wwPDB X-ray Structure Validation Summary Report (i)

Oct 10, 2021 – 09:09 PM EDT

PDB ID	:	3EIH
Title	:	Crystal structure of S.cerevisiae Vps4 in the presence of ATPgammaS
Authors	:	Gonciarz, M.D.; Whitby, F.G.; Eckert, D.M.; Kieffer, C.; Heroux, A.; Sundquist, W.I.; Hill, C.P.
Deposited on	:	2008-09-15
Resolution	:	3.25 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 following versions of software and data (see references (i)) were used in the production of this report:

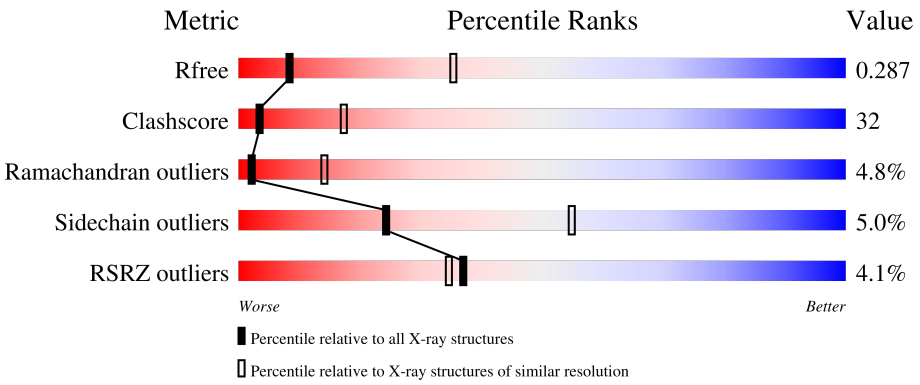
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 3.25 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	1191 (3.30-3.22)
Clashscore	141614	1251 (3.30-3.22)
Ramachandran outliers	138981	1229 (3.30-3.22)
Sidechain outliers	138945	1228 (3.30-3.22)
RSRZ outliers	127900	1154 (3.30-3.22)

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	340	2%, 53%, 36%, • • 6%
1	B	340	2%, 50%, 39%, • • 7%
1	C	340	8%, 45%, 39%, 7% • 9%

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
4	EDO	C	501	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7396 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 Vacuolar protein sorting-associated protein 4.

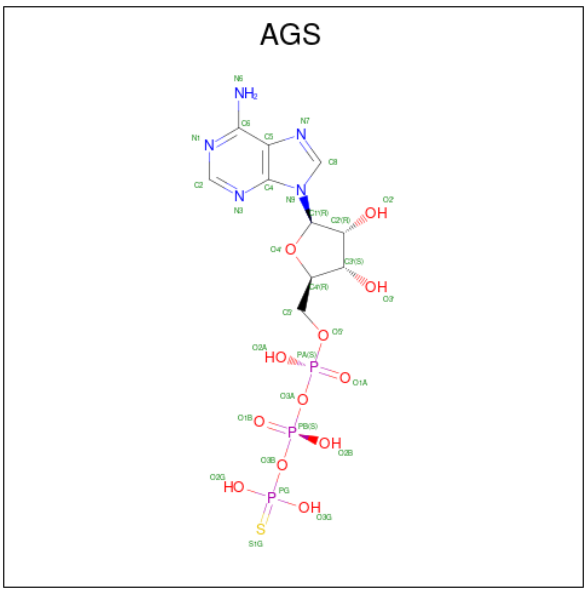
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	319	2469	1552	419	489	9	0	0	0
1	B	316	2449	1540	415	485	9	0	0	0
1	C	310	2409	1518	407	475	9	0	0	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	98	GLY	-	expression tag	UNP P52917
A	99	ILE	-	expression tag	UNP P52917
A	100	ASP	-	expression tag	UNP P52917
A	101	PRO	-	expression tag	UNP P52917
A	102	PHE	-	expression tag	UNP P52917
A	103	THR	-	expression tag	UNP P52917
A	233	GLN	GLU	engineered mutation	UNP P52917
B	98	GLY	-	expression tag	UNP P52917
B	99	ILE	-	expression tag	UNP P52917
B	100	ASP	-	expression tag	UNP P52917
B	101	PRO	-	expression tag	UNP P52917
B	102	PHE	-	expression tag	UNP P52917
B	103	THR	-	expression tag	UNP P52917
B	233	GLN	GLU	engineered mutation	UNP P52917
C	98	GLY	-	expression tag	UNP P52917
C	99	ILE	-	expression tag	UNP P52917
C	100	ASP	-	expression tag	UNP P52917
C	101	PRO	-	expression tag	UNP P52917
C	102	PHE	-	expression tag	UNP P52917
C	103	THR	-	expression tag	UNP P52917
C	233	GLN	GLU	engineered mutation	UNP P52917

- Molecule 2 is PHOSPHOTHIOPHOSPHORIC ACID-ADENYLATE ESTER (three-letter

code: AGS) (formula: $C_{10}H_{16}N_5O_{12}P_3S$).

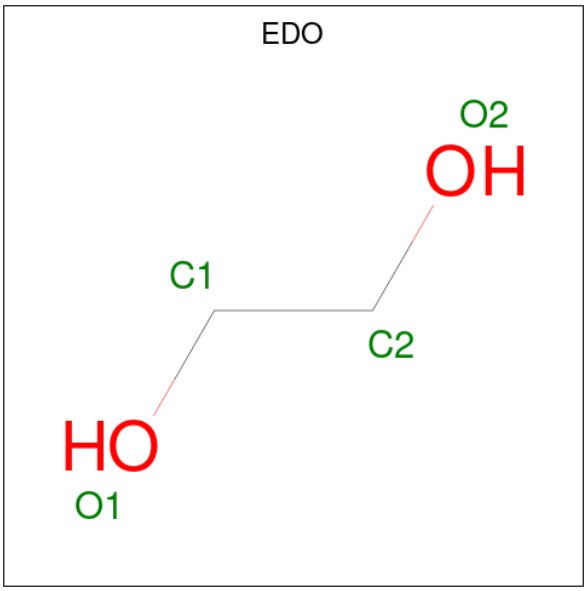


Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	N	O	P	S		
2	A	1	31	10	5	12	3	1	0	0
2	B	1	31	10	5	12	3	1	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



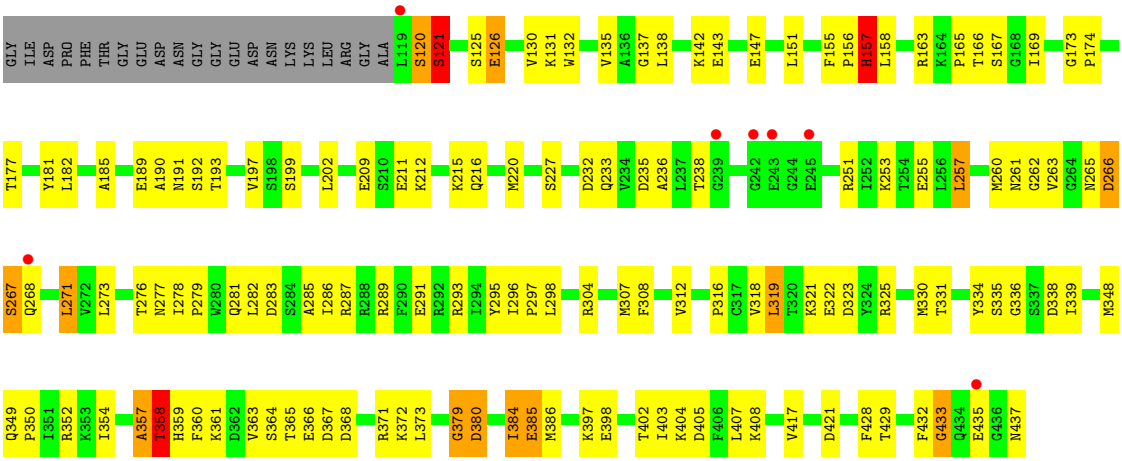
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	C	1	4	2	2	0	0

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.

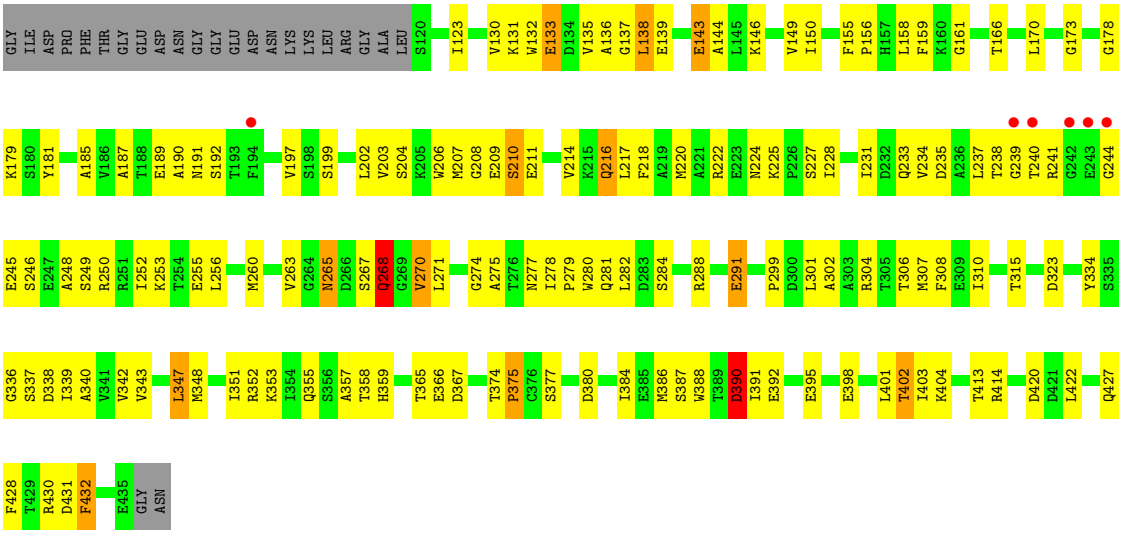
• Molecule 1: Vacuolar protein sorting-associated protein 4

Chain A:



• Molecule 1: Vacuolar protein sorting-associated protein 4

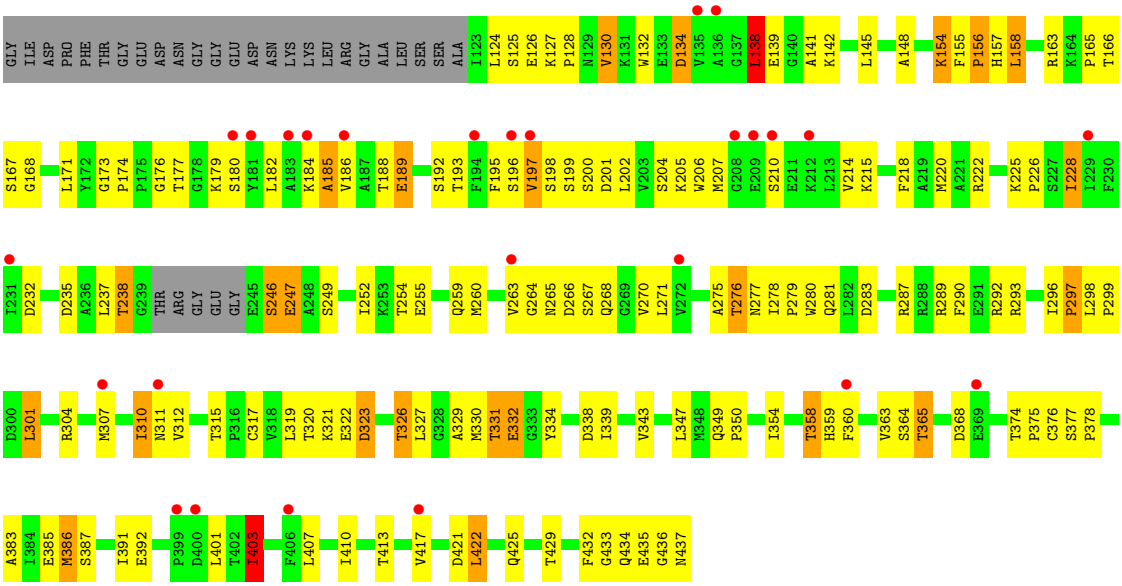
Chain B:



• Molecule 1: Vacuolar protein sorting-associated protein 4

Chain C:

8%
45%
39%
7%
9%


GLY ILE ASP PRO PHE THR GLY GLU ASP ASN GLY GLY GLU ASP ASN LYS LYS LEU ARG GLY ALA LEU SER SER ALA I123 L124 S125 E126 K127 P128 N129 V130 K131 W132 E133 D134 V135 A136 G137 L138 E139 G140 A141 K142 L145 A148 K154 F155 P156 H157 L158 R163 K164 P165 T166
S167 G168 L171 Y172 G173 P174 P175 G176 T177 G178 K179 S180 Y181 L182 A183 K184 A185 V186 A187 T188 E189 S192 T193 F194 F195 S196 V197 S198 S199 S200 D201 L202 V203 S204 K205 W206 M207 G208 E209 S210 E211 K212 L213 V214 K215 F218 A219 M220 A221 R222 K225 P226 S227 I228 I229 F230
I231 D232 D235 A236 L237 T238 G239 THR ARG GLY GLU GLY E245 S246 E247 A248 S249 I252 K253 T254 E255 Q259 M260 V263 G264 N265 D266 S267 Q268 G269 V270 L271 V272 A275 T276 N277 I278 P279 W280 Q281 L282 D283 R287 R288 R289 F290 E291 R292 R293 I296 P297 L298 P299
D300 L301 R304 M307 I310 N311 V312 T315 P316 C317 V318 L319 T320 K321 E322 D323 T326 L327 G328 A329 M330 T331 E332 G333 Y334 D338 I339 V343 L347 M348 Q349 P350 I354 T358 H359 F360 V363 S364 T365 D368 E369 T374 P375 C376 S377 P378
A383 I384 E385 M386 S387 I391 E392 P399 D400 L401 T402 I403 F406 L407 I410 T413 V417 D421 L422 Q425 T429 F432 G433 Q434 E435 G436 N437

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	76.96Å 119.80Å 156.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.00 – 3.25 49.93 – 3.25	Depositor EDS
% Data completeness (in resolution range)	99.7 (33.00-3.25) 99.6 (49.93-3.25)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.77 (at 3.25Å)	Xtriage
Refinement program	REFMAC 5.4.0054	Depositor
R, R_{free}	0.226 , 0.305 0.242 , 0.287	Depositor DCC
R_{free} test set	1208 reflections (5.14%)	wwPDB-VP
Wilson B-factor (Å^2)	79.8	Xtriage
Anisotropy	0.164	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 78.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	7396	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.99% of the height of the origin peak. No significant pseudotranslation is detected.*

[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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: AGS, MG, EDO

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	0.64	0/2511	0.84	6/3395 (0.2%)
1	B	0.60	0/2491	0.81	5/3368 (0.1%)
1	C	0.55	0/2450	0.70	0/3312
All	All	0.60	0/7452	0.79	11/10075 (0.1%)

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
1	B	0	1
All	All	0	3

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	156	PRO	N-CA-C	13.60	147.47	112.10
1	A	157	HIS	N-CA-C	-9.61	85.06	111.00
1	B	375	PRO	CA-N-CD	-9.44	98.29	111.50
1	A	319	LEU	CA-CB-CG	6.63	130.54	115.30
1	B	216	GLN	N-CA-C	6.25	127.87	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	157	HIS	Peptide
1	A	357	ALA	Peptide
1	B	216	GLN	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	2469	0	2476	122	0
1	B	2449	0	2457	145	0
1	C	2409	0	2417	216	0
2	A	31	0	12	1	0
2	B	31	0	12	4	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
4	C	4	0	6	10	0
All	All	7396	0	7380	476	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 32.

The worst 5 of 476 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:C:154:LYS:HD3	1:C:155:PHE:CE1	1.61	1.34
1:B:386:MET:CE	1:B:391:ILE:HG13	1.63	1.27
1:B:374:THR:HG23	1:B:375:PRO:CD	1.68	1.24
1:B:386:MET:HE1	1:B:391:ILE:CG1	1.67	1.24
1:C:154:LYS:HE3	1:C:155:PHE:CZ	1.75	1.21

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	317/340 (93%)	267 (84%)	40 (13%)	10 (3%)	4 23
1	B	314/340 (92%)	252 (80%)	47 (15%)	15 (5%)	2 14
1	C	306/340 (90%)	227 (74%)	59 (19%)	20 (6%)	1 9
All	All	937/1020 (92%)	746 (80%)	146 (16%)	45 (5%)	2 14

5 of 45 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	266	ASP
1	B	191	ASN
1	B	210	SER
1	B	291	GLU
1	B	366	GLU

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	272/287 (95%)	262 (96%)	10 (4%)	34 62
1	B	270/287 (94%)	260 (96%)	10 (4%)	34 62
1	C	266/287 (93%)	246 (92%)	20 (8%)	13 39
All	All	808/861 (94%)	768 (95%)	40 (5%)	24 55

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	270	VAL
1	C	364	SER
1	C	276	THR
1	C	331	THR
1	C	377	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	359	HIS
1	C	425	GLN
1	B	268	GLN
1	C	261	ASN
1	C	265	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 6 ligands modelled in this entry, 3 are monoatomic - leaving 3 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	$\#\|Z\| > 2$
2	AGS	A	501	-	26,33,33	1.79	4 (15%)	26,52,52	1.29	5 (19%)
4	EDO	C	501	-	3,3,3	0.49	0	2,2,2	0.24	0
2	AGS	B	501	-	26,33,33	1.94	4 (15%)	26,52,52	1.29	2 (7%)

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
2	AGS	A	501	-	-	4/17/38/38	0/3/3/3
4	EDO	C	501	-	-	1/1/1/1	-
2	AGS	B	501	-	-	8/17/38/38	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	501	AGS	PG-S1G	8.18	2.08	1.90
2	A	501	AGS	PG-S1G	7.47	2.06	1.90
2	B	501	AGS	C5-C4	2.73	1.48	1.40
2	A	501	AGS	C5-C4	2.60	1.47	1.40
2	B	501	AGS	C2-N3	2.09	1.35	1.32

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	501	AGS	N3-C2-N1	-3.79	122.75	128.68
2	A	501	AGS	N3-C2-N1	-3.04	123.93	128.68
2	B	501	AGS	C4-C5-N7	-2.51	106.78	109.40
2	A	501	AGS	O3G-PG-O3B	2.29	112.29	104.64
2	A	501	AGS	PA-O3A-PB	-2.09	125.66	132.83

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	501	AGS	C5'-O5'-PA-O3A
2	B	501	AGS	C5'-O5'-PA-O3A
2	B	501	AGS	PG-O3B-PB-O1B
2	B	501	AGS	PA-O3A-PB-O3B
2	B	501	AGS	PB-O3A-PA-O5'

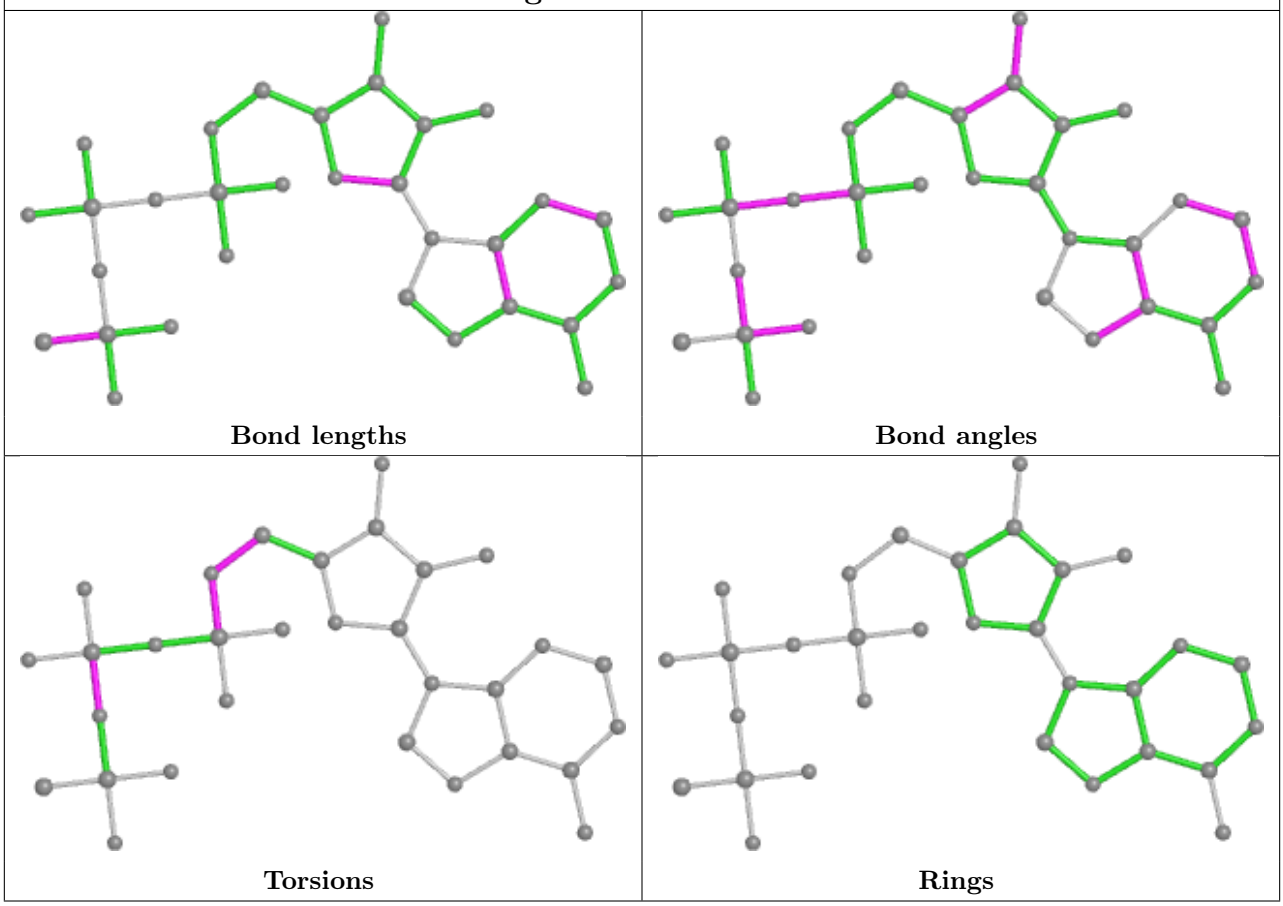
There are no ring outliers.

3 monomers are involved in 15 short contacts:

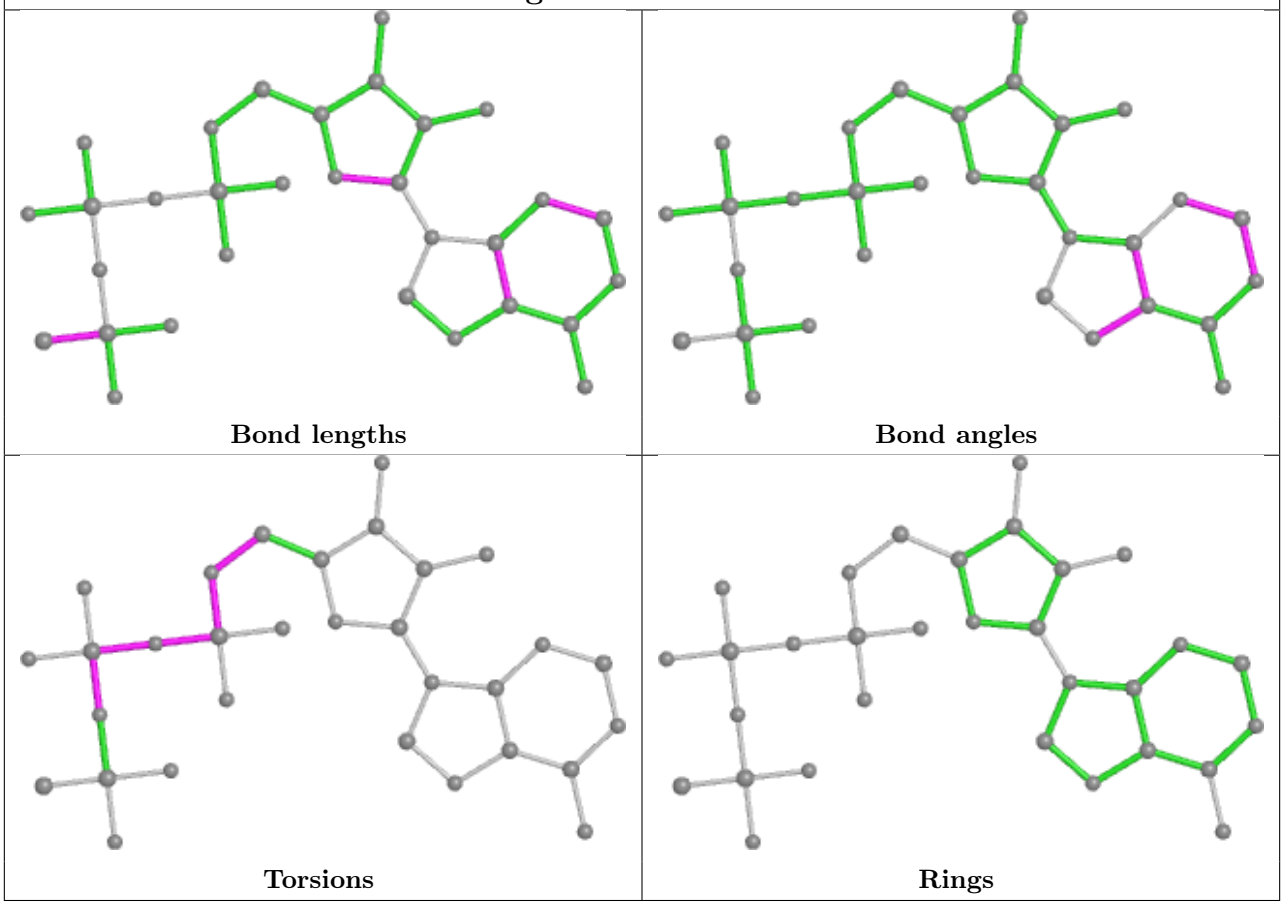
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	AGS	1	0
4	C	501	EDO	10	0
2	B	501	AGS	4	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand AGS A 501



Ligand AGS B 501



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	319/340 (93%)	0.23	7 (2%)	62	59	42, 68, 108, 124	0
1	B	316/340 (92%)	0.28	6 (1%)	66	64	51, 82, 122, 138	0
1	C	310/340 (91%)	0.58	26 (8%)	11	11	74, 105, 130, 136	0
All	All	945/1020 (92%)	0.36	39 (4%)	37	34	42, 84, 126, 138	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	243	GLU	5.6
1	B	240	THR	5.2
1	C	181	TYR	4.8
1	A	242	GLY	4.7
1	B	242	GLY	4.3

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)

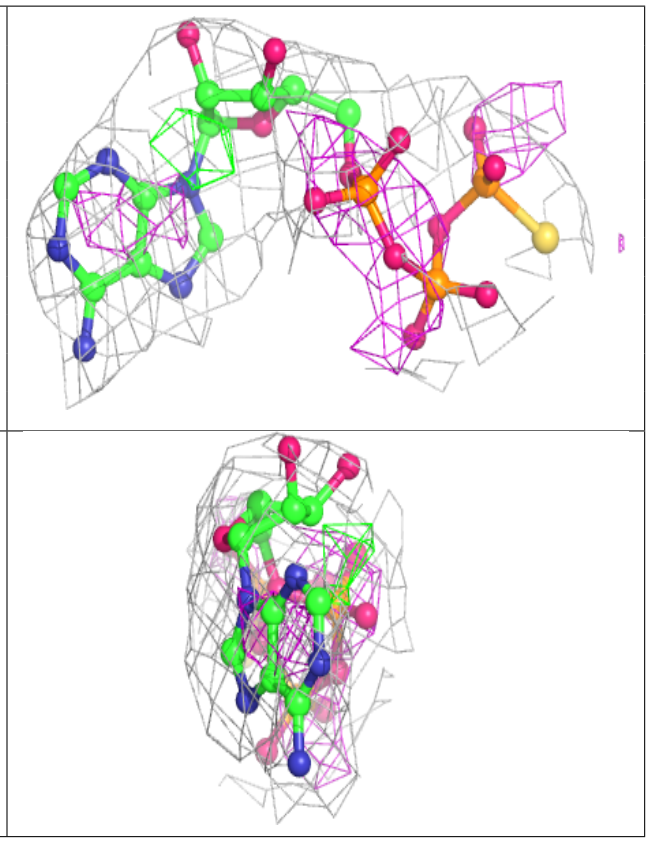
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	MG	C	502	1/1	0.84	0.12	100,100,100,100	0
2	AGS	B	501	31/31	0.91	0.16	103,107,113,114	0
3	MG	A	502	1/1	0.94	0.18	51,51,51,51	0
2	AGS	A	501	31/31	0.94	0.16	64,67,81,82	0
4	EDO	C	501	4/4	0.95	0.08	54,56,58,61	0
3	MG	B	502	1/1	0.96	0.14	59,59,59,59	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.

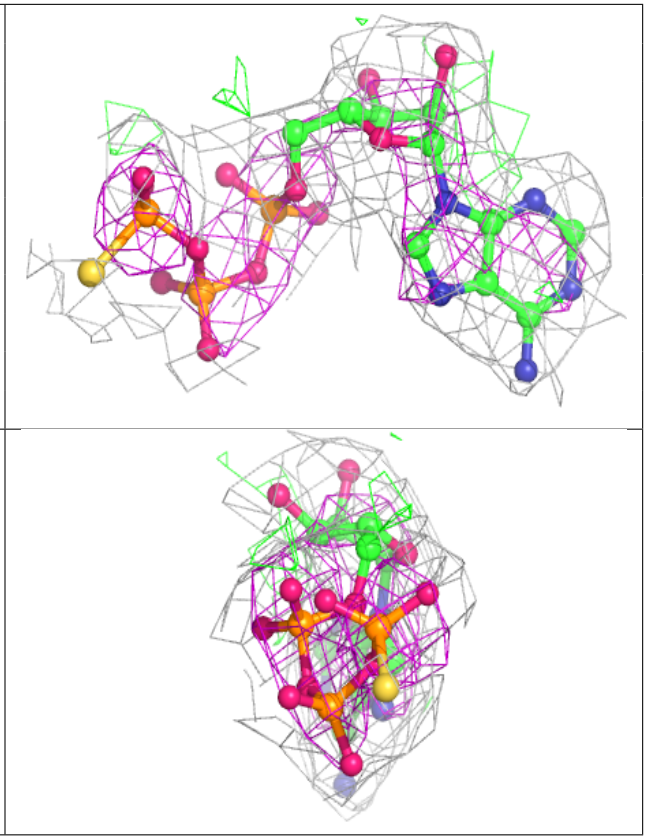
Electron density around AGS B 501:

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 AGS A 501:

$2mF_o$-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.