



## wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Jul 5, 2018 – 12:47 PM EDT

PDB ID : 6D05  
EMDB ID: : EMD-7785  
Title : Cryo-EM structure of a Plasmodium vivax invasion complex essential for entry into human reticulocytes; two molecules of parasite ligand, subclass 2.  
Authors : Gruszczyk, J.; Huang, R.K.; Hong, C.; Yu, Z.; Tham, W.H.  
Deposited on : 2018-04-10  
Resolution : 3.80 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report  
for a publicly released PDB/EMDB 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/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Mogul : 1.7.3 (157068), CSD as539be (2018)  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20031172



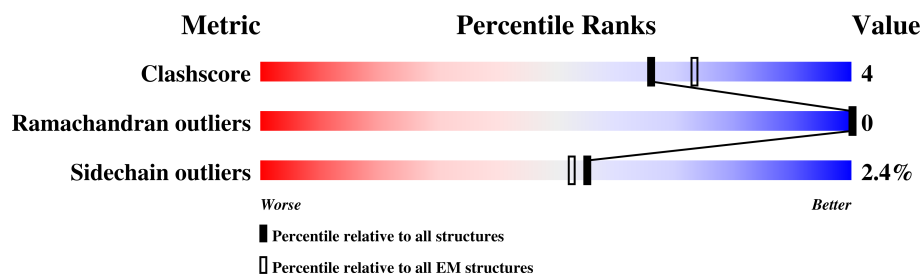
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.80 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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	659	84% 13% .
1	B	659	85% 12% .
2	C	698	84% 13% .
2	D	698	84% 13% .
3	E	820	48% 8% . 43%
3	F	820	48% 8% . 43%



## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 28748 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Transferrin receptor protein 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	641	Total	C	N	O	S	0	0
			5081	3260	855	952	14		
1	B	641	Total	C	N	O	S	0	0
			5081	3260	855	952	14		

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	ALA	-	expression tag	UNP P02786
A	103	ASP	-	expression tag	UNP P02786
A	104	PRO	-	expression tag	UNP P02786
A	105	HIS	-	expression tag	UNP P02786
A	106	HIS	-	expression tag	UNP P02786
A	107	HIS	-	expression tag	UNP P02786
A	108	HIS	-	expression tag	UNP P02786
A	109	HIS	-	expression tag	UNP P02786
A	110	HIS	-	expression tag	UNP P02786
A	111	SER	-	expression tag	UNP P02786
A	112	SER	-	expression tag	UNP P02786
A	113	GLY	-	expression tag	UNP P02786
A	114	ILE	-	expression tag	UNP P02786
A	115	GLU	-	expression tag	UNP P02786
A	116	GLY	-	expression tag	UNP P02786
A	117	ARG	-	expression tag	UNP P02786
A	118	GLY	-	expression tag	UNP P02786
A	119	GLU	-	expression tag	UNP P02786
A	120	PHE	-	expression tag	UNP P02786
A	142	SER	GLY	variant	UNP P02786
B	102	ALA	-	expression tag	UNP P02786
B	103	ASP	-	expression tag	UNP P02786
B	104	PRO	-	expression tag	UNP P02786
B	105	HIS	-	expression tag	UNP P02786
B	106	HIS	-	expression tag	UNP P02786
B	107	HIS	-	expression tag	UNP P02786

*Continued on next page...*



*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	108	HIS	-	expression tag	UNP P02786
B	109	HIS	-	expression tag	UNP P02786
B	110	HIS	-	expression tag	UNP P02786
B	111	SER	-	expression tag	UNP P02786
B	112	SER	-	expression tag	UNP P02786
B	113	GLY	-	expression tag	UNP P02786
B	114	ILE	-	expression tag	UNP P02786
B	115	GLU	-	expression tag	UNP P02786
B	116	GLY	-	expression tag	UNP P02786
B	117	ARG	-	expression tag	UNP P02786
B	118	GLY	-	expression tag	UNP P02786
B	119	GLU	-	expression tag	UNP P02786
B	120	PHE	-	expression tag	UNP P02786
B	142	SER	GLY	variant	UNP P02786

- Molecule 2 is a protein called Serotransferrin.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	679	Total	C	N	O	S	0	0
			5266	3305	912	1002	47		
2	D	679	Total	C	N	O	S	0	0
			5266	3305	912	1002	47		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	429	VAL	ILE	variant	UNP P02787
D	429	VAL	ILE	variant	UNP P02787

- Molecule 3 is a protein called Reticulocyte binding protein 2, putative.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	E	466	Total	C	N	O	S	0	0
			3904	2495	650	749	10		
3	F	466	Total	C	N	O	S	0	0
			3904	2495	650	749	10		

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	150	GLY	-	expression tag	UNP A5K736
E	151	ALA	-	expression tag	UNP A5K736

*Continued on next page...*



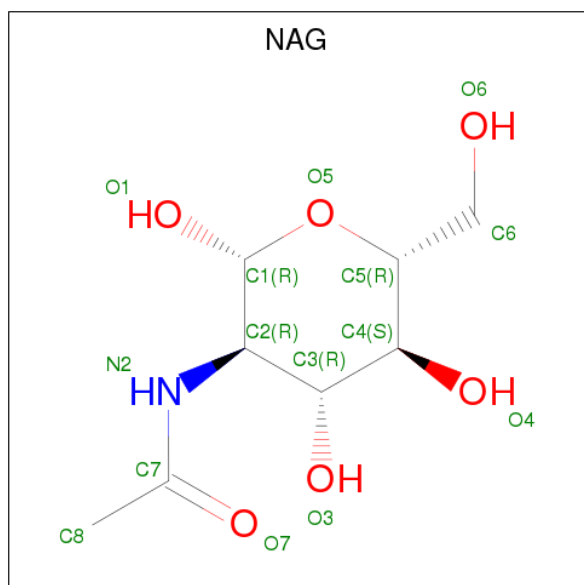
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	152	MET	-	expression tag	UNP A5K736
E	153	GLY	-	expression tag	UNP A5K736
E	154	SER	-	expression tag	UNP A5K736
E	168	SER	ILE	variant	UNP A5K736
F	150	GLY	-	expression tag	UNP A5K736
F	151	ALA	-	expression tag	UNP A5K736
F	152	MET	-	expression tag	UNP A5K736
F	153	GLY	-	expression tag	UNP A5K736
F	154	SER	-	expression tag	UNP A5K736
F	168	SER	ILE	variant	UNP A5K736

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	AltConf
4	B	1	Total Ca 1 1	0
4	A	1	Total Ca 1 1	0

- Molecule 5 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



Mol	Chain	Residues	Atoms	AltConf
5	A	1	Total C N O 70 40 5 25	0
5	A	1	Total C N O 70 40 5 25	0

Continued on next page...



*Continued from previous page...*

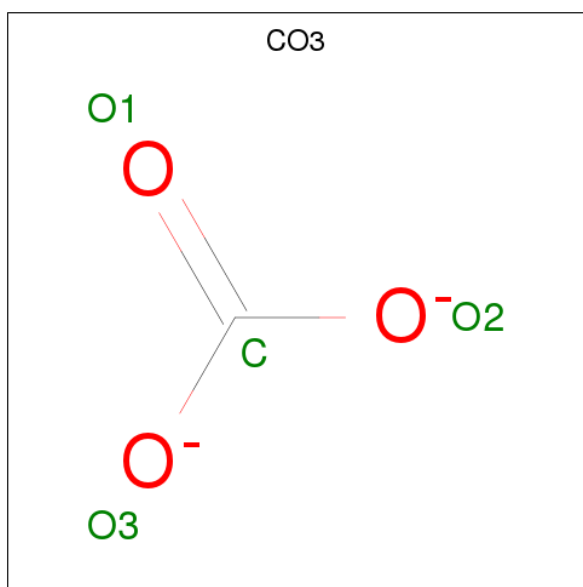
Mol	Chain	Residues	Atoms				AltConf
5	A	1	Total	C	N	O	0
			70	40	5	25	
5	A	1	Total	C	N	O	0
			70	40	5	25	
5	A	1	Total	C	N	O	0
			70	40	5	25	
5	B	1	Total	C	N	O	0
			70	40	5	25	
5	B	1	Total	C	N	O	0
			70	40	5	25	
5	B	1	Total	C	N	O	0
			70	40	5	25	
5	B	1	Total	C	N	O	0
			70	40	5	25	
5	B	1	Total	C	N	O	0
			70	40	5	25	
5	C	1	Total	C	N	O	0
			42	24	3	15	
5	C	1	Total	C	N	O	0
			42	24	3	15	
5	C	1	Total	C	N	O	0
			42	24	3	15	
5	D	1	Total	C	N	O	0
			42	24	3	15	
5	D	1	Total	C	N	O	0
			42	24	3	15	
5	D	1	Total	C	N	O	0
			42	24	3	15	

- Molecule 6 is FE (III) ION (three-letter code: FE) (formula: Fe).

Mol	Chain	Residues	Atoms		AltConf
6	D	2	Total	Fe	0
			2	2	
6	C	2	Total	Fe	0
			2	2	

- Molecule 7 is CARBONATE ION (three-letter code: CO3) (formula: CO<sub>3</sub>).





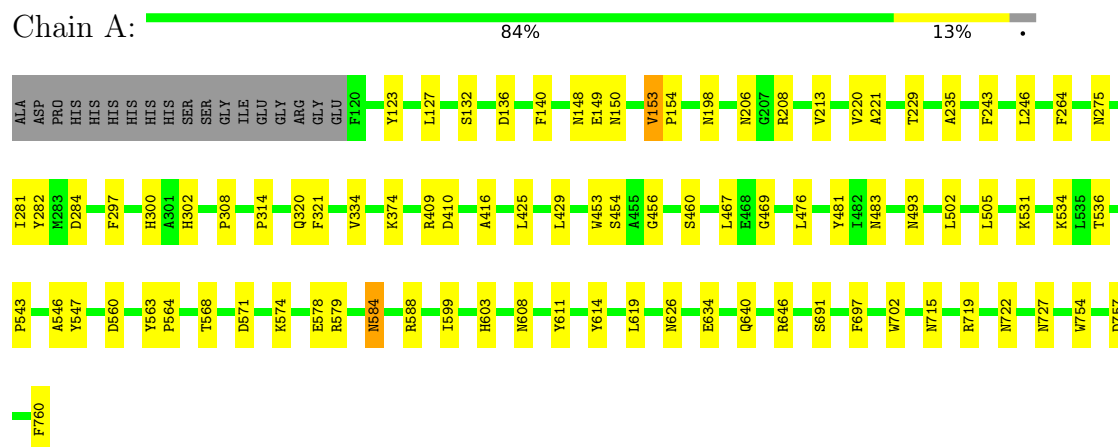
Mol	Chain	Residues	Atoms			AltConf
7	C	1	Total	C	O	0
			8	2	6	
7	C	1	Total	C	O	0
			8	2	6	
7	D	1	Total	C	O	0
			8	2	6	
7	D	1	Total	C	O	0
			8	2	6	



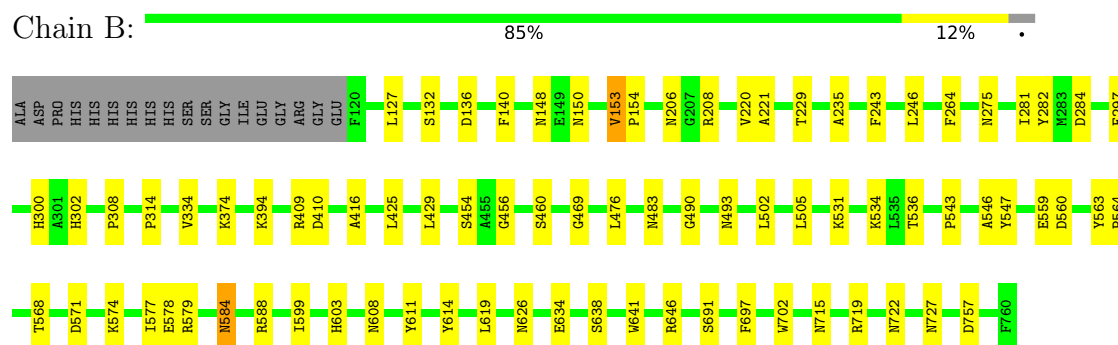
### 3 Residue-property plots

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

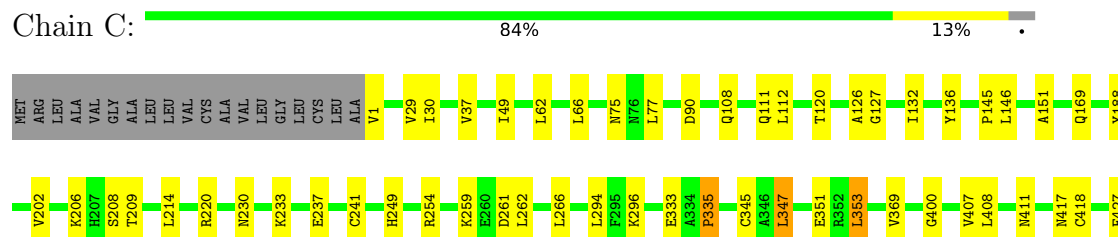
#### • Molecule 1: Transferrin receptor protein 1



#### • Molecule 1: Transferrin receptor protein 1



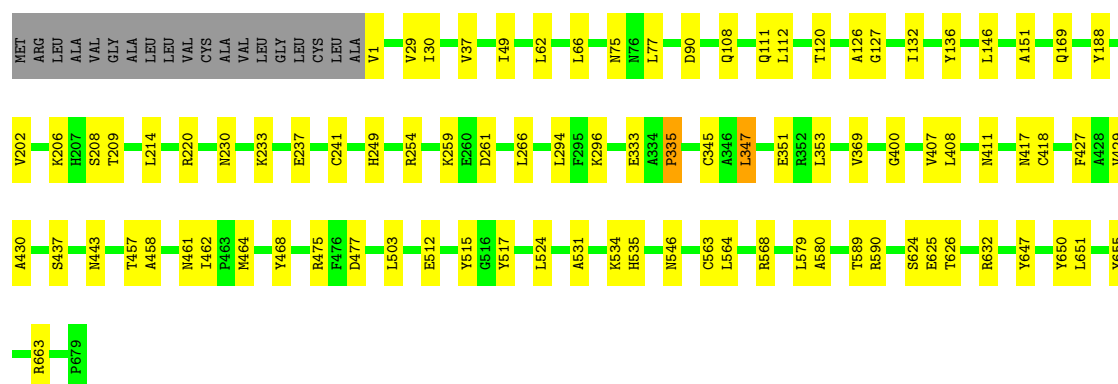
#### • Molecule 2: Serotransferrin





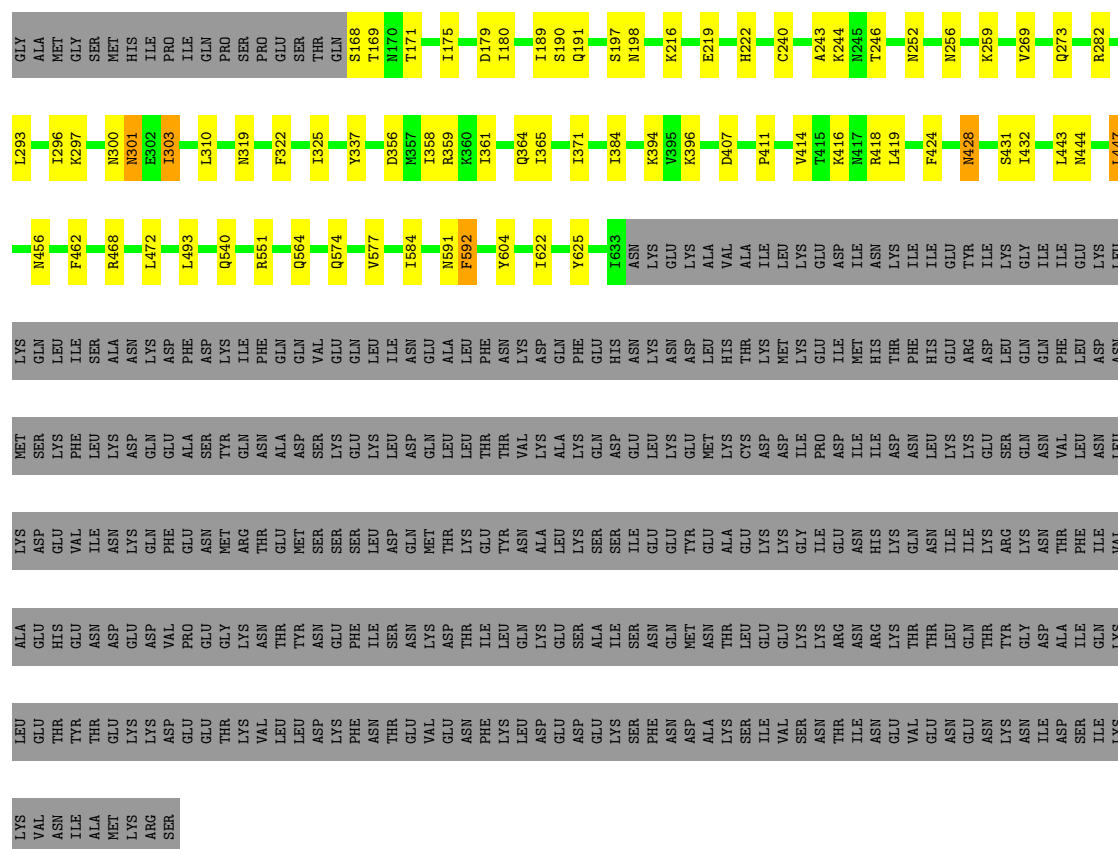
- Molecule 2: Serotransferrin

Chain D:  84% 13% .



- Molecule 3: Reticulocyte binding protein 2, putative

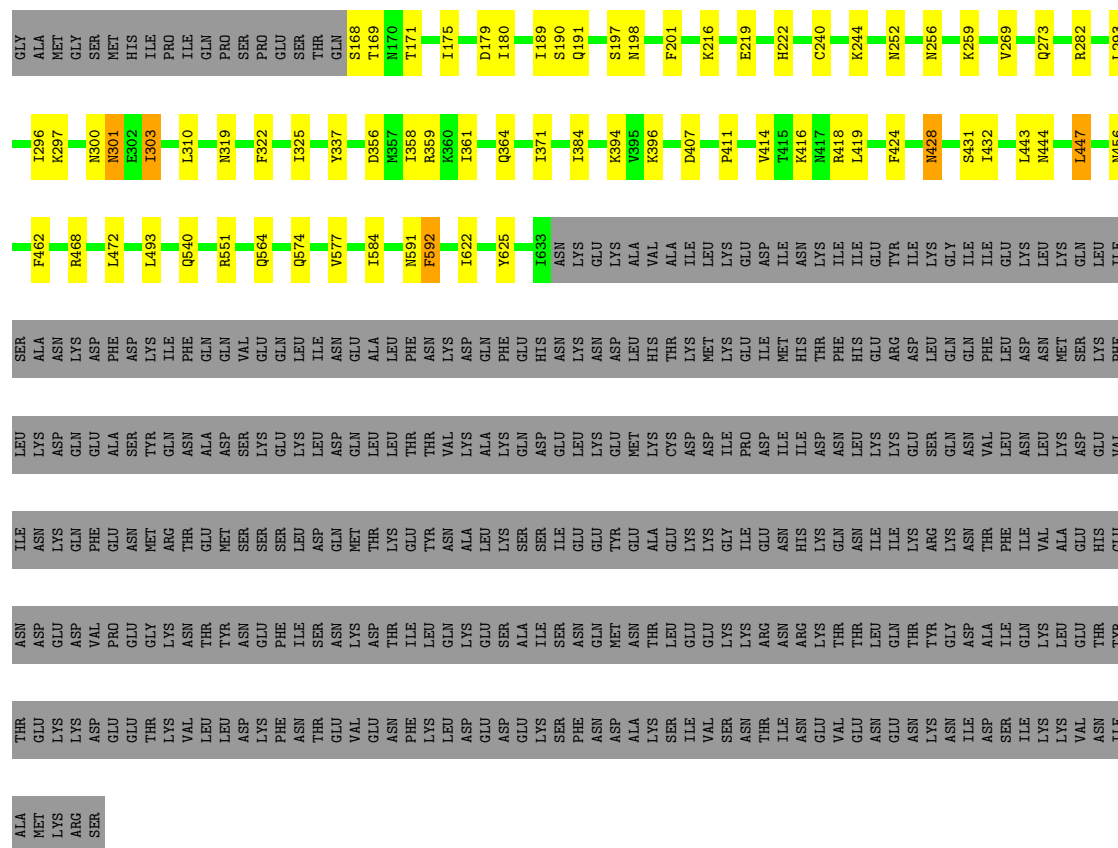
Chain E:  48% 8% 43%





- Molecule 3: Reticulocyte binding protein 2, putative

Chain F:  48% 8% . 43%





## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C2	Depositor
Number of particles used	302858	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	80	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor



## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CO3, FE, CA, NAG

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	$\# Z  > 2$	RMSZ	$\# Z  > 2$
1	A	0.52	0/5203	0.59	1/7053 (0.0%)
1	B	0.52	0/5203	0.59	1/7053 (0.0%)
2	C	0.43	0/5386	0.60	4/7280 (0.1%)
2	D	0.43	0/5386	0.60	4/7280 (0.1%)
3	E	0.38	0/3973	0.56	1/5337 (0.0%)
3	F	0.38	0/3973	0.56	1/5337 (0.0%)
All	All	0.45	0/29124	0.59	12/39340 (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
2	C	0	2
2	D	0	2
3	E	0	1
3	F	0	1
All	All	0	8

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	476	LEU	CA-CB-CG	6.02	129.14	115.30
1	A	476	LEU	CA-CB-CG	6.00	129.11	115.30
2	D	503	LEU	CA-CB-CG	5.78	128.60	115.30
2	C	503	LEU	CA-CB-CG	5.77	128.57	115.30

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
3	F	447	LEU	CA-CB-CG	5.22	127.31	115.30

There are no chirality outliers.

5 of 8 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	235	ALA	Peptide
1	B	235	ALA	Peptide
2	C	1	VAL	Peptide
2	C	335	PRO	Peptide
2	D	1	VAL	Peptide

## 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	5081	0	5010	47	0
1	B	5081	0	5010	41	0
2	C	5266	0	5084	52	0
2	D	5266	0	5084	47	0
3	E	3904	0	3918	37	0
3	F	3904	0	3918	36	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	70	0	63	1	0
5	B	70	0	63	0	0
5	C	42	0	38	0	0
5	D	42	0	38	0	0
6	C	2	0	0	0	0
6	D	2	0	0	0	0
7	C	8	0	0	1	0
7	D	8	0	0	1	0
All	All	28748	0	28226	248	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 4.

The worst 5 of 248 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:B:153:VAL:HG22	3:F:592:PHE:HE1	1.45	0.79
1:A:425:LEU:O	1:A:429:LEU:HB2	1.86	0.76
1:B:425:LEU:O	1:B:429:LEU:HB2	1.86	0.76
3:F:269:VAL:HG12	3:F:273:GLN:HE22	1.58	0.67
3:E:269:VAL:HG12	3:E:273:GLN:HE22	1.58	0.67

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 PDB entries followed by that with respect to all EM entries.

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	639/659 (97%)	605 (95%)	34 (5%)	0	100	100
1	B	639/659 (97%)	604 (94%)	35 (6%)	0	100	100
2	C	677/698 (97%)	638 (94%)	39 (6%)	0	100	100
2	D	677/698 (97%)	638 (94%)	39 (6%)	0	100	100
3	E	464/820 (57%)	456 (98%)	8 (2%)	0	100	100
3	F	464/820 (57%)	456 (98%)	8 (2%)	0	100	100
All	All	3560/4354 (82%)	3397 (95%)	163 (5%)	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 PDB entries followed by that with respect to all EM entries.

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	550/564 (98%)	532 (97%)	18 (3%)	41	72
1	B	550/564 (98%)	533 (97%)	17 (3%)	43	72
2	C	572/585 (98%)	566 (99%)	6 (1%)	78	89
2	D	572/585 (98%)	566 (99%)	6 (1%)	78	89
3	E	442/775 (57%)	428 (97%)	14 (3%)	42	72
3	F	442/775 (57%)	428 (97%)	14 (3%)	42	72
All	All	3128/3848 (81%)	3053 (98%)	75 (2%)	55	77

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	722	ASN
2	D	254	ARG
3	F	456	ASN
1	B	727	ASN
2	C	443	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 53 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	443	ASN
2	D	417	ASN
3	F	444	ASN
2	C	461	ASN
2	D	207	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.



## 5.6 Ligand geometry

Of 26 ligands modelled in this entry, 6 are monoatomic - leaving 20 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	NAG	A	802	1	14,14,15	1.08	1 (7%)	17,19,21	1.80	2 (11%)
5	NAG	A	803	1,5	14,14,15	0.76	1 (7%)	17,19,21	1.08	1 (5%)
5	NAG	A	804	5	14,14,15	0.35	0	17,19,21	0.42	0
5	NAG	A	805	1,5	14,14,15	1.15	1 (7%)	17,19,21	1.29	4 (23%)
5	NAG	A	806	5	14,14,15	0.52	0	17,19,21	0.54	0
5	NAG	B	802	1	14,14,15	0.55	0	17,19,21	0.44	0
5	NAG	B	803	1,5	14,14,15	1.24	2 (14%)	17,19,21	1.35	3 (17%)
5	NAG	B	804	5	14,14,15	0.32	0	17,19,21	0.55	0
5	NAG	B	805	1,5	14,14,15	0.74	1 (7%)	17,19,21	1.75	2 (11%)
5	NAG	B	806	5	14,14,15	0.37	0	17,19,21	0.56	0
7	CO3	C	702	6	0,3,3	0.00	-	0,3,3	0.00	-
7	CO3	C	704	6	0,3,3	0.00	-	0,3,3	0.00	-
5	NAG	C	705	2,5	14,14,15	0.28	0	17,19,21	0.78	0
5	NAG	C	706	5	14,14,15	0.39	0	17,19,21	0.74	1 (5%)
5	NAG	C	707	2	14,14,15	0.50	0	17,19,21	1.30	1 (5%)
7	CO3	D	702	6	0,3,3	0.00	-	0,3,3	0.00	-
7	CO3	D	704	6	0,3,3	0.00	-	0,3,3	0.00	-
5	NAG	D	705	2,5	14,14,15	0.53	0	17,19,21	1.24	2 (11%)
5	NAG	D	706	5	14,14,15	1.12	2 (14%)	17,19,21	2.19	2 (11%)
5	NAG	D	707	2	14,14,15	0.45	0	17,19,21	0.56	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
5	NAG	A	802	1	-	0/6/23/26	0/1/1/1
5	NAG	A	803	1,5	-	0/6/23/26	0/1/1/1

*Continued on next page...*



*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	NAG	A	804	5	-	0/6/23/26	0/1/1/1
5	NAG	A	805	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	806	5	-	0/6/23/26	0/1/1/1
5	NAG	B	802	1	-	0/6/23/26	0/1/1/1
5	NAG	B	803	1,5	-	0/6/23/26	0/1/1/1
5	NAG	B	804	5	-	0/6/23/26	0/1/1/1
5	NAG	B	805	1,5	-	0/6/23/26	0/1/1/1
5	NAG	B	806	5	-	0/6/23/26	0/1/1/1
7	CO3	C	702	6	-	0/0/0/0	0/0/0/0
7	CO3	C	704	6	-	0/0/0/0	0/0/0/0
5	NAG	C	705	2,5	-	0/6/23/26	0/1/1/1
5	NAG	C	706	5	-	0/6/23/26	0/1/1/1
5	NAG	C	707	2	-	0/6/23/26	0/1/1/1
7	CO3	D	702	6	-	0/0/0/0	0/0/0/0
7	CO3	D	704	6	-	0/0/0/0	0/0/0/0
5	NAG	D	705	2,5	-	0/6/23/26	0/1/1/1
5	NAG	D	706	5	-	0/6/23/26	0/1/1/1
5	NAG	D	707	2	-	0/6/23/26	0/1/1/1

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	803	NAG	O5-C1	-3.87	1.37	1.43
5	A	805	NAG	O5-C1	-3.67	1.37	1.43
5	A	803	NAG	O5-C1	-2.63	1.39	1.43
5	B	803	NAG	C1-C2	-2.23	1.49	1.52
5	D	706	NAG	O5-C1	2.49	1.47	1.43

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	803	NAG	O5-C5-C6	-2.08	103.85	107.15
5	A	805	NAG	C1-C2-N2	2.00	113.91	110.49
5	A	805	NAG	C3-C4-C5	2.11	114.02	110.24
5	B	803	NAG	C3-C4-C5	2.14	114.06	110.24
5	C	706	NAG	C1-O5-C5	2.20	115.22	112.19

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:



Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	803	NAG	1	0
7	C	704	CO3	1	0
7	D	704	CO3	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.