



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:02 PM GMT

PDB ID : 3SO4
Title : Methionine-adenosyltransferase from *Entamoeba histolytica*
Authors : Merritt, E.A.; Medical Structural Genomics of Pathogenic Protozoa (MSGPP)
Deposited on : 2011-06-29
Resolution : 3.18 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
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

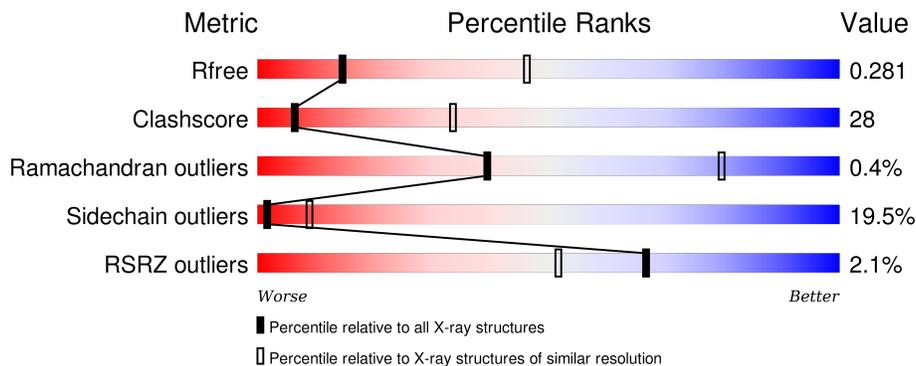
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.18 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}	91344	1115 (3.22-3.14)
Clashscore	102246	1125 (3.20-3.16)
Ramachandran outliers	100387	1105 (3.20-3.16)
Sidechain outliers	100360	1104 (3.20-3.16)
RSRZ outliers	91569	1120 (3.22-3.14)

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. 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	415	 2% 54% 29% 8% 9%
1	B	415	 2% 50% 32% 9% 9%
1	C	415	 3% 54% 27% 10% 9%
1	D	415	 2% 53% 32% 6% 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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	ACT	B	400	-	-	-	X
2	ACT	C	400	-	-	X	X

2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 11498 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 Methionine-adenosyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	379	2858	1804	487	549	18	0	0	0
1	B	379	2881	1818	491	554	18	0	0	0
1	C	379	2865	1807	490	550	18	0	0	0
1	D	379	2874	1815	491	550	18	0	0	0

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-17	MET	-	EXPRESSION TAG	UNP C4M272
A	-16	ALA	-	EXPRESSION TAG	UNP C4M272
A	-15	HIS	-	EXPRESSION TAG	UNP C4M272
A	-14	HIS	-	EXPRESSION TAG	UNP C4M272
A	-13	HIS	-	EXPRESSION TAG	UNP C4M272
A	-12	HIS	-	EXPRESSION TAG	UNP C4M272
A	-11	HIS	-	EXPRESSION TAG	UNP C4M272
A	-10	HIS	-	EXPRESSION TAG	UNP C4M272
A	-9	MET	-	EXPRESSION TAG	UNP C4M272
A	-8	GLY	-	EXPRESSION TAG	UNP C4M272
A	-7	THR	-	EXPRESSION TAG	UNP C4M272
A	-6	LEU	-	EXPRESSION TAG	UNP C4M272
A	-5	GLU	-	EXPRESSION TAG	UNP C4M272
A	-4	ALA	-	EXPRESSION TAG	UNP C4M272
A	-3	GLN	-	EXPRESSION TAG	UNP C4M272
A	-2	THR	-	EXPRESSION TAG	UNP C4M272
A	-1	GLN	-	EXPRESSION TAG	UNP C4M272
A	0	GLY	-	EXPRESSION TAG	UNP C4M272
A	1	PRO	-	EXPRESSION TAG	UNP C4M272
A	2	GLY	-	EXPRESSION TAG	UNP C4M272
A	3	SER	-	EXPRESSION TAG	UNP C4M272

Continued on next page...

Continued from previous page...

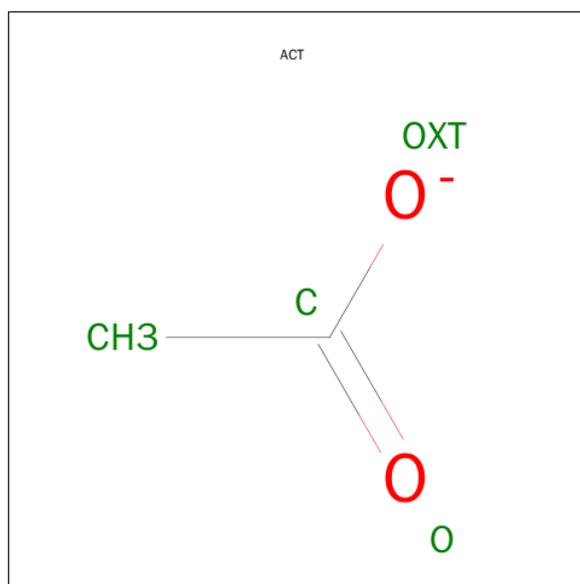
Chain	Residue	Modelled	Actual	Comment	Reference
B	-17	MET	-	EXPRESSION TAG	UNP C4M272
B	-16	ALA	-	EXPRESSION TAG	UNP C4M272
B	-15	HIS	-	EXPRESSION TAG	UNP C4M272
B	-14	HIS	-	EXPRESSION TAG	UNP C4M272
B	-13	HIS	-	EXPRESSION TAG	UNP C4M272
B	-12	HIS	-	EXPRESSION TAG	UNP C4M272
B	-11	HIS	-	EXPRESSION TAG	UNP C4M272
B	-10	HIS	-	EXPRESSION TAG	UNP C4M272
B	-9	MET	-	EXPRESSION TAG	UNP C4M272
B	-8	GLY	-	EXPRESSION TAG	UNP C4M272
B	-7	THR	-	EXPRESSION TAG	UNP C4M272
B	-6	LEU	-	EXPRESSION TAG	UNP C4M272
B	-5	GLU	-	EXPRESSION TAG	UNP C4M272
B	-4	ALA	-	EXPRESSION TAG	UNP C4M272
B	-3	GLN	-	EXPRESSION TAG	UNP C4M272
B	-2	THR	-	EXPRESSION TAG	UNP C4M272
B	-1	GLN	-	EXPRESSION TAG	UNP C4M272
B	0	GLY	-	EXPRESSION TAG	UNP C4M272
B	1	PRO	-	EXPRESSION TAG	UNP C4M272
B	2	GLY	-	EXPRESSION TAG	UNP C4M272
B	3	SER	-	EXPRESSION TAG	UNP C4M272
C	-17	MET	-	EXPRESSION TAG	UNP C4M272
C	-16	ALA	-	EXPRESSION TAG	UNP C4M272
C	-15	HIS	-	EXPRESSION TAG	UNP C4M272
C	-14	HIS	-	EXPRESSION TAG	UNP C4M272
C	-13	HIS	-	EXPRESSION TAG	UNP C4M272
C	-12	HIS	-	EXPRESSION TAG	UNP C4M272
C	-11	HIS	-	EXPRESSION TAG	UNP C4M272
C	-10	HIS	-	EXPRESSION TAG	UNP C4M272
C	-9	MET	-	EXPRESSION TAG	UNP C4M272
C	-8	GLY	-	EXPRESSION TAG	UNP C4M272
C	-7	THR	-	EXPRESSION TAG	UNP C4M272
C	-6	LEU	-	EXPRESSION TAG	UNP C4M272
C	-5	GLU	-	EXPRESSION TAG	UNP C4M272
C	-4	ALA	-	EXPRESSION TAG	UNP C4M272
C	-3	GLN	-	EXPRESSION TAG	UNP C4M272
C	-2	THR	-	EXPRESSION TAG	UNP C4M272
C	-1	GLN	-	EXPRESSION TAG	UNP C4M272
C	0	GLY	-	EXPRESSION TAG	UNP C4M272
C	1	PRO	-	EXPRESSION TAG	UNP C4M272
C	2	GLY	-	EXPRESSION TAG	UNP C4M272
C	3	SER	-	EXPRESSION TAG	UNP C4M272

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-17	MET	-	EXPRESSION TAG	UNP C4M272
D	-16	ALA	-	EXPRESSION TAG	UNP C4M272
D	-15	HIS	-	EXPRESSION TAG	UNP C4M272
D	-14	HIS	-	EXPRESSION TAG	UNP C4M272
D	-13	HIS	-	EXPRESSION TAG	UNP C4M272
D	-12	HIS	-	EXPRESSION TAG	UNP C4M272
D	-11	HIS	-	EXPRESSION TAG	UNP C4M272
D	-10	HIS	-	EXPRESSION TAG	UNP C4M272
D	-9	MET	-	EXPRESSION TAG	UNP C4M272
D	-8	GLY	-	EXPRESSION TAG	UNP C4M272
D	-7	THR	-	EXPRESSION TAG	UNP C4M272
D	-6	LEU	-	EXPRESSION TAG	UNP C4M272
D	-5	GLU	-	EXPRESSION TAG	UNP C4M272
D	-4	ALA	-	EXPRESSION TAG	UNP C4M272
D	-3	GLN	-	EXPRESSION TAG	UNP C4M272
D	-2	THR	-	EXPRESSION TAG	UNP C4M272
D	-1	GLN	-	EXPRESSION TAG	UNP C4M272
D	0	GLY	-	EXPRESSION TAG	UNP C4M272
D	1	PRO	-	EXPRESSION TAG	UNP C4M272
D	2	GLY	-	EXPRESSION TAG	UNP C4M272
D	3	SER	-	EXPRESSION TAG	UNP C4M272

- Molecule 2 is ACETATE ION (three-letter code: ACT) (formula: C₂H₃O₂).



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

Continued on next page...

Continued from previous page...

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

- Molecule 3 is water.

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

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	58.49Å 113.16Å 220.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.18 29.92 – 3.18	Depositor EDS
% Data completeness (in resolution range)	91.1 (50.00-3.18) 91.3 (29.92-3.18)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.50 (at 3.18Å)	Xtrriage
Refinement program	REFMAC 5.6.0095	Depositor
R, R_{free}	0.228 , 0.284 0.230 , 0.281	Depositor DCC
R_{free} test set	1295 reflections (5.86%)	DCC
Wilson B-factor (Å ²)	76.0	Xtrriage
Anisotropy	0.453	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 54.3	EDS
Estimated twinning fraction	No twinning to report.	Xtrriage
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtrriage
Outliers	1 of 23419 reflections (0.004%)	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	11498	wwPDB-VP
Average B, all atoms (Å ²)	80.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.43% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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:
ACT

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	0.44	0/2914	0.69	2/3952 (0.1%)
1	B	0.46	0/2937	0.72	1/3978 (0.0%)
1	C	0.43	0/2921	0.70	1/3959 (0.0%)
1	D	0.42	0/2930	0.68	1/3969 (0.0%)
All	All	0.44	0/11702	0.70	5/15858 (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	B	0	1

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	173	LEU	CA-CB-CG	5.73	128.47	115.30
1	C	55	LEU	CA-CB-CG	5.64	128.28	115.30
1	A	213	VAL	CB-CA-C	-5.60	100.76	111.40
1	D	224	ASP	CB-CG-OD2	-5.44	113.41	118.30
1	B	343	ARG	NE-CZ-NH1	5.04	122.82	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	239	ILE	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	2858	0	2785	150	0
1	B	2881	0	2830	182	0
1	C	2865	0	2799	180	0
1	D	2874	0	2824	161	0
2	A	4	0	3	1	0
2	B	4	0	3	0	0
2	C	4	0	3	2	0
2	D	4	0	3	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
All	All	11498	0	11250	630	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 28.

The worst 5 of 630 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:209:LEU:O	1:A:213:VAL:HG22	1.24	1.29
1:A:213:VAL:CG2	1:A:214:THR:H	1.37	1.28
1:D:47:THR:CG2	1:D:52:ILE:HD13	1.63	1.27
1:A:213:VAL:HG23	1:A:214:THR:N	1.26	1.20
1:C:326:ILE:HD12	1:C:327:CYS:N	1.58	1.16

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	375/415 (90%)	352 (94%)	22 (6%)	1 (0%)	46	84
1	B	375/415 (90%)	351 (94%)	24 (6%)	0	100	100
1	C	375/415 (90%)	349 (93%)	22 (6%)	4 (1%)	17	62
1	D	375/415 (90%)	349 (93%)	25 (7%)	1 (0%)	46	84
All	All	1500/1660 (90%)	1401 (93%)	93 (6%)	6 (0%)	39	80

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	121	ASP
1	A	102	GLN
1	C	327	CYS
1	C	392	LEU
1	C	117	ILE

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	306/351 (87%)	248 (81%)	58 (19%)	2	10
1	B	312/351 (89%)	249 (80%)	63 (20%)	1	7
1	C	308/351 (88%)	241 (78%)	67 (22%)	1	6
1	D	310/351 (88%)	257 (83%)	53 (17%)	2	12
All	All	1236/1404 (88%)	995 (80%)	241 (20%)	2	9

5 of 241 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	348	ILE
1	C	82	THR
1	D	251	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	357	ILE
1	C	19	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	122	GLN
1	C	304	GLN
1	B	215	GLN
1	A	339	ASN
1	B	339	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	$\# Z > 2$
2	ACT	A	400	-	1,3,3	1.47	0	0,3,3	0.00	-
2	ACT	B	400	-	1,3,3	1.09	0	0,3,3	0.00	-

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	ACT	C	400	-	1,3,3	1.42	0	0,3,3	0.00	-
2	ACT	D	400	-	1,3,3	1.44	0	0,3,3	0.00	-

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	ACT	A	400	-	-	0/0/0/0	0/0/0/0
2	ACT	B	400	-	-	0/0/0/0	0/0/0/0
2	ACT	C	400	-	-	0/0/0/0	0/0/0/0
2	ACT	D	400	-	-	0/0/0/0	0/0/0/0

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.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	400	ACT	1	0
2	C	400	ACT	2	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	379/415 (91%)	-0.16	5 (1%) 79 67	47, 68, 99, 129	0
1	B	379/415 (91%)	-0.12	6 (1%) 74 62	44, 73, 108, 124	0
1	C	379/415 (91%)	0.03	11 (2%) 55 40	49, 80, 119, 139	0
1	D	379/415 (91%)	0.07	10 (2%) 59 45	54, 89, 122, 142	0
All	All	1516/1660 (91%)	-0.05	32 (2%) 67 52	44, 76, 114, 142	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	103	ASP	5.2
1	D	199	ASN	4.5
1	A	327	CYS	4.1
1	C	104	ILE	4.1
1	D	200	VAL	3.1

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
2	ACT	C	400	4/4	0.94	0.38	3.87	43,47,47,48	0
2	ACT	B	400	4/4	0.97	0.35	3.36	31,32,32,35	0
2	ACT	A	400	4/4	0.94	0.30	1.72	39,40,41,43	0
2	ACT	D	400	4/4	0.93	0.21	-0.06	35,37,38,40	0

6.5 Other polymers [i](#)

There are no such residues in this entry.