wwPDB X-ray Structure Validation Summary Report (i)

Nov 22, 2023 – 09:25 PM JST

PDB ID	:	7EZT
Title	:	The structure and functional mechanism of nucleotide regulated acetylhexosaminidase Am2136 from Akkermansia muciniphila
Authors	:	Bao, R.; Li, C.C.; Tang, X.Y.; Zhu, Y.B.; Song, Y.J.; Zhao, N.L.; Huang, Q.; Mou, X.Y.; Luo, G.H.; Liu, T.G.
Deposited on	:	2021-06-02
Resolution	:	2.81 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 (i) 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 (i)) were used in the production of this report:

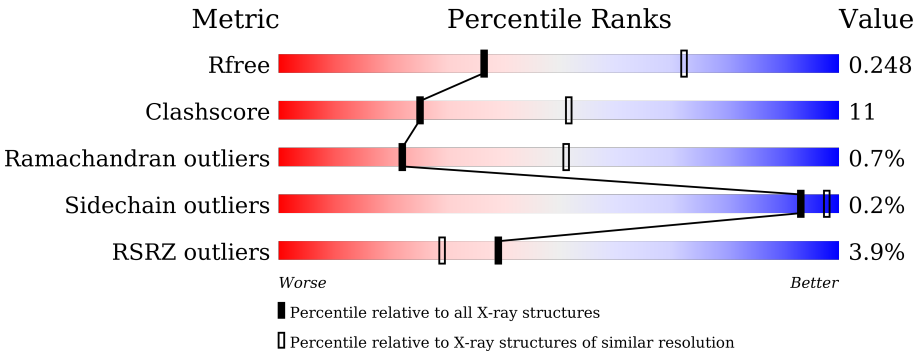
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.81 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	3617 (2.84-2.80)
Clashscore	141614	4060 (2.84-2.80)
Ramachandran outliers	138981	3978 (2.84-2.80)
Sidechain outliers	138945	3980 (2.84-2.80)
RSRZ outliers	127900	3552 (2.84-2.80)

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 >=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 <=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	737	4% 77% 23%
2	B	727	5% 75% 24%
3	C	731	5% 72% 26%
4	D	734	% 81% 18%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 23173 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 Beta-N-acetylhexosaminidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	737	5773	3694	995	1061	4	19	4	2	0

- Molecule 2 is a protein called Beta-N-acetylhexosaminidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
2	B	727	5701	3651	982	1045	4	19	1	3	0

- Molecule 3 is a protein called Beta-N-acetylhexosaminidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
3	C	731	5731	3669	986	1053	4	19	0	4	0

- Molecule 4 is a protein called Beta-N-acetylhexosaminidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
4	D	734	5762	3689	991	1059	4	19	0	5	0

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
5	A	1	1	1	0	0
5	B	1	Total	Mg	0	0
			1	1		
5	C	1	Total	Mg	0	0
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	D	1	Total Mg 1 1	0	0

- Molecule 6 is water.

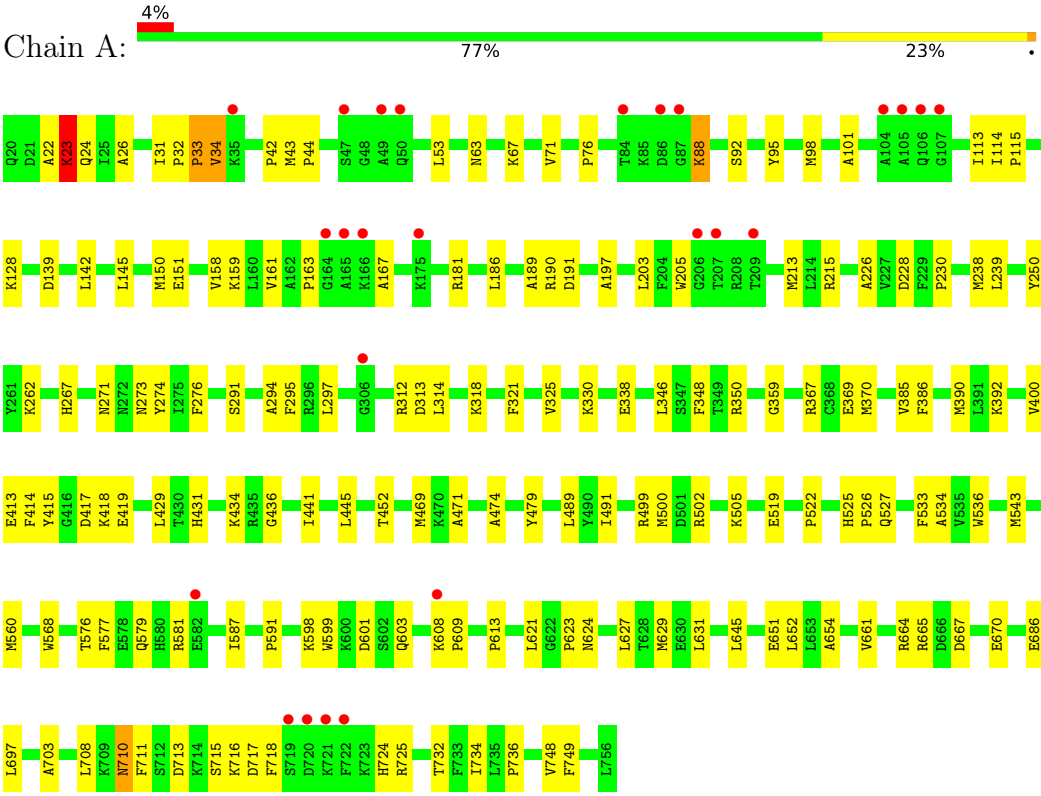
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	48	Total O 48 48	0	0
6	B	40	Total O 40 40	0	0
6	C	52	Total O 52 52	0	0
6	D	62	Total O 62 62	0	0

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.

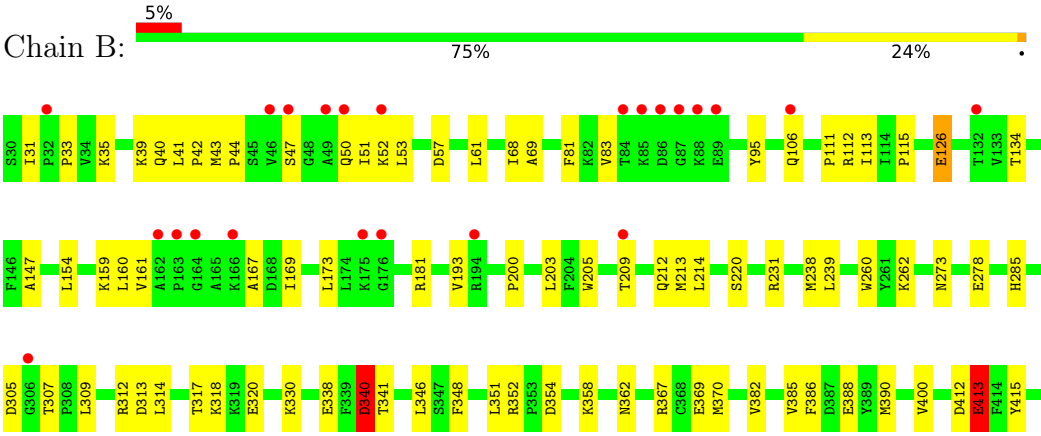
- Molecule 1: Beta-N-acetylhexosaminidase

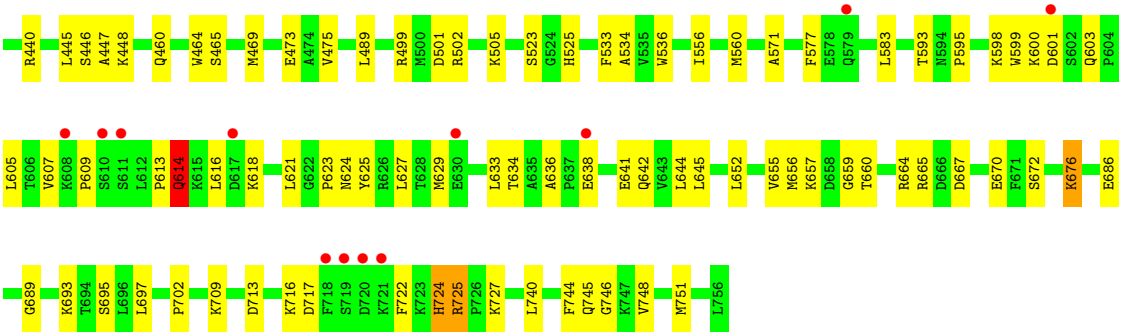
Chain A:



- Molecule 2: Beta-N-acetylhexosaminidase

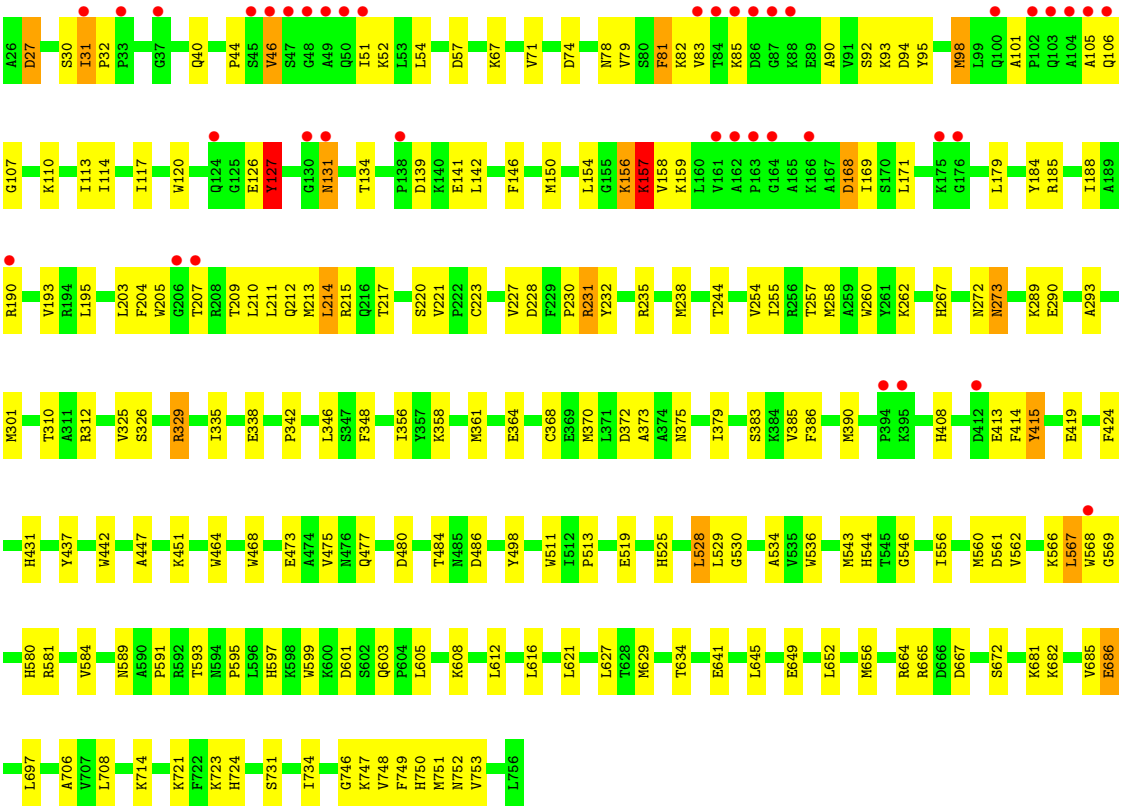
Chain B:





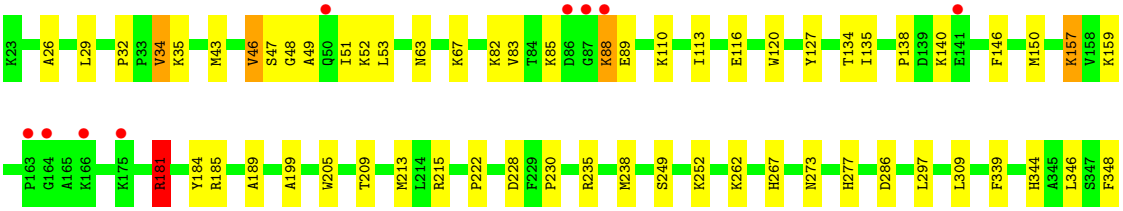
• Molecule 3: Beta-N-acetylhexosaminidase

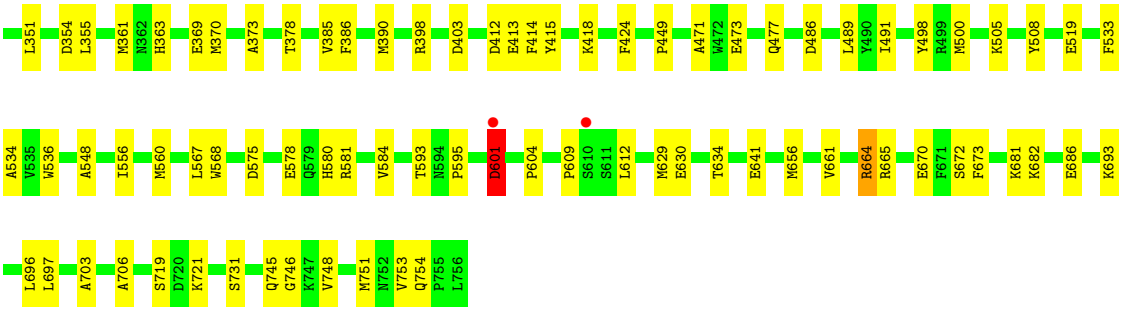
Chain C:



• Molecule 4: Beta-N-acetylhexosaminidase

Chain D:




L351
D354
L355
M361
N362
H363
E369
M370
A373
T378
V385
F386
M390
R398
D403
D412
E413
F414
Y415
K418
F424
P449
A471
W472
E473
Q477
D486
L489
Y490
I491
Y498
R499
M500
K505
Y508
E519
F533
A534
V535
W536
A548
I556
M560
L567
W568
D575
E578
Q579
H580
R581
V584
T593
N594
P595
D601
P604
P609
S610
S611
L612
M629
E630
T634
E641
M656
V661
R664
R665
E670
F671
S672
F673
K681
K682
E686
K693
L696
L697
A703
A706
S719
D720
K721
S731
Q745
G746
K747
V748
M751
N752
V753
Q754
P755
L756

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	96.20Å 119.50Å 161.93Å 90.00° 103.40° 90.00°	Depositor
Resolution (Å)	26.44 – 2.81 29.88 – 2.81	Depositor EDS
% Data completeness (in resolution range)	98.9 (26.44-2.81) 99.0 (29.88-2.81)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.67 (at 2.80Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.194 , 0.248 0.194 , 0.248	Depositor DCC
R_{free} test set	2001 reflections (2.32%)	wwPDB-VP
Wilson B-factor (Å^2)	43.8	Xtriage
Anisotropy	0.736	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 48.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	23173	wwPDB-VP
Average B, all atoms (Å^2)	50.0	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 19.88 % 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 9.8182e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: MG

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	2/5906 (0.0%)	0.65	5/7969 (0.1%)
2	B	0.47	3/5837 (0.1%)	0.90	10/7876 (0.1%)
3	C	0.44	0/5870	0.91	20/7921 (0.3%)
4	D	0.44	1/5904 (0.0%)	0.84	14/7966 (0.2%)
All	All	0.44	6/23517 (0.0%)	0.83	49/31732 (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	2
2	B	0	3
3	C	0	5
4	D	0	3
All	All	0	13

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	126	GLU	CD-OE2	8.45	1.34	1.25
2	B	413	GLU	CG-CD	7.43	1.63	1.51
1	A	88	LYS	CE-NZ	-6.04	1.33	1.49
2	B	413	GLU	CD-OE1	-5.97	1.19	1.25
4	D	157	LYS	CE-NZ	-5.67	1.34	1.49

The worst 5 of 49 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	340	ASP	CB-CG-OD1	36.94	151.54	118.30
2	B	340	ASP	CB-CG-OD2	-34.24	87.48	118.30
3	C	686	GLU	OE1-CD-OE2	-33.91	82.61	123.30
4	D	181	ARG	NE-CZ-NH2	-29.19	105.71	120.30
4	D	181	ARG	CD-NE-CZ	25.97	159.96	123.60

There are no chirality outliers.

5 of 13 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	23	LYS	Peptide
1	A	711	PHE	Sidechain
2	B	340	ASP	Sidechain
2	B	614	GLN	Sidechain
2	B	724	HIS	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	5773	0	5798	112	0
2	B	5701	0	5729	125	1
3	C	5731	0	5759	158	0
4	D	5762	0	5795	103	1
5	A	1	0	0	0	0
5	B	1	0	0	0	0
5	C	1	0	0	0	0
5	D	1	0	0	0	0
6	A	48	0	0	3	0
6	B	40	0	0	0	0
6	C	52	0	0	6	0
6	D	62	0	0	4	0
All	All	23173	0	23081	491	1

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 11.

The worst 5 of 491 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:386:PHE:HB3	1:A:390:MSE:HE2	1.37	1.06
3:C:46:VAL:HG12	3:C:51:ILE:HD11	1.46	0.96
2:B:386:PHE:HB3	2:B:390:MSE:HE2	1.50	0.94
3:C:214:LEU:HD12	6:C:901:HOH:O	1.70	0.90
2:B:609:PRO:HG3	2:B:748:VAL:HG23	1.53	0.89

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:362:ASN:OD1	4:D:612:LEU:N[2_555]	2.16	0.04

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	737/737 (100%)	711 (96%)	20 (3%)	6 (1%)	19	47
2	B	728/727 (100%)	687 (94%)	35 (5%)	6 (1%)	19	47
3	C	733/731 (100%)	693 (94%)	35 (5%)	5 (1%)	22	51
4	D	737/734 (100%)	706 (96%)	29 (4%)	2 (0%)	41	70
All	All	2935/2929 (100%)	2797 (95%)	119 (4%)	19 (1%)	22	54

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	23	LYS
2	B	31	ILE
3	C	27	ASP
3	C	46	VAL
3	C	168	ASP

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	619/598 (104%)	618 (100%)	1 (0%)	93	98
2	B	612/590 (104%)	611 (100%)	1 (0%)	93	98
3	C	616/593 (104%)	615 (100%)	1 (0%)	93	98
4	D	620/596 (104%)	619 (100%)	1 (0%)	93	98
All	All	2467/2377 (104%)	2463 (100%)	4 (0%)	93	98

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	312	ARG
2	B	413	GLU
3	C	157	LYS
4	D	181	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	273	ASN
4	D	277	HIS
4	D	754	GLN
4	D	363	HIS
2	B	273	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 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

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	718/737 (97%)	-0.08	26 (3%)	42	32	30, 46, 78, 141	0
2	B	708/727 (97%)	0.01	35 (4%)	29	20	28, 50, 97, 157	0
3	C	712/731 (97%)	0.01	40 (5%)	24	16	26, 50, 97, 145	0
4	D	715/734 (97%)	-0.35	11 (1%)	73	67	21, 37, 69, 123	0
All	All	2853/2929 (97%)	-0.10	112 (3%)	39	29	21, 46, 89, 157	0

The worst 5 of 112 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	720	ASP	8.7
2	B	50	GLN	6.6
1	A	104	ALA	6.2
3	C	164	GLY	5.6
1	A	720	ASP	5.6

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 monosaccharides in this entry.

6.4 Ligands

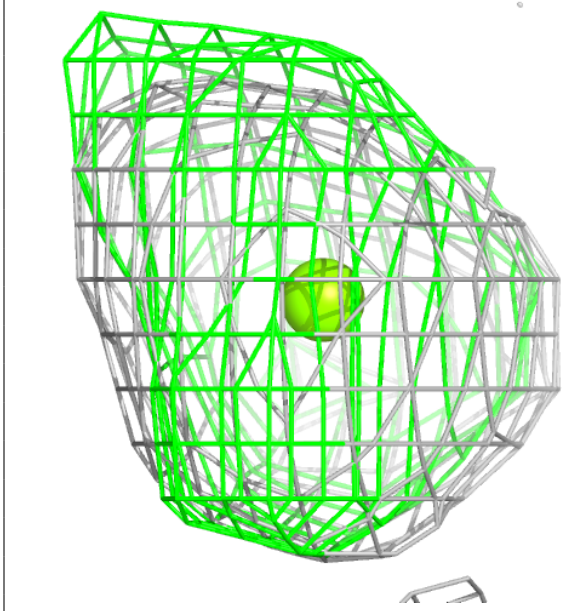
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

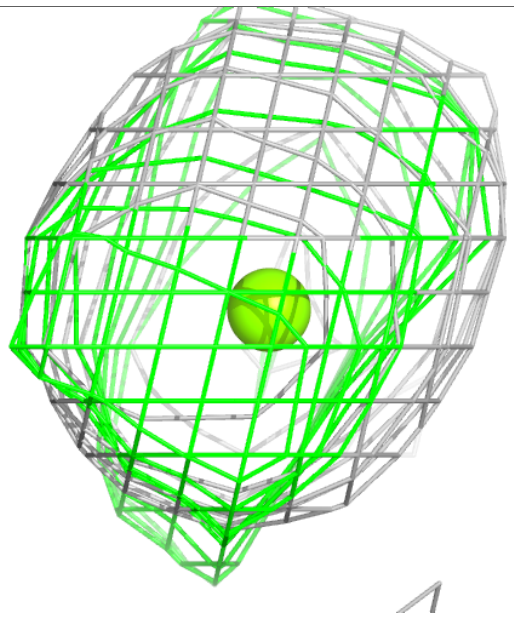
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	MG	C	801	1/1	0.97	0.27	29,29,29,29	0
5	MG	B	801	1/1	0.98	0.26	23,23,23,23	0
5	MG	A	801	1/1	0.99	0.26	22,22,22,22	0
5	MG	D	801	1/1	0.99	0.30	24,24,24,24	0

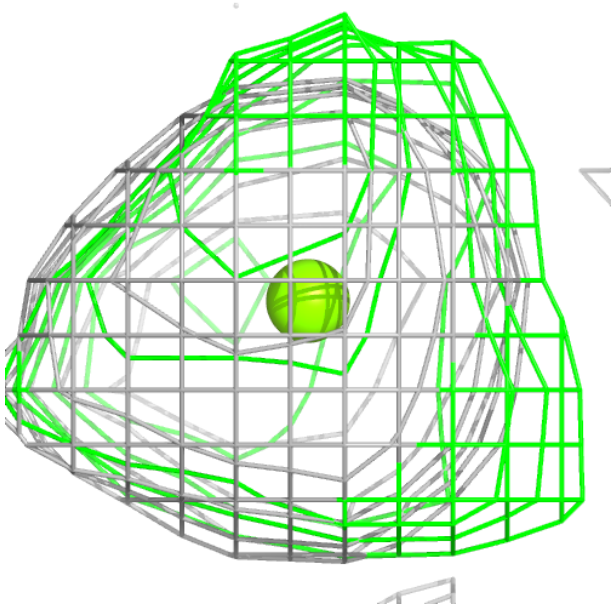
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around MG C 801:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

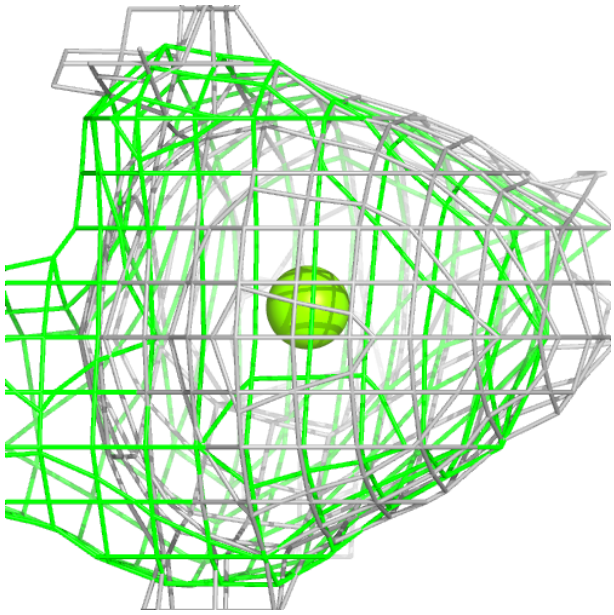


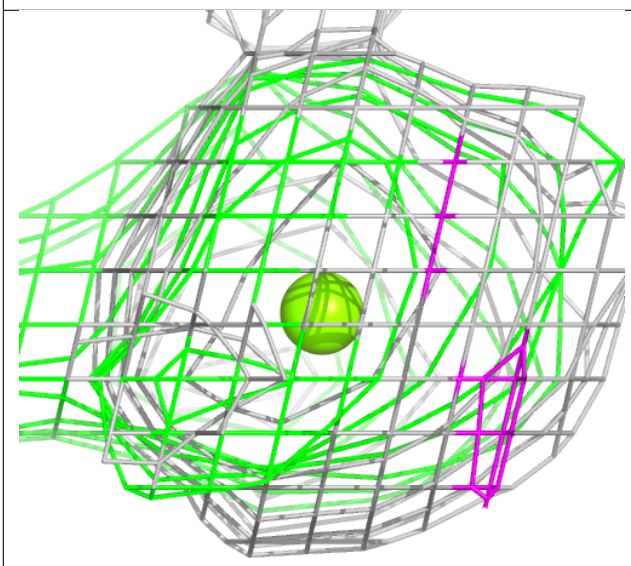


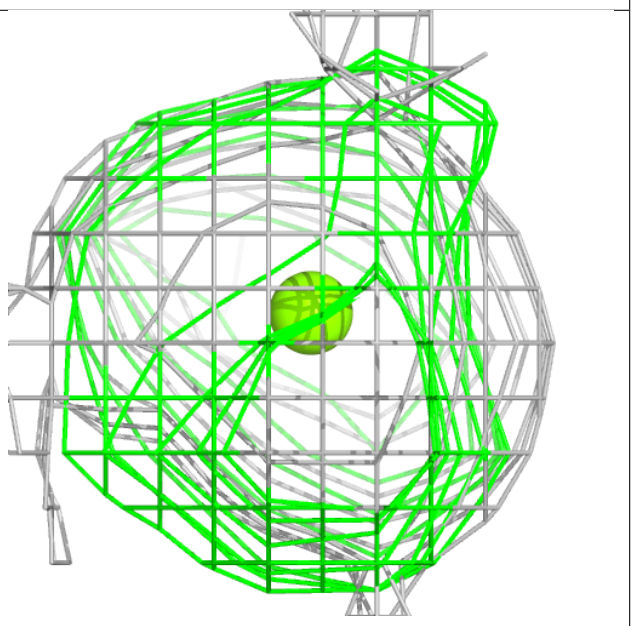


Electron density around MG B 801:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

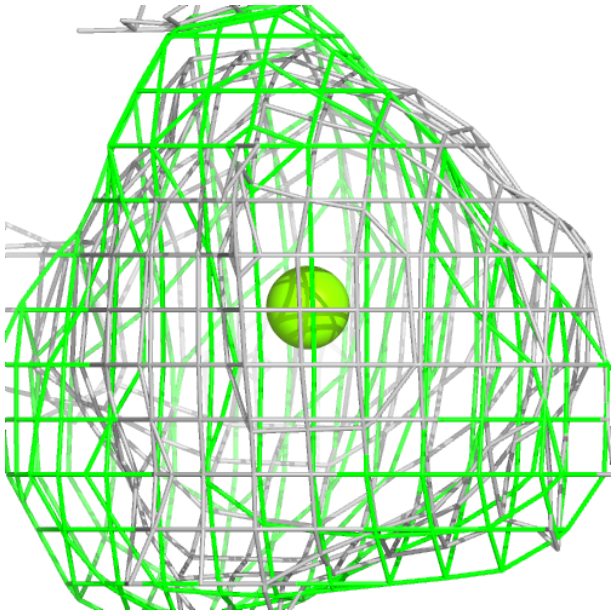


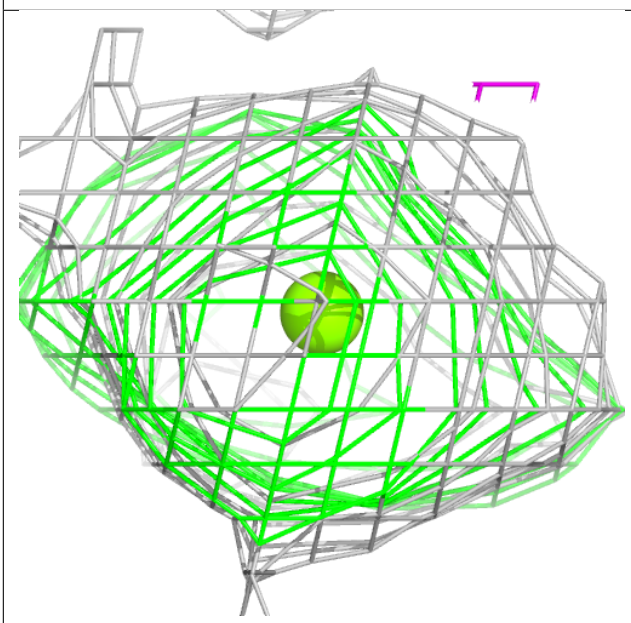


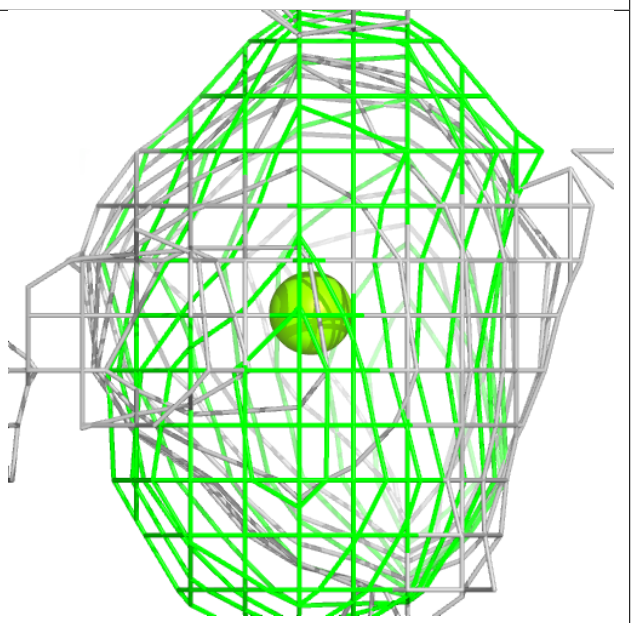


Electron density around MG A 801:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

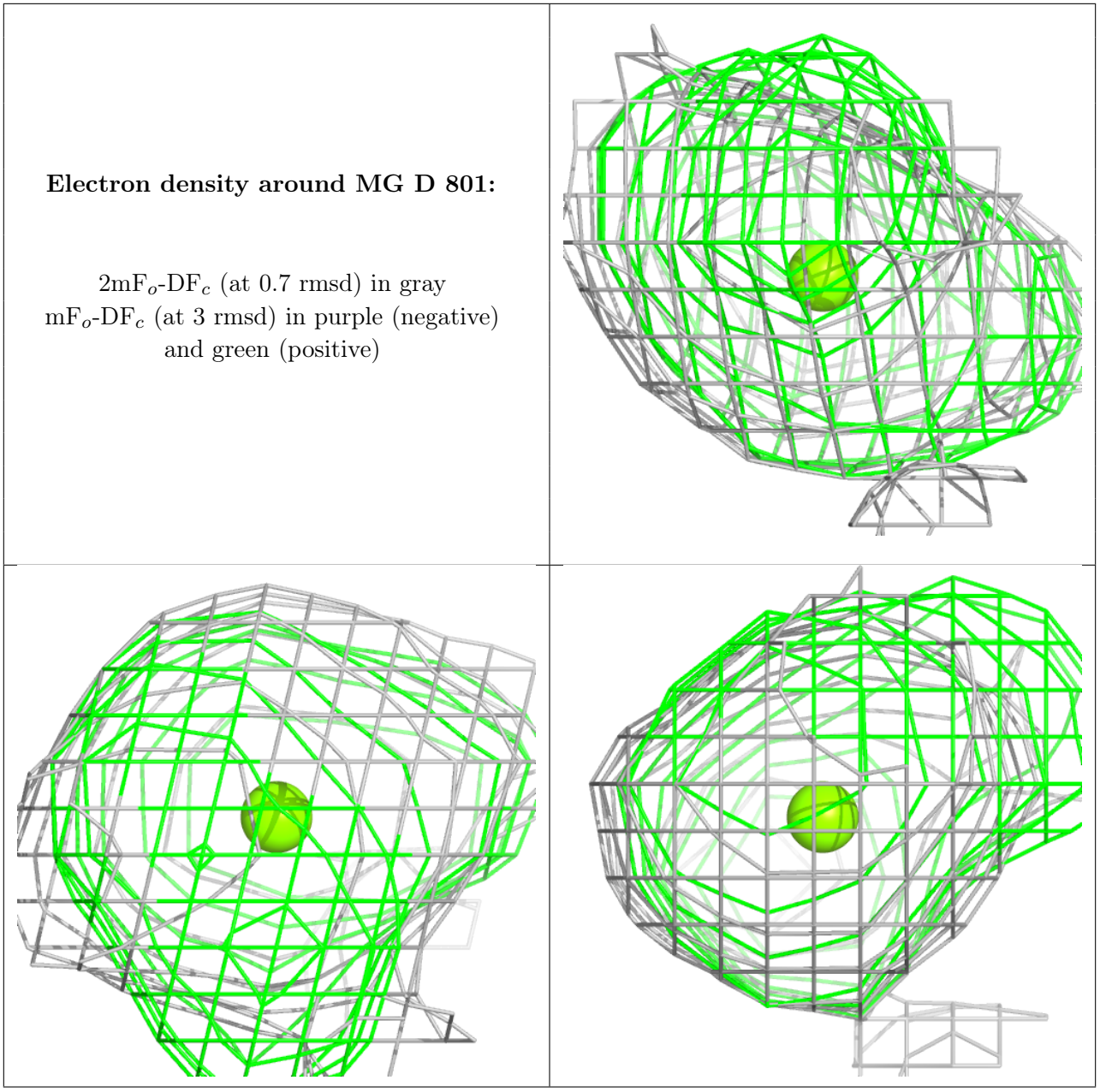






Electron density around MG D 801:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.