WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 01:54 am BST

PDB ID	:	1SSM
Title	:	Serine Acetyltransferase- Apoenzyme (truncated)
Authors	:	Olsen, L.R.; Huang, B.; Vetting, M.W.; Roderick, S.L.
Deposited on	:	2004-03-24
Resolution	:	2.15 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:

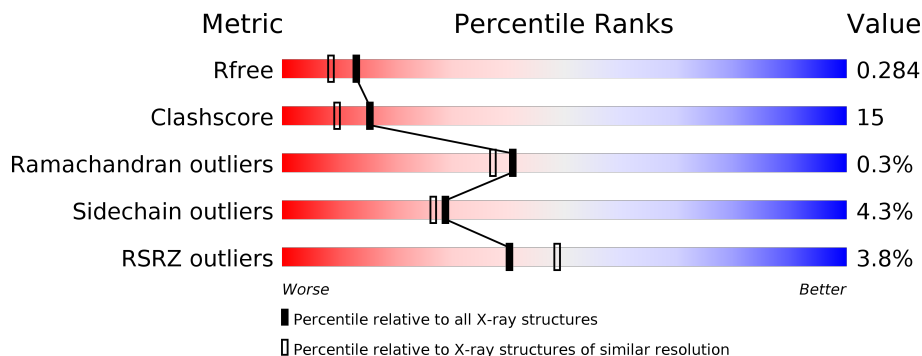|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 2.15 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	Percentile Ranks	Value
Rfree		0.284
Clashscore		15
Ramachandran outliers		0.3%
Sidechain outliers		4.3%
RSRZ outliers		3.8%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	242	5% 71% 26% ··
1	B	242	3% 71% 26% ··
1	C	242	3% 71% 27% ··
1	D	242	3% 76% 22% ··
1	E	242	5% 70% 27% ··
1	F	242	4% 65% 31% ··

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 11514 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 Serine acetyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	240	Total 1824	C 1162	N 320	O 336	S 2	Se 4	0	0	0
1	B	240	Total 1826	C 1163	N 319	O 337	S 2	Se 5	0	0	0
1	C	240	Total 1819	C 1159	N 318	O 336	S 2	Se 4	0	0	0
1	D	240	Total 1823	C 1162	N 319	O 336	S 2	Se 4	0	0	0
1	E	240	Total 1831	C 1166	N 321	O 337	S 2	Se 5	0	0	0
1	F	240	Total 1822	C 1162	N 319	O 335	S 2	Se 4	0	0	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP P43886
A	22	MSE	MET	MODIFIED RESIDUE	UNP P43886
A	54	MSE	MET	MODIFIED RESIDUE	UNP P43886
A	151	MSE	MET	MODIFIED RESIDUE	UNP P43886
A	197	MSE	MET	MODIFIED RESIDUE	UNP P43886
B	1	MSE	MET	MODIFIED RESIDUE	UNP P43886
B	22	MSE	MET	MODIFIED RESIDUE	UNP P43886
B	54	MSE	MET	MODIFIED RESIDUE	UNP P43886
B	151	MSE	MET	MODIFIED RESIDUE	UNP P43886
B	197	MSE	MET	MODIFIED RESIDUE	UNP P43886
C	1	MSE	MET	MODIFIED RESIDUE	UNP P43886
C	22	MSE	MET	MODIFIED RESIDUE	UNP P43886
C	54	MSE	MET	MODIFIED RESIDUE	UNP P43886
C	151	MSE	MET	MODIFIED RESIDUE	UNP P43886
C	197	MSE	MET	MODIFIED RESIDUE	UNP P43886
D	1	MSE	MET	MODIFIED RESIDUE	UNP P43886
D	22	MSE	MET	MODIFIED RESIDUE	UNP P43886

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	54	MSE	MET	MODIFIED RESIDUE	UNP P43886
D	151	MSE	MET	MODIFIED RESIDUE	UNP P43886
D	197	MSE	MET	MODIFIED RESIDUE	UNP P43886
E	1	MSE	MET	MODIFIED RESIDUE	UNP P43886
E	22	MSE	MET	MODIFIED RESIDUE	UNP P43886
E	54	MSE	MET	MODIFIED RESIDUE	UNP P43886
E	151	MSE	MET	MODIFIED RESIDUE	UNP P43886
E	197	MSE	MET	MODIFIED RESIDUE	UNP P43886
F	1	MSE	MET	MODIFIED RESIDUE	UNP P43886
F	22	MSE	MET	MODIFIED RESIDUE	UNP P43886
F	54	MSE	MET	MODIFIED RESIDUE	UNP P43886
F	151	MSE	MET	MODIFIED RESIDUE	UNP P43886
F	197	MSE	MET	MODIFIED RESIDUE	UNP P43886

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	101	Total 101	O 101	0	0
2	B	96	Total 96	O 96	0	0
2	C	89	Total 89	O 89	0	0
2	D	91	Total 91	O 91	0	0
2	E	86	Total 86	O 86	0	0
2	F	106	Total 106	O 106	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 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: Serine acetyltransferase

- Molecule 1: Serine acetyltransferase

- Molecule 1: Serine acetyltransferase

L124 Y127 L128 I132 D137 I140 H141 P142 A143 G149 I150 D153 H154 A155 V160 G161 E162 D169 L173 V176 T177 L178 G179 G182 K183 E184 S185 G186 D187 R188 H189 P190 K191 V192 R193 E194 M197 I198 G199 K203 K212 Y213 R238 I239 V240

SER GLN

• **Molecule 1: Serine acetyltransferase**

Chain D:

3%
76% 22%

M1 T2 L3 D4 V5 K14 E15 L16 A17 E18 L44 K48 N51 P52 I57 R60 E61 I62 E65 A77 A78 I81 P89 A90 V91 E92 T96 P97 L101 Q108 S109 Y110 Y115 L124 I132 D137 H141 P142 A143 D153 H154

A155 I158 V159 L173 Q174 G175 V176 T177 L178 G182 K183 E184 S185 G186 D187 R188 H189 P190 K191 E194 M197 A202 K203 I204 K212 Y213 A214 K215 N224 P225 E228 Y229 A230 T231 R238 I239 V240 SER GLN

• **Molecule 1: Serine acetyltransferase**

Chain E:

5%
70% 27%

M1 T2 L3 D4 V5 I9 L16 E20 L23 A24 S25 H28 L32 Q35 L41 A46 N47 K48 L49 A50 N51 P52 L59 I63 E64 E65 A66 Y67 Q68 S69 C76 Q82 A83 H86 R87 D88 P89 A90 W94 S95 T96 P97 F104 H105

Q108 S109 Y110 R111 I112 T113 L128 I132 S133 D137 H141 P142 A143 G149 F152 D153 H154 A155 T156 E162 I172 L173 V176 T177 L178 G182 K183 E184 S185 G186 D187 P190 K191 E194 G195 V196 M197 I204 Y213 V226 P227 E228 Y229

R238 I239 V240 SER GLN

• **Molecule 1: Serine acetyltransferase**

Chain F:

4%
65% 31%

M1 T2 Q7 Q11 E12 A13 K14 E15 L16 A17 E18 L23 F26 F27 I31 L32 K33 Y43 L44 K48 N51 P52 I53 M54 S58 L59 R60 I63 Q68 S69 N70 I73 I74 D75 Q82 R85 H86 R87 D88 P89 A90 V91 E92 T96 P97

L98 L99 F104 Q108 S109 Y110 R111 I112 Y115 L116 W117 K122 I132 D137 I140 H141 P142 A143 A144 K145 G149 F152 D153 H154 A155 T156 N168 D169 I172 L173 V176 T177 G182 K183 E184 S185 G186 D187 R188 K191 E194 G195 V196 M197 I198

A202 K203 Y213 A214 K215 S220 P225 V226 P227 T231 A232 A233 I239 V240 SER GLN

4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	107.10Å 126.60Å 107.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.74 – 2.15 29.74 – 2.13	Depositor EDS
% Data completeness (in resolution range)	88.9 (29.74-2.15) 90.5 (29.74-2.13)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.52 (at 2.14Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.227 , 0.278 0.231 , 0.284	Depositor DCC
R_{free} test set	7181 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å^2)	22.4	Xtriage
Anisotropy	0.336	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 41.5	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.000 for l,-k,h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11514	wwPDB-VP
Average B, all atoms (Å^2)	27.0	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 47.91 % 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 9.2745e-05. 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 (i)

5.1 Standard geometry (i)

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	0.83	2/1859 (0.1%)	0.93	3/2525 (0.1%)
1	B	0.84	1/1860 (0.1%)	0.95	2/2525 (0.1%)
1	C	0.84	0/1853	0.93	2/2517 (0.1%)
1	D	0.82	0/1857	0.93	2/2521 (0.1%)
1	E	0.85	1/1866 (0.1%)	0.93	2/2533 (0.1%)
1	F	0.81	1/1857 (0.1%)	0.92	3/2522 (0.1%)
All	All	0.83	5/11152 (0.0%)	0.93	14/15143 (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	B	0	1
1	C	0	1
1	E	0	1
All	All	0	3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	76	CYS	CB-SG	7.41	1.94	1.82
1	A	15	GLU	CB-CG	5.52	1.62	1.52
1	A	15	GLU	CG-CD	5.36	1.59	1.51
1	F	43	TYR	CD2-CE2	5.22	1.47	1.39
1	B	92	GLU	CD-OE2	5.06	1.31	1.25

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	183	LYS	N-CA-C	-7.08	91.89	111.00
1	B	178	LEU	N-CA-C	-6.80	92.63	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	88	ASP	CB-CG-OD1	-6.55	112.41	118.30
1	E	3	LEU	CB-CG-CD1	-6.50	99.96	111.00
1	A	178	LEU	N-CA-C	-6.35	93.85	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	213	TYR	Sidechain
1	C	127	TYR	Sidechain
1	E	213	TYR	Sidechain

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	1824	0	1821	50	0
1	B	1826	0	1829	69	0
1	C	1819	0	1816	54	0
1	D	1823	0	1827	53	0
1	E	1831	0	1834	52	0
1	F	1822	0	1823	66	0
2	A	101	0	0	14	0
2	B	96	0	0	14	0
2	C	89	0	0	11	0
2	D	91	0	0	9	0
2	E	86	0	0	4	0
2	F	106	0	0	9	0
All	All	11514	0	10950	324	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 15.

The worst 5 of 324 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:53:ILE:HB	2:A:7567:HOH:O	1.47	1.12
1:E:141:HIS:HD2	1:E:143:ALA:H	1.00	0.98
1:E:3:LEU:HG	2:E:7410:HOH:O	1.65	0.95
1:E:141:HIS:CD2	1:E:143:ALA:H	1.85	0.94
1:C:141:HIS:HD2	1:C:143:ALA:H	1.04	0.94

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	238/242 (98%)	229 (96%)	8 (3%)	1 (0%)	34	29
1	B	238/242 (98%)	225 (94%)	13 (6%)	0	100	100
1	C	238/242 (98%)	229 (96%)	9 (4%)	0	100	100
1	D	238/242 (98%)	224 (94%)	13 (6%)	1 (0%)	34	29
1	E	238/242 (98%)	225 (94%)	11 (5%)	2 (1%)	19	12
1	F	238/242 (98%)	227 (95%)	11 (5%)	0	100	100
All	All	1428/1452 (98%)	1359 (95%)	65 (5%)	4 (0%)	41	37

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	229	TYR
1	E	183	LYS
1	E	89	PRO
1	A	89	PRO

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	190/192 (99%)	183 (96%)	7 (4%)	34	32
1	B	191/192 (100%)	184 (96%)	7 (4%)	34	32
1	C	189/192 (98%)	181 (96%)	8 (4%)	30	28
1	D	190/192 (99%)	183 (96%)	7 (4%)	34	32
1	E	192/192 (100%)	183 (95%)	9 (5%)	26	23
1	F	190/192 (99%)	179 (94%)	11 (6%)	20	15
All	All	1142/1152 (99%)	1093 (96%)	49 (4%)	29	27

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	16	LEU
1	D	188	ARG
1	F	115	TYR
1	D	115	TYR
1	D	212	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	141	HIS
1	D	7	GLN
1	F	141	HIS
1	C	148	HIS
1	C	154	HIS

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)

There are no ligands in this entry.

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	235/242 (97%)	0.04	11 (4%)	31	41	16, 25, 41, 56	0
1	B	235/242 (97%)	0.05	7 (2%)	50	59	16, 25, 41, 55	0
1	C	235/242 (97%)	0.09	7 (2%)	50	59	17, 25, 42, 52	0
1	D	235/242 (97%)	0.05	8 (3%)	45	53	15, 24, 40, 50	0
1	E	235/242 (97%)	0.12	11 (4%)	31	41	16, 25, 42, 52	0
1	F	235/242 (97%)	-0.02	9 (3%)	40	49	16, 23, 41, 53	0
All	All	1410/1452 (97%)	0.05	53 (3%)	40	49	15, 25, 42, 56	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	184	GLU	5.8
1	B	185	SER	5.7
1	C	185	SER	5.6
1	A	184	GLU	5.5
1	E	184	GLU	5.5

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.