wwPDB X-ray Structure Validation Summary Report (i)

May 17, 2020 – 08:01 pm BST

PDB ID	:	5HQ3
Title	:	Stable, high-expression variant of human acetylcholinesterase
Authors	:	Goldenzweig, A.; Goldsmith, M.; Hill, S.E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Albeck, S.; Unger, T.; Prilusky, J.; Lieberman, R.L.; Aharoni, A.; Silman, I.; Sussman, J.L.; Tawfik, D.S.; Fleishman, S.J.
Deposited on	:	2016-01-21
Resolution	:	2.60 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 (1)) were used in the production of this report:

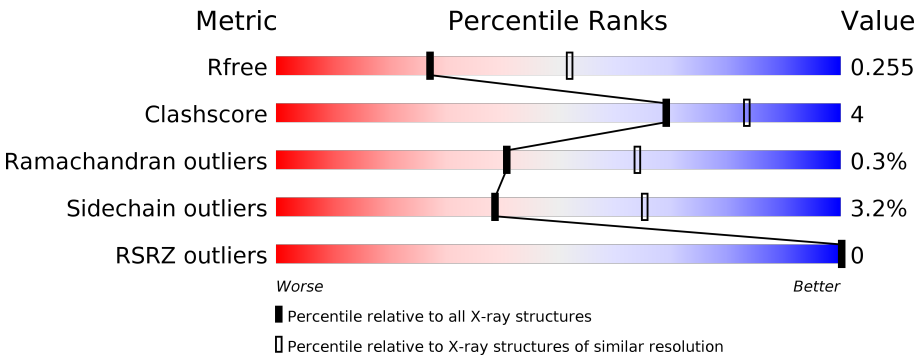
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 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	548	87% 9% •
2	B	549	83% 12% • •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 8356 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 Acetylcholinesterase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	529	4173	2680	729	751	13	0	1	0

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP P22303
A	12	THR	VAL	engineered mutation	UNP P22303
A	23	THR	LYS	engineered mutation	UNP P22303
A	42	VAL	MET	engineered mutation	UNP P22303
A	48	ARG	LEU	engineered mutation	UNP P22303
A	60	TRP	VAL	engineered mutation	UNP P22303
A	67	ASN	SER	engineered mutation	UNP P22303
A	91	ASN	GLU	engineered mutation	UNP P22303
A	109	LYS	THR	engineered mutation	UNP P22303
A	110	ASN	SER	engineered mutation	UNP P22303
A	112	ALA	THR	engineered mutation	UNP P22303
A	115	MET	LEU	engineered mutation	UNP P22303
A	127	SER	ALA	engineered mutation	UNP P22303
A	140	ARG	GLN	engineered mutation	UNP P22303
A	141	THR	ALA	engineered mutation	UNP P22303
A	144	VAL	THR	engineered mutation	UNP P22303
A	187	ILE	VAL	engineered mutation	UNP P22303
A	226	ILE	VAL	engineered mutation	UNP P22303
A	234	ALA	GLY	engineered mutation	UNP P22303
A	238	TYR	THR	engineered mutation	UNP P22303
A	240	SER	GLY	engineered mutation	UNP P22303
A	241	ARG	MET	engineered mutation	UNP P22303
A	242	GLU	GLY	engineered mutation	UNP P22303
A	249	LEU	THR	engineered mutation	UNP P22303
A	253	LYS	HIS	engineered mutation	UNP P22303
A	275	ASN	THR	engineered mutation	UNP P22303
A	278	PRO	ALA	engineered mutation	UNP P22303

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	GLU	VAL	engineered mutation	UNP P22303
A	309	PRO	SER	engineered mutation	UNP P22303
A	318	ASN	ALA	engineered mutation	UNP P22303
A	322	LYS	HIS	engineered mutation	UNP P22303
A	325	ASP	GLN	engineered mutation	UNP P22303
A	331	ASN	VAL	engineered mutation	UNP P22303
A	357	GLU	ALA	engineered mutation	UNP P22303
A	361	GLU	ALA	engineered mutation	UNP P22303
A	378	ILE	VAL	engineered mutation	UNP P22303
A	393	LYS	ARG	engineered mutation	UNP P22303
A	394	ASN	LEU	engineered mutation	UNP P22303
A	396	ASP	GLU	engineered mutation	UNP P22303
A	408	ILE	VAL	engineered mutation	UNP P22303
A	414	PHE	LEU	engineered mutation	UNP P22303
A	416	GLN	GLY	engineered mutation	UNP P22303
A	418	TYR	LEU	engineered mutation	UNP P22303
A	421	ASN	GLN	engineered mutation	UNP P22303
A	434	SER	ALA	engineered mutation	UNP P22303
A	438	PRO	SER	engineered mutation	UNP P22303
A	441	GLU	LEU	engineered mutation	UNP P22303
A	467	LYS	ALA	engineered mutation	UNP P22303
A	474	ARG	GLN	engineered mutation	UNP P22303
A	506	ASP	GLY	engineered mutation	UNP P22303
A	507	GLU	ALA	engineered mutation	UNP P22303
A	509	LYS	GLN	engineered mutation	UNP P22303

- Molecule 2 is a protein called Acetylcholinesterase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	528	4146	2666	721	746	13	0	1	0

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP P22303
B	12	THR	VAL	engineered mutation	UNP P22303
B	23	THR	LYS	engineered mutation	UNP P22303
B	42	VAL	MET	engineered mutation	UNP P22303
B	48	ARG	LEU	engineered mutation	UNP P22303
B	60	TRP	VAL	engineered mutation	UNP P22303
B	67	ASN	SER	engineered mutation	UNP P22303

Continued on next page...

Continued from previous page...

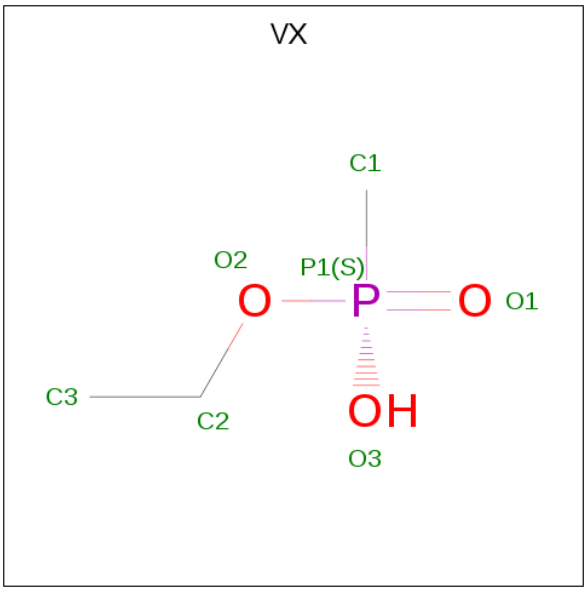
Chain	Residue	Modelled	Actual	Comment	Reference
B	91	ASN	GLU	engineered mutation	UNP P22303
B	109	LYS	THR	engineered mutation	UNP P22303
B	110	ASN	SER	engineered mutation	UNP P22303
B	112	ALA	THR	engineered mutation	UNP P22303
B	115	MET	LEU	engineered mutation	UNP P22303
B	127	SER	ALA	engineered mutation	UNP P22303
B	140	ARG	GLN	engineered mutation	UNP P22303
B	141	THR	ALA	engineered mutation	UNP P22303
B	144	VAL	THR	engineered mutation	UNP P22303
B	187	ILE	VAL	engineered mutation	UNP P22303
B	226	ILE	VAL	engineered mutation	UNP P22303
B	234	ALA	GLY	engineered mutation	UNP P22303
B	238	TYR	THR	engineered mutation	UNP P22303
B	240	SER	GLY	engineered mutation	UNP P22303
B	241	ARG	MET	engineered mutation	UNP P22303
B	242	GLU	GLY	engineered mutation	UNP P22303
B	249	LEU	THR	engineered mutation	UNP P22303
B	253	LYS	HIS	engineered mutation	UNP P22303
B	275	ASN	THR	engineered mutation	UNP P22303
B	278	PRO	ALA	engineered mutation	UNP P22303
B	280	GLU	VAL	engineered mutation	UNP P22303
B	309	PRO	SER	engineered mutation	UNP P22303
B	318	ASN	ALA	engineered mutation	UNP P22303
B	322	LYS	HIS	engineered mutation	UNP P22303
B	325	ASP	GLN	engineered mutation	UNP P22303
B	331	ASN	VAL	engineered mutation	UNP P22303
B	357	GLU	ALA	engineered mutation	UNP P22303
B	361	GLU	ALA	engineered mutation	UNP P22303
B	378	ILE	VAL	engineered mutation	UNP P22303
B	393	LYS	ARG	engineered mutation	UNP P22303
B	394	ASN	LEU	engineered mutation	UNP P22303
B	396	ASP	GLU	engineered mutation	UNP P22303
B	408	ILE	VAL	engineered mutation	UNP P22303
B	414	PHE	LEU	engineered mutation	UNP P22303
B	416	GLN	GLY	engineered mutation	UNP P22303
B	418	TYR	LEU	engineered mutation	UNP P22303
B	421	ASN	GLN	engineered mutation	UNP P22303
B	434	SER	ALA	engineered mutation	UNP P22303
B	438	PRO	SER	engineered mutation	UNP P22303
B	441	GLU	LEU	engineered mutation	UNP P22303
B	467	LYS	ALA	engineered mutation	UNP P22303
B	474	ARG	GLN	engineered mutation	UNP P22303

Continued on next page...

Continued from previous page...

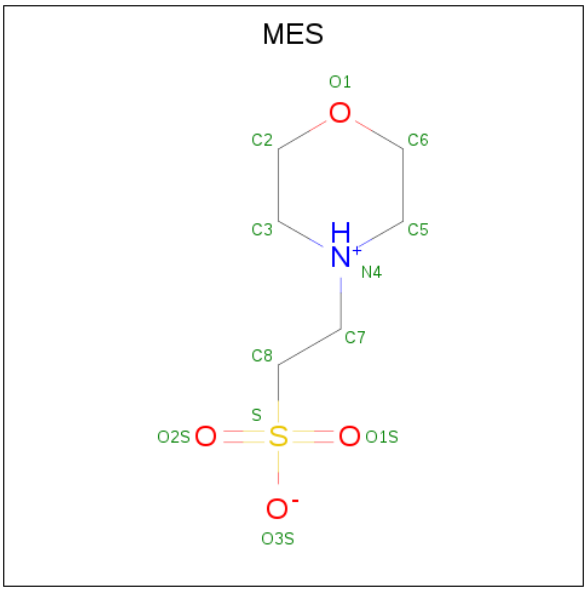
Chain	Residue	Modelled	Actual	Comment	Reference
B	506	ASP	GLY	engineered mutation	UNP P22303
B	507	GLU	ALA	engineered mutation	UNP P22303
B	509	LYS	GLN	engineered mutation	UNP P22303

- Molecule 3 is O-ETHYLMETHYLPHOSPHONIC ACID ESTER GROUP (three-letter code: VX) (formula: $C_3H_9O_3P$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	O	P		
3	A	1	6	3	2	1	0	0
			Total	C	O	P		
3	B	1	6	3	2	1	0	0

- Molecule 4 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
4	A	1	12	6	1	4	1	0	0
			Total	C	N	O	S		
4	B	1	12	6	1	4	1	0	0

- Molecule 5 is water.

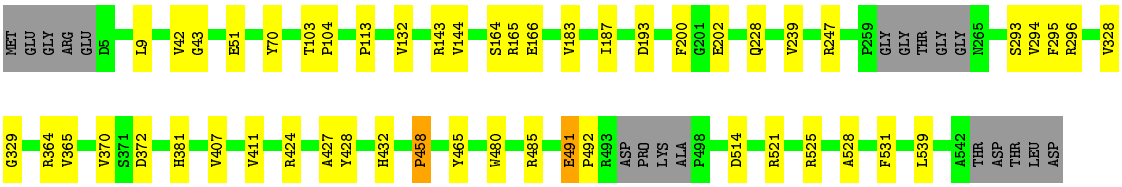
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	1	1	1	0	0

3 Residue-property plots (i)

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 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.

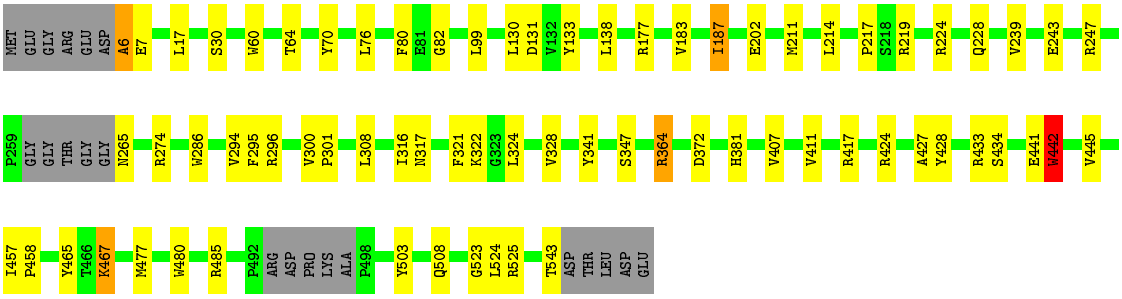
- Molecule 1: Acetylcholinesterase

Chain A:



- Molecule 2: Acetylcholinesterase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	89.53Å 89.53Å 395.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.60 49.42 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.8 (50.00-2.60) 99.8 (49.42-2.60)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	4.22 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.201 , 0.253 0.205 , 0.255	Depositor DCC
R_{free} test set	2498 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	43.9	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 21.9	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	8356	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.93% 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: VX, MES

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	0.75	0/4308	0.88	8/5886 (0.1%)
2	B	0.74	0/4280	0.87	7/5849 (0.1%)
All	All	0.75	0/8588	0.88	15/11735 (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
2	B	0	2

There are no bond length outliers.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	274	ARG	NE-CZ-NH2	-8.75	115.92	120.30
1	A	296	ARG	NE-CZ-NH1	7.12	123.86	120.30
1	A	424	ARG	NE-CZ-NH1	6.19	123.40	120.30
2	B	296	ARG	NE-CZ-NH1	5.79	123.19	120.30
1	A	296	ARG	NE-CZ-NH2	-5.76	117.42	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	265	ASN	Peptide
2	B	6	ALA	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	4173	0	4007	23	0
2	B	4146	0	3986	36	0
3	A	6	0	8	0	0
3	B	6	0	8	0	0
4	A	12	0	13	2	0
4	B	12	0	13	4	0
5	A	1	0	0	0	0
All	All	8356	0	8035	61	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 61 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)
2:B:138:LEU:HA	2:B:477:MET:HE2	1.61	0.83
2:B:138:LEU:HA	2:B:477:MET:CE	2.12	0.79
2:B:211:MET:HG2	2:B:308:LEU:HD21	1.77	0.66
1:A:432:HIS:HB2	1:A:514:ASP:HB2	1.83	0.61
4:A:602:MES:H31	4:A:602:MES:O2S	2.02	0.58

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	524/548 (96%)	499 (95%)	25 (5%)	0	100	100
2	B	523/549 (95%)	502 (96%)	18 (3%)	3 (1%)	25	47
All	All	1047/1097 (95%)	1001 (96%)	43 (4%)	3 (0%)	41	64

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	442	TRP
2	B	347	SER
2	B	523	GLY

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	435/455 (96%)	422 (97%)	13 (3%)	41	67
2	B	432/456 (95%)	417 (96%)	15 (4%)	36	62
All	All	867/911 (95%)	839 (97%)	28 (3%)	39	65

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	491	GLU
2	B	70	TYR
2	B	467	LYS
2	B	30	SER
2	B	64	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	317	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)

4 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
3	VX	B	601	2	2,5,6	0.49	0	1,5,8	1.04	0
3	VX	A	601	1	2,5,6	1.08	0	1,5,8	0.54	0
4	MES	A	602	-	12,12,12	2.71	5 (41%)	14,16,16	2.38	6 (42%)
4	MES	B	602	-	12,12,12	2.77	6 (50%)	14,16,16	2.25	5 (35%)

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
3	VX	B	601	2	-	0/1/3/4	-
3	VX	A	601	1	-	1/1/3/4	-
4	MES	A	602	-	-	2/6/14/14	0/1/1/1
4	MES	B	602	-	-	6/6/14/14	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	602	MES	C8-S	6.19	1.86	1.77
4	A	602	MES	C8-S	6.15	1.86	1.77
4	B	602	MES	O2S-S	3.70	1.56	1.45
4	B	602	MES	O1S-S	3.52	1.55	1.45
4	A	602	MES	O1S-S	3.38	1.55	1.45

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	602	MES	O2S-S-C8	6.08	114.23	106.92
4	B	602	MES	O2S-S-C8	4.36	112.17	106.92
4	A	602	MES	O1S-S-C8	4.15	111.91	106.92
4	B	602	MES	C6-O1-C2	3.72	122.30	109.89
4	B	602	MES	C5-N4-C3	3.24	116.12	108.83

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	602	MES	N4-C7-C8-S
4	B	602	MES	N4-C7-C8-S
4	B	602	MES	C7-C8-S-O2S
4	B	602	MES	C7-C8-S-O3S
4	B	602	MES	C8-C7-N4-C5

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	602	MES	2	0
4	B	602	MES	4	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, 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	529/548 (96%)	-0.71	0	100	100	21, 36, 60, 96	0
2	B	528/549 (96%)	-0.62	0	100	100	23, 41, 67, 105	0
All	All	1057/1097 (96%)	-0.66	0	100	100	21, 38, 65, 105	0

There are no RSRZ outliers to report.

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. 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
4	MES	B	602	12/12	0.75	0.24	39,51,88,94	0
4	MES	A	602	12/12	0.80	0.21	43,57,103,105	0
3	VX	B	601	6/7	0.98	0.16	32,36,51,51	0
3	VX	A	601	6/7	0.98	0.13	39,42,45,46	0

6.5 Other polymers (i)

There are no such residues in this entry.