



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 09:58 PM GMT

PDB ID : 5A23  
Title : SdsA sulfatase triclinic form  
Authors : De la Mora, E.; Flores-Hernandez, E.; Jakoncik, J.; Stojanoff, V.; Sanchez-Puig, N.; Moreno, A.  
Deposited on : 2015-05-11  
Resolution : 2.41 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](mailto: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 : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026982  
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) : rb-20026982

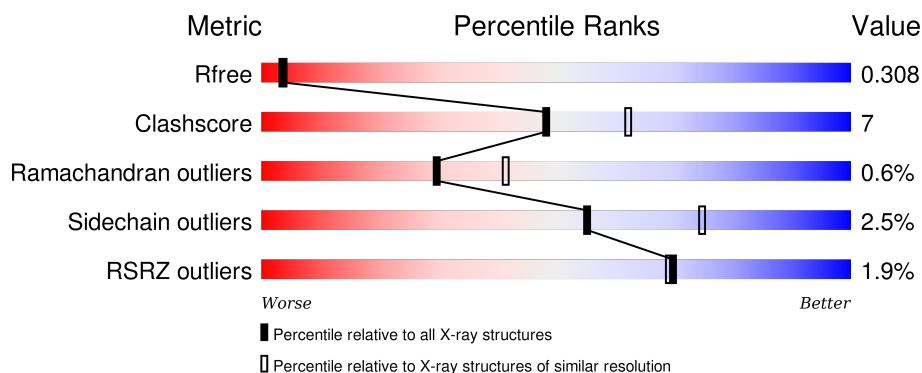
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.41 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	3386 (2.44-2.40)
Clashscore	102246	3897 (2.44-2.40)
Ramachandran outliers	100387	3837 (2.44-2.40)
Sidechain outliers	100360	3838 (2.44-2.40)
RSRZ outliers	91569	3396 (2.44-2.40)

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  $\geq 3$ , 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	658	<div> <div>2%</div> <div>79%</div> <div>15%</div> <div>• 5%</div> </div>
1	B	658	<div> <div>3%</div> <div>77%</div> <div>17%</div> <div>• 5%</div> </div>
1	C	658	<div> <div>2%</div> <div>75%</div> <div>19%</div> <div>• 5%</div> </div>
1	D	658	<div> <div>%</div> <div>78%</div> <div>16%</div> <div>• 5%</div> </div>

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
2	ZN	A	1001	-	-	-	X

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 19632 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.

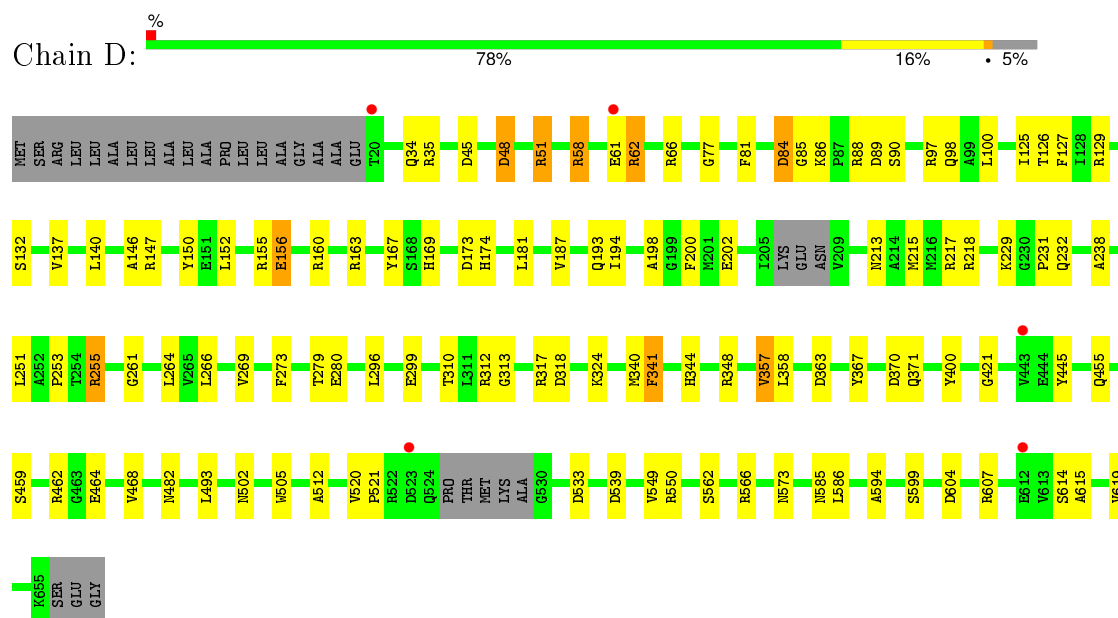
- Molecule 1 is a protein called SDS HYDROLASE SDSA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	628	Total	C	N	O	S	0	0	1
			4906	3096	887	913	10			
1	B	628	Total	C	N	O	S	0	0	1
			4906	3096	887	913	10			
1	C	628	Total	C	N	O	S	0	0	1
			4906	3096	887	913	10			
1	D	628	Total	C	N	O	S	0	0	1
			4906	3096	887	913	10			

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	2	Total	Zn	0	0
			2	2		
2	A	2	Total	Zn	0	0
			2	2		
2	D	2	Total	Zn	0	0
			2	2		
2	C	2	Total	Zn	0	0
			2	2		





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	86.90 Å   86.80 Å   129.03 Å 107.59°   102.96°   95.59°	Depositor
Resolution (Å)	118.77 – 2.41 19.73 – 2.41	Depositor EDS
% Data completeness (in resolution range)	71.7 (118.77-2.41) 57.3 (19.73-2.41)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.27 (at 2.41 Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, $R_{free}$	0.253   ,   0.307 0.258   ,   0.308	Depositor DCC
$R_{free}$ test set	4958 reflections (5.47%)	DCC
Wilson B-factor (Å <sup>2</sup> )	41.1	Xtriage
Anisotropy	0.154	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 20.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.39$ , $\langle L^2 \rangle = 0.22$	Xtriage
Outliers	0 of 95605 reflections	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	19632	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.32% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical 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: ZN

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	0.97	1/5012 (0.0%)	1.07	21/6805 (0.3%)
1	B	0.82	2/5012 (0.0%)	0.95	4/6805 (0.1%)
1	C	0.85	0/5012	1.00	18/6805 (0.3%)
1	D	0.92	0/5012	1.02	18/6805 (0.3%)
All	All	0.89	3/20048 (0.0%)	1.01	61/27220 (0.2%)

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	1
1	B	0	1
1	C	0	1
All	All	0	3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	299	GLU	CD-OE2	5.89	1.32	1.25
1	B	299	GLU	CD-OE2	5.68	1.31	1.25
1	B	338	GLU	CD-OE1	5.33	1.31	1.25

The worst 5 of 61 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	59	ARG	NE-CZ-NH1	10.62	125.61	120.30
1	A	370	ASP	CB-CG-OD2	9.75	127.07	118.30
1	A	59	ARG	NE-CZ-NH2	-9.45	115.58	120.30

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	370	ASP	CB-CG-OD1	-9.26	109.97	118.30
1	C	217	ARG	NE-CZ-NH1	8.78	124.69	120.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	40	GLU	Peptide
1	B	523	ASP	Peptide
1	C	134	TRP	Peptide

## 5.2 Too-close contacts

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	4906	0	4818	52	0
1	B	4906	0	4818	100	0
1	C	4906	0	4818	87	0
1	D	4906	0	4818	66	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	1	0
All	All	19632	0	19272	280	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 7.

The worst 5 of 280 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:D:299:GLU:OE2	2:D:1001:ZN:ZN	1.17	0.91
1:C:35:ARG:C	1:C:38:GLU:HG2	2.00	0.82
1:C:35:ARG:NH2	1:C:89:ASP:OD1	2.12	0.81
1:C:35:ARG:O	1:C:38:GLU:HG2	1.82	0.78
1:B:146:ALA:HB1	1:B:181:LEU:HD21	1.65	0.77

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	622/658 (94%)	581 (93%)	37 (6%)	4 (1%)	30	41
1	B	622/658 (94%)	586 (94%)	33 (5%)	3 (0%)	34	47
1	C	622/658 (94%)	571 (92%)	49 (8%)	2 (0%)	46	62
1	D	622/658 (94%)	575 (92%)	41 (7%)	6 (1%)	19	27
All	All	2488/2632 (94%)	2313 (93%)	160 (6%)	15 (1%)	30	41

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	140	LEU
1	B	81	PHE
1	B	138	ASP
1	B	173	ASP
1	C	38	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	499/521 (96%)	486 (97%)	13 (3%)	54	74
1	B	499/521 (96%)	487 (98%)	12 (2%)	57	76
1	C	499/521 (96%)	486 (97%)	13 (3%)	54	74

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	499/521 (96%)	488 (98%)	11 (2%)	60	78
All	All	1996/2084 (96%)	1947 (98%)	49 (2%)	55	75

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

Mol	Chain	Res	Type
1	B	550	ARG
1	C	62	ARG
1	D	520	VAL
1	B	614	SER
1	C	71	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	102	ASN
1	C	185	GLN
1	C	345	ASN
1	B	524	GLN
1	C	307	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](#)

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	628/658 (95%)	-0.33	13 (2%) 67 66	13, 38, 73, 104	0
1	B	628/658 (95%)	-0.08	19 (3%) 54 52	23, 50, 84, 118	0
1	C	628/658 (95%)	-0.25	10 (1%) 74 74	20, 45, 75, 123	0
1	D	628/658 (95%)	-0.30	5 (0%) 87 87	18, 42, 73, 107	0
All	All	2512/2632 (95%)	-0.24	47 (1%) 70 69	13, 44, 79, 123	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	522	ARG	7.6
1	B	492	ALA	5.3
1	C	524	GLN	4.3
1	B	530	GLY	4.2
1	B	484	ALA	3.9

### 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
2	ZN	A	1001	1/1	0.99	0.15	2.74	39,39,39,39	0
2	ZN	C	1002	1/1	0.96	0.16	1.87	50,50,50,50	0
2	ZN	D	1001	1/1	0.98	0.14	0.80	27,27,27,27	1
2	ZN	C	1001	1/1	0.98	0.12	0.53	52,52,52,52	0
2	ZN	A	1002	1/1	1.00	0.12	0.49	44,44,44,44	0
2	ZN	D	1002	1/1	0.98	0.12	-0.72	48,48,48,48	1
2	ZN	B	1002	1/1	0.97	0.10	-0.75	52,52,52,52	0
2	ZN	B	1001	1/1	0.99	0.10	-0.90	57,57,57,57	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.