



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 11:41 PM GMT

PDB ID : 1Y8R
Title : SUMO E1 ACTIVATING ENZYME SAE1-SAE2-SUMO1-MG-ATP COMPLEX
Authors : Lois, L.M.; Lima, C.D.
Deposited on : 2004-12-13
Resolution : 2.75 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

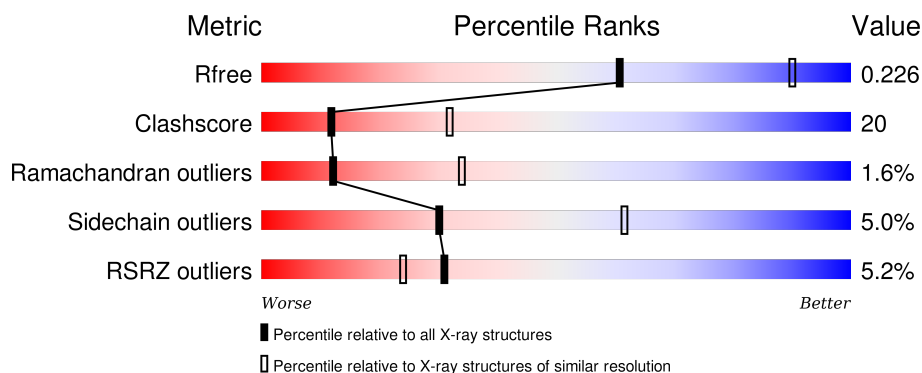
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.75 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	3340 (2.80-2.72)
Clashscore	102246	3829 (2.80-2.72)
Ramachandran outliers	100387	3767 (2.80-2.72)
Sidechain outliers	100360	3770 (2.80-2.72)
RSRZ outliers	91569	3352 (2.80-2.72)

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	346	<div> <div>3%</div> <div> <div></div> <div>55%</div> <div>34%</div> <div>• 9%</div> </div> </div>
1	D	346	<div> <div>•</div> <div> <div></div> <div>59%</div> <div>29%</div> <div>• 10%</div> </div> </div>
2	B	640	<div> <div>4%</div> <div> <div></div> <div>49%</div> <div>29%</div> <div>• 20%</div> </div> </div>
2	E	640	<div> <div>5%</div> <div> <div></div> <div>50%</div> <div>28%</div> <div>• 20%</div> </div> </div>
3	C	97	<div> <div>5%</div> <div> <div></div> <div>42%</div> <div>41%</div> <div>• • 13%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	97	<p>22% 33% 48% 13%</p>

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 14695 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 Ubiquitin-like 1 activating enzyme E1A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	314	Total	C	N	O	S	0	0	0
			2462	1559	418	472	13			
1	D	311	Total	C	N	O	S	0	0	0
			2437	1544	414	466	13			

- Molecule 2 is a protein called Ubiquitin-like 2 activating enzyme E1B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	512	Total	C	N	O	S	0	0	0
			4022	2556	695	752	19			
2	E	514	Total	C	N	O	S	0	0	0
			4031	2562	696	754	19			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	173	ALA	CYS	ENGINEERED	UNP Q9UBT2
E	173	ALA	CYS	ENGINEERED	UNP Q9UBT2

- Molecule 3 is a protein called Ubiquitin-like protein SMT3C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	84	Total	C	N	O	S	0	0	0
			680	425	117	134	4			
3	F	84	Total	C	N	O	S	0	0	0
			680	425	117	134	4			

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		

Continued on next page...

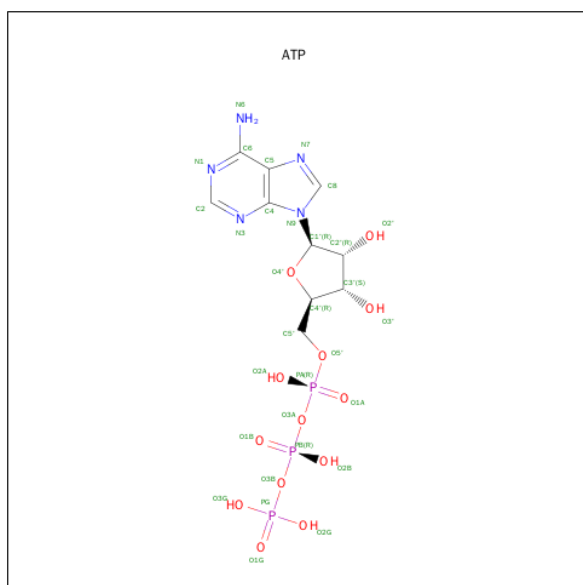
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total	Mg	0	0
			1	1		

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Zn	0	0
			1	1		
5	E	1	Total	Zn	0	0
			1	1		

- Molecule 6 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: C₁₀H₁₆N₅O₁₃P₃).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	E	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
6	B	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	61	Total	O	0	0
			61	61		

Continued on next page...

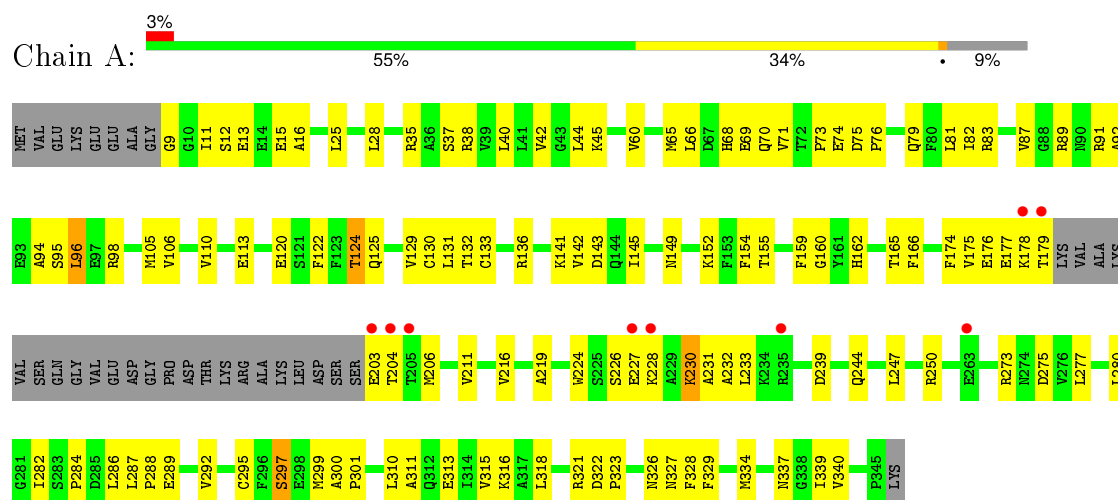
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	B	89	Total 89	O 89	0	0
7	C	13	Total 13	O 13	0	0
7	D	62	Total 62	O 62	0	0
7	E	85	Total 85	O 85	0	0
7	F	7	Total 7	O 7	0	0

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 errors 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: Ubiquitin-like 1 activating enzyme E1A

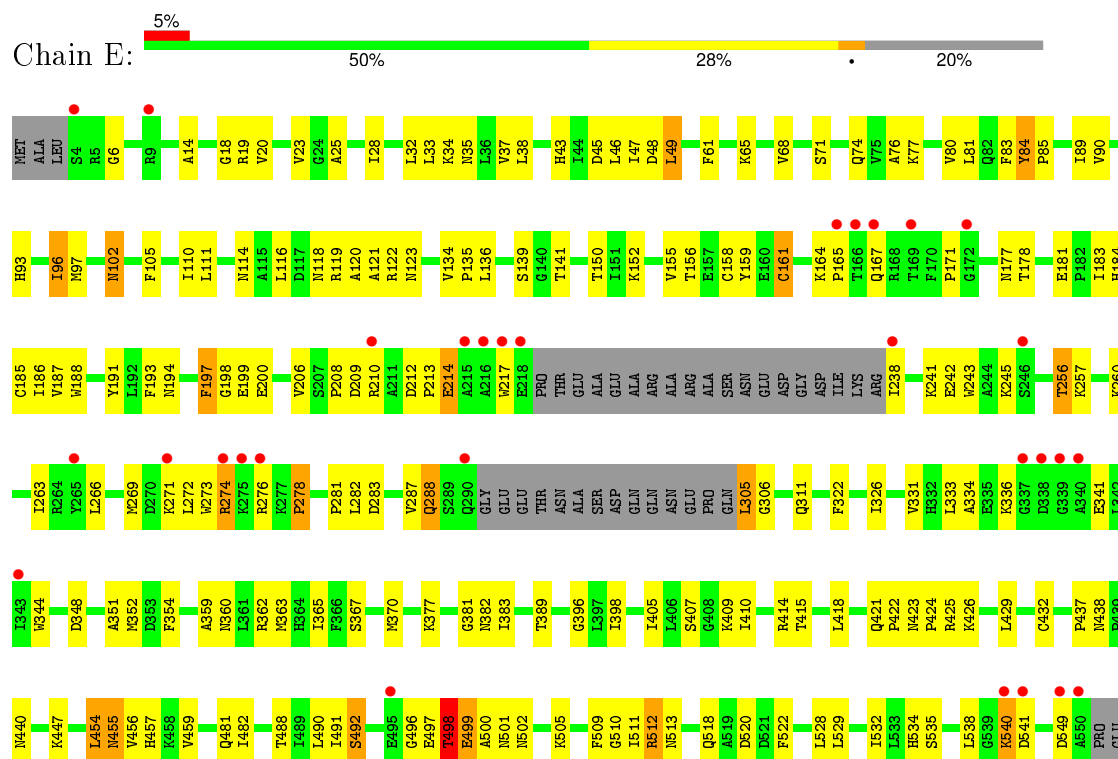


- Molecule 1: Ubiquitin-like 1 activating enzyme E1A



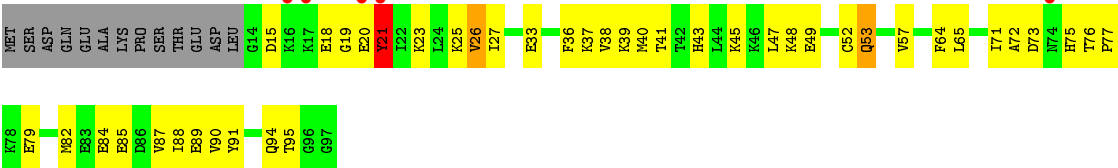
- Molecule 2: Ubiquitin-like 2 activating enzyme E1B



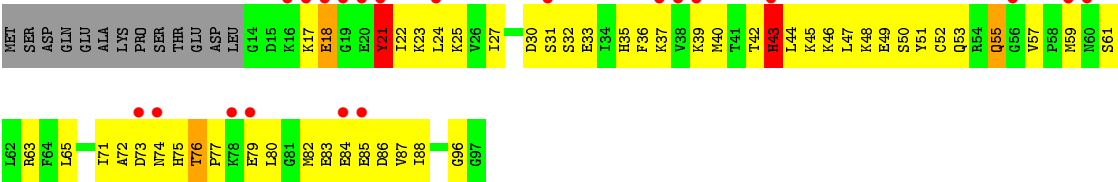


LYS	VAL	GLY	PRO	LYS	GLN	ALA	GLU	ASP	ALA	LYS	SER	ILE	THR	ASN	GLN	LYS	ASP	GLY	ASP	GLN	ALA	PRO	SER	THR	SER	THR	ALA	GLN	GLN	GLU	GLU	VAL	LEU	ILE	VAL	VAL	ASP	SER	ASP	GLU	GLU	ASP	SER	SER	ASN	ASN	ALA	ASP	VAL	SER	GLU	GLU	GLU	ARG	SER	ARG	LYS	ARG
LYS	LEU	ASP	GLU	LYS	GLU	ASN	LEU	SER	ALA	LYS	ARG	SER	ARG	ILE	GLU	GLN	LYS	GLU	LEU	ASP	VAL	ILE	ALA	LEU	ASP																																	

• Molecule 3: Ubiquitin-like protein SMT3C



• Molecule 3: Ubiquitin-like protein SMT3C



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2	Depositor
Cell constants a, b, c, α , β , γ	117.20Å 215.15Å 100.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.99 – 2.75 19.99 – 2.75	Depositor EDS
% Data completeness (in resolution range)	94.9 (19.99-2.75) 95.1 (19.99-2.75)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.25 (at 2.75Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.220 , 0.276 0.222 , 0.226	Depositor DCC
R_{free} test set	3228 reflections (5.08%)	DCC
Wilson B-factor (Å ²)	43.3	Xtriage
Anisotropy	0.114	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 59.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.25$	Xtriage
Outliers	0 of 63862 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	14695	wwPDB-VP
Average B, all atoms (Å ²)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 18.76% 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: MG, ZN, ATP

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	0.47	0/2505	0.59	0/3379
1	D	0.51	0/2480	0.61	0/3346
2	B	0.42	0/4099	0.60	0/5548
2	E	0.46	0/4108	0.61	0/5560
3	C	0.40	0/690	0.55	0/918
3	F	0.38	0/690	0.52	0/918
All	All	0.45	0/14572	0.60	0/19669

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	2462	0	2451	94	0
1	D	2437	0	2425	82	0
2	B	4022	0	4058	160	0
2	E	4031	0	4067	159	0
3	C	680	0	674	39	0
3	F	680	0	674	60	0
4	B	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	1	0	0	0	0
5	B	1	0	0	0	0
5	E	1	0	0	0	0
6	B	31	0	12	0	0
6	E	31	0	12	1	0
7	A	61	0	0	3	0
7	B	89	0	0	6	0
7	C	13	0	0	0	0
7	D	62	0	0	4	0
7	E	85	0	0	7	0
7	F	7	0	0	0	0
All	All	14695	0	14373	577	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 20.

The worst 5 of 577 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:465:GLN:HB2	2:B:491:ILE:HD13	1.43	0.99
1:A:40:LEU:HD12	1:A:129:VAL:HG13	1.51	0.92
2:E:490:LEU:HD11	2:E:511:ILE:HD11	1.52	0.91
3:F:21:TYR:HB3	3:F:22:ILE:HD12	1.54	0.88
1:D:22:GLN:HE21	1:D:31:GLN:HE22	1.23	0.85

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	310/346 (90%)	264 (85%)	43 (14%)	3 (1%)	19 48

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	307/346 (89%)	264 (86%)	40 (13%)	3 (1%)	19	48
2	B	506/640 (79%)	433 (86%)	67 (13%)	6 (1%)	16	43
2	E	508/640 (79%)	452 (89%)	46 (9%)	10 (2%)	9	27
3	C	82/97 (84%)	70 (85%)	9 (11%)	3 (4%)	4	12
3	F	82/97 (84%)	68 (83%)	10 (12%)	4 (5%)	3	7
All	All	1795/2166 (83%)	1551 (86%)	215 (12%)	29 (2%)	12	34

5 of 29 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	8	PRO
1	D	205	THR
2	B	7	LEU
2	B	205	GLU
2	B	206	VAL

5.3.2 Protein sidechains ⓘ

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	270/296 (91%)	258 (96%)	12 (4%)	35	67
1	D	267/296 (90%)	257 (96%)	10 (4%)	41	74
2	B	442/551 (80%)	419 (95%)	23 (5%)	29	60
2	E	443/551 (80%)	421 (95%)	22 (5%)	30	61
3	C	75/87 (86%)	71 (95%)	4 (5%)	28	58
3	F	75/87 (86%)	68 (91%)	7 (9%)	11	29
All	All	1572/1868 (84%)	1494 (95%)	78 (5%)	30	61

5 of 78 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	21	TYR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	216	VAL
3	F	21	TYR
3	C	26	VAL
1	D	26	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 42 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	74	ASN
1	D	244	GLN
3	F	60	ASN
3	C	75	HIS
1	D	31	GLN

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](#)

Of 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 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
6	ATP	B	802	4	24,33,33	1.28	3 (12%)	31,52,52	2.62	7 (22%)
6	ATP	E	801	4	24,33,33	1.18	2 (8%)	31,52,52	2.67	7 (22%)

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
6	ATP	B	802	4	-	0/18/38/38	0/3/3/3
6	ATP	E	801	4	-	0/18/38/38	0/3/3/3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	E	801	ATP	C2-N1	2.02	1.37	1.33
6	B	802	ATP	C4-N3	2.38	1.39	1.35
6	B	802	ATP	C2-N1	2.59	1.38	1.33
6	E	801	ATP	C2-N3	3.60	1.38	1.32
6	B	802	ATP	C2-N3	3.76	1.38	1.32

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	B	802	ATP	N3-C2-N1	-11.69	119.94	128.89
6	E	801	ATP	N3-C2-N1	-11.61	120.00	128.89
6	B	802	ATP	C4-C5-N7	-4.70	105.16	109.48
6	E	801	ATP	C4-C5-N7	-4.51	105.33	109.48
6	B	802	ATP	PB-O3B-PG	-3.19	121.96	132.67

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	E	801	ATP	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.

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, 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	314/346 (90%)	-0.20	9 (2%) 55 48	13, 41, 82, 112	0
1	D	311/346 (89%)	-0.28	4 (1%) 79 75	11, 39, 73, 94	0
2	B	512/640 (80%)	-0.00	26 (5%) 32 24	17, 53, 101, 131	0
2	E	514/640 (80%)	-0.01	30 (5%) 26 20	15, 44, 102, 133	0
3	C	84/97 (86%)	0.17	5 (5%) 25 19	26, 63, 91, 101	0
3	F	84/97 (86%)	1.14	21 (25%) 1 0	29, 86, 107, 115	0
All	All	1819/2166 (83%)	-0.02	95 (5%) 31 24	11, 46, 97, 133	0

The worst 5 of 95 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	19	GLY	6.5
3	F	20	GLU	5.9
2	B	217	TRP	5.4
2	E	290	GLN	5.1
1	A	204	THR	4.9

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 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(Å ²)	Q<0.9
6	ATP	B	802	31/31	0.98	0.10	-1.19	19,28,34,35	0
6	ATP	E	801	31/31	0.99	0.10	-1.35	20,26,29,32	0
5	ZN	B	642	1/1	0.99	0.07	-2.13	37,37,37,37	0
5	ZN	E	642	1/1	0.99	0.04	-2.43	31,31,31,31	0
4	MG	B	641	1/1	0.96	0.04	-	16,16,16,16	0
4	MG	E	641	1/1	0.99	0.10	-	31,31,31,31	0

6.5 Other polymers [i](#)

There are no such residues in this entry.