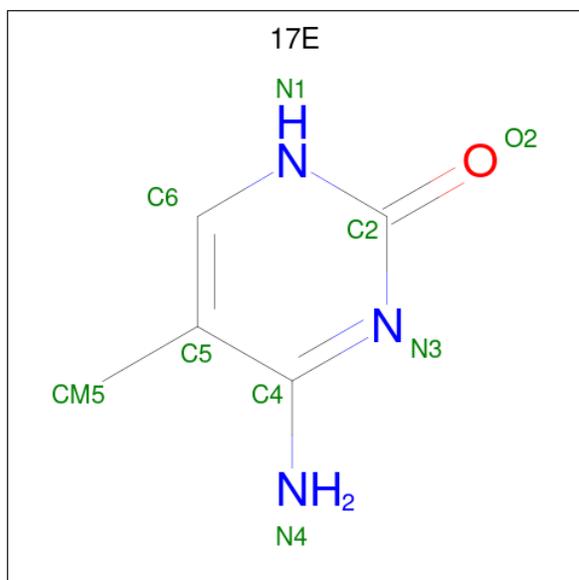
Chain	Residue	Modelled	Actual	Comment	Reference
D	415	SER	-	expression tag	UNP W8V4R8
D	416	SER	-	expression tag	UNP W8V4R8
D	417	VAL	-	expression tag	UNP W8V4R8
D	418	ASP	-	expression tag	UNP W8V4R8
D	419	LYS	-	expression tag	UNP W8V4R8
D	420	LEU	-	expression tag	UNP W8V4R8
D	421	ALA	-	expression tag	UNP W8V4R8
D	422	ALA	-	expression tag	UNP W8V4R8
D	423	ALA	-	expression tag	UNP W8V4R8
D	424	LEU	-	expression tag	UNP W8V4R8
D	425	GLU	-	expression tag	UNP W8V4R8
D	426	HIS	-	expression tag	UNP W8V4R8
D	427	HIS	-	expression tag	UNP W8V4R8
D	428	HIS	-	expression tag	UNP W8V4R8
D	429	HIS	-	expression tag	UNP W8V4R8
D	430	HIS	-	expression tag	UNP W8V4R8
D	431	HIS	-	expression tag	UNP W8V4R8
E	413	ASN	-	expression tag	UNP W8V4R8
E	414	SER	-	expression tag	UNP W8V4R8
E	415	SER	-	expression tag	UNP W8V4R8
E	416	SER	-	expression tag	UNP W8V4R8
E	417	VAL	-	expression tag	UNP W8V4R8
E	418	ASP	-	expression tag	UNP W8V4R8
E	419	LYS	-	expression tag	UNP W8V4R8
E	420	LEU	-	expression tag	UNP W8V4R8
E	421	ALA	-	expression tag	UNP W8V4R8
E	422	ALA	-	expression tag	UNP W8V4R8
E	423	ALA	-	expression tag	UNP W8V4R8
E	424	LEU	-	expression tag	UNP W8V4R8
E	425	GLU	-	expression tag	UNP W8V4R8
E	426	HIS	-	expression tag	UNP W8V4R8
E	427	HIS	-	expression tag	UNP W8V4R8
E	428	HIS	-	expression tag	UNP W8V4R8
E	429	HIS	-	expression tag	UNP W8V4R8
E	430	HIS	-	expression tag	UNP W8V4R8
E	431	HIS	-	expression tag	UNP W8V4R8
F	413	ASN	-	expression tag	UNP W8V4R8
F	414	SER	-	expression tag	UNP W8V4R8
F	415	SER	-	expression tag	UNP W8V4R8
F	416	SER	-	expression tag	UNP W8V4R8
F	417	VAL	-	expression tag	UNP W8V4R8
F	418	ASP	-	expression tag	UNP W8V4R8

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	419	LYS	-	expression tag	UNP W8V4R8
F	420	LEU	-	expression tag	UNP W8V4R8
F	421	ALA	-	expression tag	UNP W8V4R8
F	422	ALA	-	expression tag	UNP W8V4R8
F	423	ALA	-	expression tag	UNP W8V4R8
F	424	LEU	-	expression tag	UNP W8V4R8
F	425	GLU	-	expression tag	UNP W8V4R8
F	426	HIS	-	expression tag	UNP W8V4R8
F	427	HIS	-	expression tag	UNP W8V4R8
F	428	HIS	-	expression tag	UNP W8V4R8
F	429	HIS	-	expression tag	UNP W8V4R8
F	430	HIS	-	expression tag	UNP W8V4R8
F	431	HIS	-	expression tag	UNP W8V4R8

- Molecule 2 is 5-methylcytosine (three-letter code: 17E) (formula: C<sub>5</sub>H<sub>7</sub>N<sub>3</sub>O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	9	5	3	1	0	0
2	B	1	9	5	3	1	0	0
2	C	1	9	5	3	1	0	0
2	D	1	9	5	3	1	0	0
2	E	1	9	5	3	1	0	0

Continued on next page...

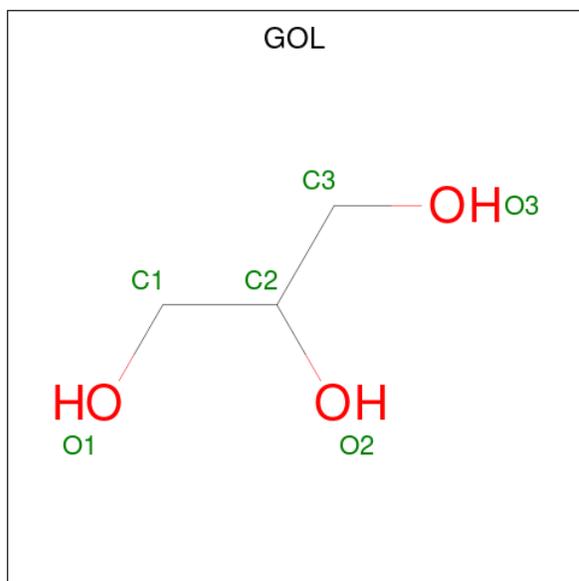
Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	F	1	Total	C	N	O	0	0
			9	5	3	1		

- Molecule 3 is FE (II) ION (three-letter code: FE2) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Fe	0	0
			1	1		
3	B	1	Total	Fe	0	0
			1	1		
3	C	1	Total	Fe	0	0
			1	1		
3	D	1	Total	Fe	0	0
			1	1		
3	E	1	Total	Fe	0	0
			1	1		
3	F	1	Total	Fe	0	0
			1	1		

- Molecule 4 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
4	A	1	Total	C	O	0	0
			6	3	3		
4	A	1	Total	C	O	0	0
			6	3	3		

Continued on next page...



*Continued from previous page...*

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 12 6 6	0	1
4	B	1	Total C O 6 3 3	0	0
4	B	1	Total C O 6 3 3	0	0
4	B	1	Total C O 6 3 3	0	0
4	B	1	Total C O 6 3 3	0	0
4	C	1	Total C O 6 3 3	0	0
4	C	1	Total C O 6 3 3	0	0
4	C	1	Total C O 6 3 3	0	0
4	D	1	Total C O 6 3 3	0	0
4	D	1	Total C O 6 3 3	0	0
4	E	1	Total C O 6 3 3	0	0
4	E	1	Total C O 6 3 3	0	0
4	E	1	Total C O 6 3 3	0	0
4	F	1	Total C O 6 3 3	0	0
4	F	1	Total C O 6 3 3	0	0
4	F	1	Total C O 6 3 3	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	294	Total O 298 298	0	4
5	B	271	Total O 277 277	0	6
5	C	287	Total O 291 291	0	4

*Continued on next page...*

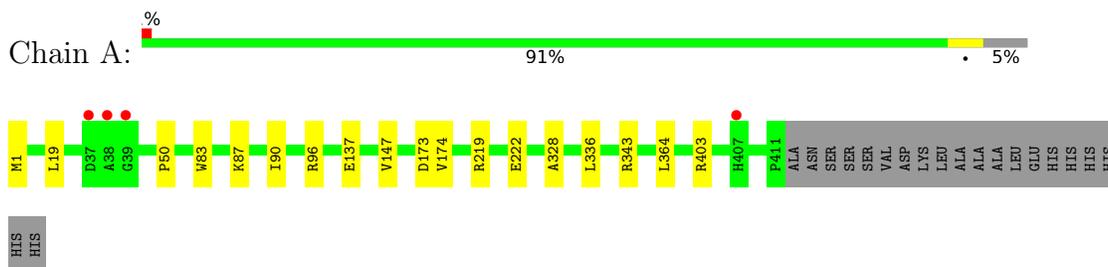
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
5	D	293	Total 300	O 300	0	7
5	E	286	Total 290	O 290	0	4
5	F	283	Total 287	O 287	0	4

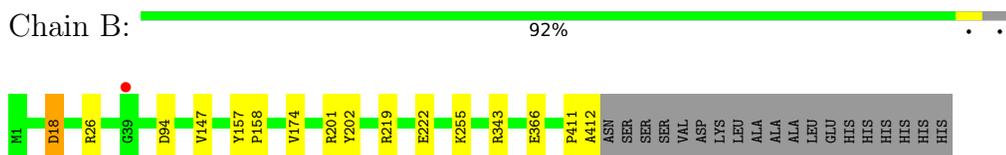
### 3 Residue-property plots [i](#)

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 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.

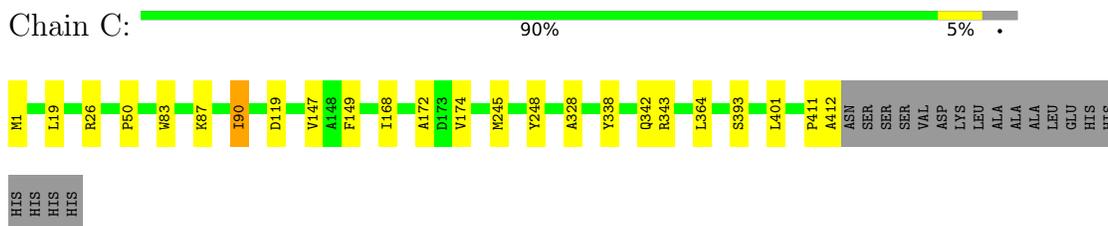
- Molecule 1: Cytosine deaminase



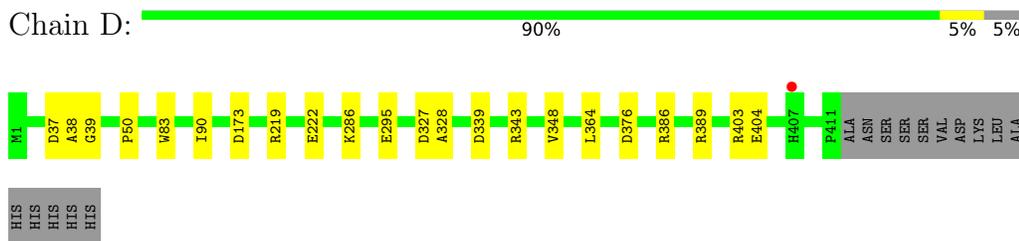
- Molecule 1: Cytosine deaminase



- Molecule 1: Cytosine deaminase



- Molecule 1: Cytosine deaminase



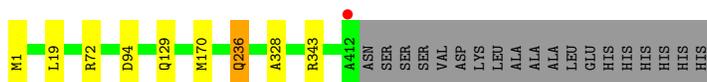
- Molecule 1: Cytosine deaminase

Chain E:  94% 5%



• Molecule 1: Cytosine deaminase

Chain F:  94%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	111.48Å 137.67Å 112.09Å 90.00° 118.03° 90.00°	Depositor
Resolution (Å)	40.17 – 1.80 40.17 – 1.80	Depositor EDS
% Data completeness (in resolution range)	96.8 (40.17-1.80) 95.1 (40.17-1.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.67 (at 1.81Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.159 , 0.187 0.159 , 0.188	Depositor DCC
$R_{free}$ test set	8065 reflections (3.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.9	Xtriage
Anisotropy	0.471	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 29.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.056 for -h-l,k,h 0.056 for l,k,-h-l 0.060 for h,-k,-h-l 0.054 for -h-l,-k,l 0.427 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	21294	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.22% of the height of the origin peak. No significant pseudotranslation is detected.*

<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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: FE2, GOL, 17E

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	0.35	0/3275	0.53	0/4436
1	B	0.35	0/3297	0.53	0/4466
1	C	0.36	0/3306	0.54	0/4477
1	D	0.36	0/3283	0.52	0/4447
1	E	0.36	0/3275	0.54	0/4436
1	F	0.36	0/3289	0.53	0/4455
All	All	0.36	0/19725	0.53	0/26717

There are no bond length outliers.

There are no bond angle outliers.

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	3217	0	3193	11	0
1	B	3239	0	3208	8	0
1	C	3248	0	3220	14	0
1	D	3225	0	3196	13	0
1	E	3217	0	3193	4	0
1	F	3231	0	3203	6	0
2	A	9	0	7	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	9	0	7	0	0
2	C	9	0	7	0	0
2	D	9	0	7	0	0
2	E	9	0	7	0	0
2	F	9	0	7	0	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
3	E	1	0	0	0	0
3	F	1	0	0	0	0
4	A	24	0	32	4	0
4	B	24	0	32	1	0
4	C	18	0	24	3	0
4	D	12	0	16	1	0
4	E	18	0	24	1	0
4	F	18	0	24	2	0
5	A	298	0	0	0	0
5	B	277	0	0	1	0
5	C	291	0	0	1	0
5	D	300	0	0	1	0
5	E	290	0	0	2	0
5	F	287	0	0	1	0
All	All	21294	0	19407	57	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 1.

The worst 5 of 57 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:1:MET:HE3	1:A:19:LEU:HD12	1.73	0.70
1:E:298:ARG:NH2	5:E:880:HOH:O	2.33	0.62
1:D:389:ARG:NH2	1:D:404:GLU:O	2.34	0.61
1:F:1:MET:N	1:F:19:LEU:O	2.32	0.59
1:C:87:LYS:HA	1:C:90:ILE:HG13	1.86	0.57

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	409/431 (95%)	395 (97%)	14 (3%)	0	100	100
1	B	412/431 (96%)	400 (97%)	12 (3%)	0	100	100
1	C	413/431 (96%)	396 (96%)	17 (4%)	0	100	100
1	D	410/431 (95%)	395 (96%)	14 (3%)	1 (0%)	47	33
1	E	409/431 (95%)	397 (97%)	12 (3%)	0	100	100
1	F	411/431 (95%)	396 (96%)	15 (4%)	0	100	100
All	All	2464/2586 (95%)	2379 (97%)	84 (3%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	38	ALA

### 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	339/355 (96%)	336 (99%)	3 (1%)	78	75
1	B	341/355 (96%)	338 (99%)	3 (1%)	78	75
1	C	342/355 (96%)	340 (99%)	2 (1%)	86	84
1	D	340/355 (96%)	337 (99%)	3 (1%)	78	75
1	E	339/355 (96%)	337 (99%)	2 (1%)	86	84
1	F	340/355 (96%)	337 (99%)	3 (1%)	78	75

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	2041/2130 (96%)	2025 (99%)	16 (1%)	81	78

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	236	GLN
1	F	94	ASP
1	D	173	ASP
1	E	343	ARG
1	C	343	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	217	GLN
1	B	241	HIS
1	C	241	HIS

### 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](#)

Of 31 ligands modelled in this entry, 6 are monoatomic - leaving 25 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
4	GOL	E	504	-	5,5,5	0.34	0	5,5,5	0.37	0
4	GOL	C	504	-	5,5,5	0.35	0	5,5,5	0.49	0
4	GOL	F	504	-	5,5,5	0.30	0	5,5,5	0.55	0
2	17E	B	501	-	8,9,9	3.12	3 (37%)	11,12,12	2.16	5 (45%)
4	GOL	E	505	-	5,5,5	0.42	0	5,5,5	0.33	0
2	17E	D	501	-	8,9,9	3.25	4 (50%)	11,12,12	1.73	4 (36%)
4	GOL	C	505	-	5,5,5	0.39	0	5,5,5	0.30	0
4	GOL	B	504	-	5,5,5	0.34	0	5,5,5	0.46	0
4	GOL	D	504	-	5,5,5	0.37	0	5,5,5	0.40	0
4	GOL	A	503	-	5,5,5	0.39	0	5,5,5	0.25	0
4	GOL	A	505[A]	-	5,5,5	0.38	0	5,5,5	0.28	0
4	GOL	A	505[B]	-	5,5,5	0.36	0	5,5,5	0.42	0
4	GOL	F	505	-	5,5,5	0.33	0	5,5,5	0.60	0
4	GOL	F	501	-	5,5,5	0.40	0	5,5,5	0.71	0
4	GOL	B	506	-	5,5,5	0.38	0	5,5,5	0.32	0
4	GOL	E	503	-	5,5,5	0.40	0	5,5,5	0.22	0
4	GOL	D	503	-	5,5,5	0.36	0	5,5,5	0.46	0
4	GOL	B	505	-	5,5,5	0.37	0	5,5,5	0.39	0
4	GOL	C	503	-	5,5,5	0.35	0	5,5,5	0.52	0
2	17E	F	502	-	8,9,9	3.13	4 (50%)	11,12,12	1.97	5 (45%)
2	17E	A	501	-	8,9,9	3.30	4 (50%)	11,12,12	1.54	2 (18%)
2	17E	C	501	-	8,9,9	3.39	4 (50%)	11,12,12	1.63	5 (45%)
4	GOL	B	503	-	5,5,5	0.39	0	5,5,5	0.37	0
2	17E	E	501	-	8,9,9	3.28	4 (50%)	11,12,12	2.00	4 (36%)
4	GOL	A	504	-	5,5,5	0.33	0	5,5,5	0.16	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
4	GOL	E	504	-	-	2/4/4/4	-
4	GOL	C	504	-	-	2/4/4/4	-
4	GOL	F	504	-	-	0/4/4/4	-
2	17E	B	501	-	-	-	0/1/1/1
4	GOL	E	505	-	-	4/4/4/4	-
4	GOL	C	505	-	-	0/4/4/4	-

*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	GOL	D	504	-	-	2/4/4/4	-
4	GOL	B	504	-	-	0/4/4/4	-
2	17E	D	501	-	-	-	0/1/1/1
4	GOL	A	503	-	-	0/4/4/4	-
4	GOL	A	505[A]	-	-	0/4/4/4	-
4	GOL	F	505	-	-	1/4/4/4	-
4	GOL	A	505[B]	-	-	2/4/4/4	-
4	GOL	F	501	-	-	3/4/4/4	-
4	GOL	B	506	-	-	2/4/4/4	-
4	GOL	E	503	-	-	0/4/4/4	-
4	GOL	D	503	-	-	1/4/4/4	-
4	GOL	B	505	-	-	0/4/4/4	-
4	GOL	C	503	-	-	0/4/4/4	-
2	17E	F	502	-	-	-	0/1/1/1
2	17E	A	501	-	-	-	0/1/1/1
2	17E	C	501	-	-	-	0/1/1/1
4	GOL	B	503	-	-	0/4/4/4	-
2	17E	E	501	-	-	-	0/1/1/1
4	GOL	A	504	-	-	2/4/4/4	-

The worst 5 of 23 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	501	17E	O2-C2	7.42	1.38	1.24
2	A	501	17E	O2-C2	7.21	1.37	1.24
2	D	501	17E	O2-C2	7.21	1.37	1.24
2	E	501	17E	O2-C2	7.20	1.37	1.24
2	F	502	17E	O2-C2	6.82	1.36	1.24

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	501	17E	C6-N1-C2	-3.87	119.09	122.02
2	E	501	17E	C6-N1-C2	-3.68	119.23	122.02
2	B	501	17E	N1-C2-N3	3.25	122.86	118.67
2	F	502	17E	N1-C2-N3	3.12	122.69	118.67
2	E	501	17E	N1-C2-N3	2.99	122.52	118.67

There are no chirality outliers.

5 of 21 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	506	GOL	O1-C1-C2-C3
4	D	504	GOL	O1-C1-C2-C3
4	E	505	GOL	O1-C1-C2-C3
4	E	505	GOL	C1-C2-C3-O3
4	A	504	GOL	O1-C1-C2-O2

There are no ring outliers.

7 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	E	505	GOL	1	0
4	A	505[A]	GOL	3	0
4	A	505[B]	GOL	1	0
4	F	501	GOL	2	0
4	D	503	GOL	1	0
4	C	503	GOL	3	0
4	B	503	GOL	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	411/431 (95%)	-0.79	4 (0%) 82 80	11, 19, 41, 78	0
1	B	412/431 (95%)	-0.77	1 (0%) 95 93	11, 18, 44, 69	0
1	C	412/431 (95%)	-0.80	0 100 100	10, 17, 36, 60	0
1	D	411/431 (95%)	-0.81	1 (0%) 95 93	10, 19, 39, 77	0
1	E	411/431 (95%)	-0.79	2 (0%) 91 89	10, 18, 40, 66	0
1	F	412/431 (95%)	-0.80	1 (0%) 95 93	10, 18, 37, 64	0
All	All	2469/2586 (95%)	-0.79	9 (0%) 92 90	10, 18, 40, 78	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	37	ASP	3.8
1	A	38	ALA	3.0
1	B	39	GLY	2.6
1	F	412	ALA	2.4
1	E	38	ALA	2.3

### 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 monosaccharides 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
4	GOL	E	505	6/6	0.82	0.20	48,51,53,54	0
4	GOL	A	505[B]	6/6	0.85	0.14	26,34,34,38	6
4	GOL	A	505[A]	6/6	0.85	0.14	31,34,34,37	6
4	GOL	B	506	6/6	0.88	0.21	46,53,57,63	0
4	GOL	A	504	6/6	0.88	0.11	32,45,48,48	0
4	GOL	F	501	6/6	0.89	0.10	30,33,39,41	0
4	GOL	B	503	6/6	0.91	0.16	39,41,45,47	0
4	GOL	C	503	6/6	0.93	0.12	31,38,41,43	0
4	GOL	D	503	6/6	0.94	0.11	24,35,41,42	0
4	GOL	E	504	6/6	0.94	0.10	19,38,41,42	0
4	GOL	F	505	6/6	0.95	0.08	20,34,38,44	0
4	GOL	C	505	6/6	0.96	0.10	19,37,42,44	0
4	GOL	C	504	6/6	0.96	0.07	21,28,32,34	0
4	GOL	E	503	6/6	0.97	0.06	15,21,24,26	0
4	GOL	B	505	6/6	0.97	0.07	19,20,25,26	0
2	17E	B	501	9/9	0.97	0.06	13,13,15,17	0
4	GOL	B	504	6/6	0.97	0.10	20,25,34,37	0
4	GOL	D	504	6/6	0.97	0.07	17,21,28,31	0
4	GOL	A	503	6/6	0.98	0.06	15,21,25,26	0
2	17E	A	501	9/9	0.98	0.05	12,13,19,20	0
2	17E	C	501	9/9	0.98	0.06	14,16,19,19	0
2	17E	D	501	9/9	0.98	0.05	13,15,19,19	0
2	17E	E	501	9/9	0.98	0.05	10,12,18,18	0
4	GOL	F	504	6/6	0.98	0.06	19,26,29,31	0
2	17E	F	502	9/9	0.98	0.06	13,16,17,20	0
3	FE2	F	503	1/1	0.99	0.09	17,17,17,17	0
3	FE2	D	502	1/1	0.99	0.07	19,19,19,19	0
3	FE2	B	502	1/1	1.00	0.07	17,17,17,17	0
3	FE2	C	502	1/1	1.00	0.08	18,18,18,18	0
3	FE2	A	502	1/1	1.00	0.07	19,19,19,19	0
3	FE2	E	502	1/1	1.00	0.07	17,17,17,17	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.