



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 14, 2017 – 01:08 am GMT

PDB ID : 1GM8  
Title : Crystal structures of penicillin acylase enzyme-substrate complexes: Structural insights into the catalytic mechanism  
Authors : McVey, C.E.; Walsh, M.A.; Dodson, G.G.; Wilson, K.S.; Brannigan, J.A.  
Deposited on : 2001-09-11  
Resolution : 2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : recalc28949

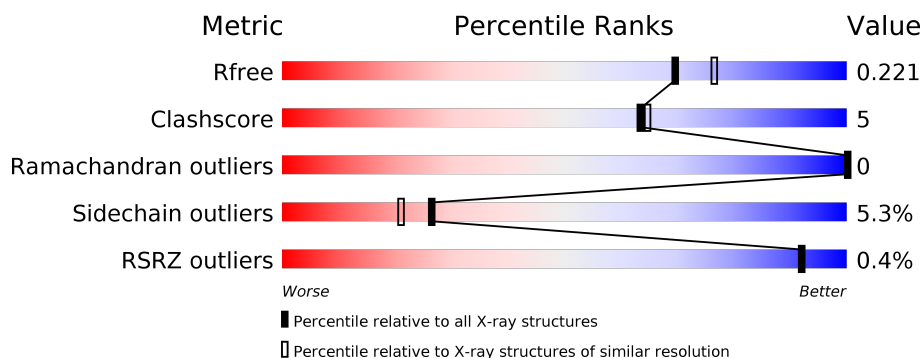
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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))
$R_{free}$	100719	6609 (2.00-2.00)
Clashscore	112137	7775 (2.00-2.00)
Ramachandran outliers	110173	7679 (2.00-2.00)
Sidechain outliers	110143	7678 (2.00-2.00)
RSRZ outliers	101464	6696 (2.00-2.00)

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	209	
2	B	557	

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
3	CA	B	1558	-	-	-	X
4	SOX	B	1559	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6754 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 PENICILLIN G ACYLASE ALPHA SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	207	Total	C	N	O	S	0	0	0
			1662	1061	279	314	8			

- Molecule 2 is a protein called PENICILLIN G ACYLASE BETA SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	557	Total	C	N	O	S	0	1	0
			4423	2810	769	834	10			

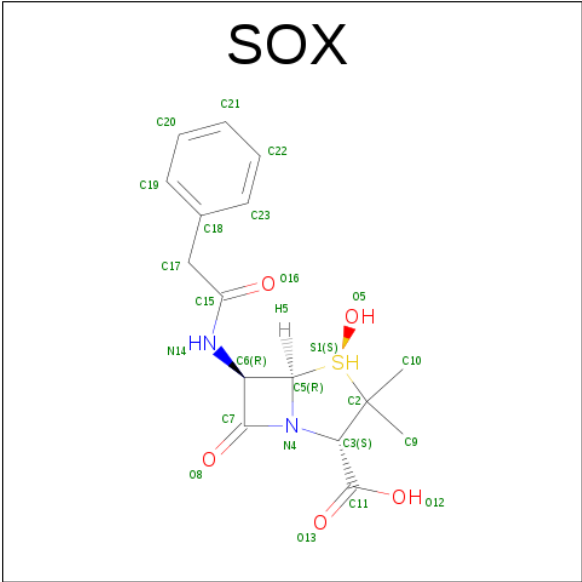
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	241	ALA	ASN	ENGINEERED MUTATION	UNP P06875

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Ca	0	0
			1	1		

- Molecule 4 is N-[(2S,4S,6R)-2-(DIHYDROXYMETHYL)-4-HYDROXY-3,3-DIMETHYL-7-OXO-4LAMBDA 4 -THIA-1-AZABICYCLO[3.2.0]HEPT-6-YL]-2-PHENYLACETAMIDE (three-letter code: SOX) (formula: C<sub>16</sub>H<sub>20</sub>N<sub>2</sub>O<sub>5</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	B	1	Total	C	N	O	S	0	0
			24	16	2	5	1		

- Molecule 5 is water.

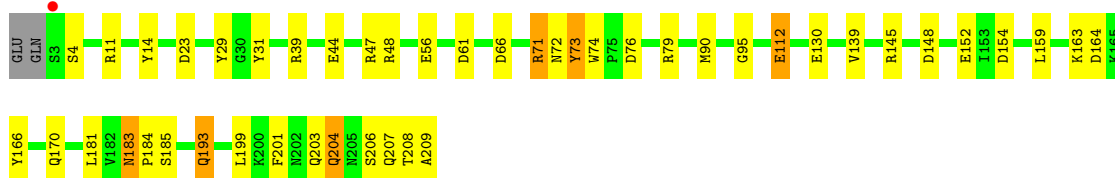
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	182	Total	O	3	3
			182	182		
5	B	462	Total	O	4	4
			462	462		

### 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.

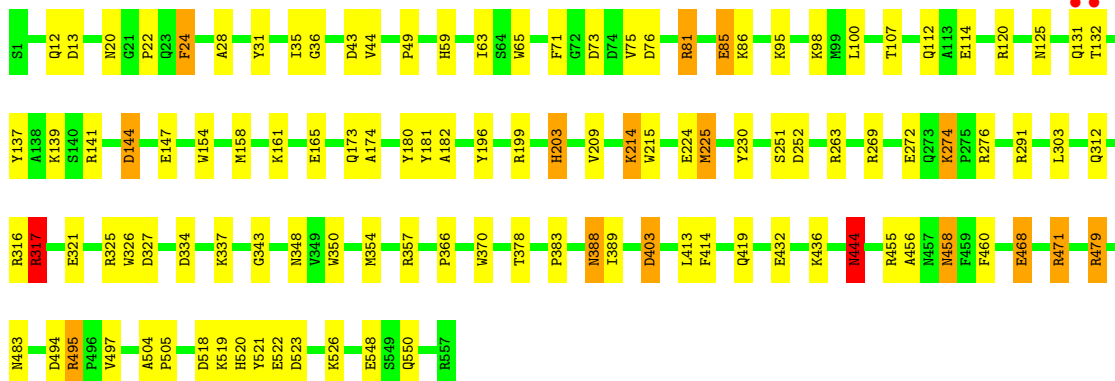
#### • Molecule 1: PENICILLIN G ACYLASE ALPHA SUBUNIT

Chain A: 



#### • Molecule 2: PENICILLIN G ACYLASE BETA SUBUNIT

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	51.20Å 131.20Å 63.90Å 90.00° 105.80° 90.00°	Depositor
Resolution (Å)	15.00 – 2.00 14.97 – 2.00	Depositor EDS
% Data completeness (in resolution range)	93.4 (15.00-2.00) 93.3 (14.97-2.00)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.63 (at 2.00Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.162 , 0.234 0.154 , 0.221	Depositor DCC
$R_{free}$ test set	1540 reflections (3.12%)	DCC
Wilson B-factor (Å <sup>2</sup> )	13.6	Xtriage
Anisotropy	0.425	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 68.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6754	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.94% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SOX, CA

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	0.88	0/1704	1.73	27/2312 (1.2%)
2	B	0.87	1/4549 (0.0%)	1.72	57/6203 (0.9%)
All	All	0.87	1/6253 (0.0%)	1.72	84/8515 (1.0%)

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	A	0	3
2	B	0	5
All	All	0	8

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	165	GLU	CD-OE2	5.57	1.31	1.25

All (84) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	471	ARG	NE-CZ-NH2	-27.30	106.65	120.30
2	B	81	ARG	CD-NE-CZ	20.99	152.99	123.60
2	B	471	ARG	CD-NE-CZ	20.12	151.76	123.60
1	A	145	ARG	NE-CZ-NH2	-15.78	112.41	120.30
2	B	76	ASP	CB-CG-OD2	13.01	130.00	118.30
1	A	66	ASP	CB-CG-OD2	-11.87	107.62	118.30
2	B	73	ASP	CB-CG-OD2	11.72	128.85	118.30
2	B	471	ARG	NH1-CZ-NH2	11.58	132.14	119.40

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	81	ARG	NE-CZ-NH1	11.27	125.94	120.30
1	A	66	ASP	CB-CG-OD1	11.24	128.42	118.30
2	B	479	ARG	NE-CZ-NH1	-11.10	114.75	120.30
1	A	23	ASP	CB-CG-OD1	11.01	128.21	118.30
2	B	230	TYR	CB-CG-CD2	10.64	127.38	121.00
2	B	317	ARG	NE-CZ-NH1	10.45	125.52	120.30
1	A	71	ARG	NE-CZ-NH2	-9.60	115.50	120.30
2	B	263	ARG	NE-CZ-NH2	-9.49	115.56	120.30
2	B	325	ARG	NE-CZ-NH2	-9.34	115.63	120.30
2	B	269	ARG	NE-CZ-NH2	-9.06	115.77	120.30
2	B	357	ARG	NE-CZ-NH2	8.91	124.76	120.30
1	A	14	TYR	CB-CG-CD2	-8.88	115.67	121.00
2	B	495[A]	ARG	NE-CZ-NH2	8.77	124.68	120.30
2	B	495[B]	ARG	NE-CZ-NH2	8.77	124.68	120.30
1	A	39	ARG	NE-CZ-NH1	8.66	124.63	120.30
2	B	141	ARG	NE-CZ-NH2	-8.66	115.97	120.30
1	A	73	TYR	CB-CG-CD2	-8.57	115.86	121.00
2	B	230	TYR	CB-CG-CD1	-8.14	116.12	121.00
1	A	73	TYR	CB-CG-CD1	8.03	125.82	121.00
1	A	145	ARG	NH1-CZ-NH2	8.02	128.23	119.40
1	A	130	GLU	OE1-CD-OE2	-7.61	114.16	123.30
2	B	76	ASP	OD1-CG-OD2	-7.56	108.93	123.30
1	A	152	GLU	OE1-CD-OE2	-7.45	114.36	123.30
2	B	199	ARG	NE-CZ-NH1	7.44	124.02	120.30
2	B	468	GLU	OE1-CD-OE2	7.42	132.20	123.30
1	A	14	TYR	CB-CG-CD1	7.33	125.40	121.00
2	B	252	ASP	CB-CG-OD1	7.31	124.88	118.30
1	A	47	ARG	NE-CZ-NH2	-7.13	116.74	120.30
1	A	90	MET	CG-SD-CE	7.12	111.59	100.20
1	A	148	ASP	CB-CG-OD2	7.09	124.69	118.30
1	A	145	ARG	CD-NE-CZ	6.93	133.31	123.60
1	A	47	ARG	CD-NE-CZ	6.66	132.92	123.60
1	A	76	ASP	CB-CG-OD1	6.60	124.24	118.30
1	A	11	ARG	NE-CZ-NH2	-6.57	117.01	120.30
2	B	455	ARG	NE-CZ-NH1	6.57	123.59	120.30
2	B	403	ASP	CB-CG-OD1	-6.57	112.39	118.30
2	B	43	ASP	CB-CG-OD2	-6.52	112.43	118.30
2	B	321	GLU	OE1-CD-OE2	6.34	130.91	123.30
2	B	521	TYR	CD1-CE1-CZ	-6.33	114.11	119.80
2	B	327	ASP	CB-CG-OD1	6.29	123.96	118.30
2	B	182	ALA	N-CA-CB	6.25	118.85	110.10
1	A	61	ASP	CB-CG-OD2	-6.19	112.73	118.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	276	ARG	NE-CZ-NH1	6.10	123.35	120.30
2	B	357	ARG	NE-CZ-NH1	-6.05	117.27	120.30
2	B	225	MET	CA-CB-CG	-5.99	103.11	113.30
2	B	24	PHE	CB-CG-CD1	-5.99	116.61	120.80
2	B	497	VAL	CA-CB-CG2	5.99	119.88	110.90
2	B	75	VAL	CA-C-O	-5.93	107.65	120.10
1	A	48	ARG	NE-CZ-NH2	-5.92	117.34	120.30
2	B	161	LYS	CA-CB-CG	5.85	126.28	113.40
2	B	161	LYS	CB-CG-CD	5.84	126.78	111.60
2	B	174	ALA	N-CA-CB	-5.83	101.94	110.10
2	B	75	VAL	CA-C-N	5.78	129.91	117.20
2	B	252	ASP	OD1-CG-OD2	-5.78	112.33	123.30
2	B	165	GLU	OE1-CD-OE2	-5.66	116.50	123.30
2	B	316	ARG	NE-CZ-NH2	5.62	123.11	120.30
2	B	144	ASP	CB-CG-OD1	5.61	123.35	118.30
1	A	184	PRO	N-CA-CB	5.60	110.02	103.30
1	A	79	ARG	NE-CZ-NH2	-5.53	117.53	120.30
2	B	334	ASP	CB-CG-OD1	5.52	123.27	118.30
2	B	28	ALA	CB-CA-C	-5.52	101.82	110.10
2	B	366	PRO	CB-CA-C	-5.51	98.22	112.00
2	B	181	TYR	N-CA-CB	-5.49	100.71	110.60
2	B	137	TYR	CB-CG-CD2	-5.46	117.72	121.00
2	B	85	GLU	OE1-CD-OE2	-5.42	116.79	123.30
1	A	209	ALA	N-CA-CB	5.42	117.68	110.10
1	A	148	ASP	OD1-CG-OD2	-5.40	113.05	123.30
1	A	204	GLN	CA-CB-CG	5.34	125.14	113.40
2	B	251	SER	N-CA-CB	-5.29	102.56	110.50
2	B	456	ALA	O-C-N	-5.24	114.32	122.70
2	B	518	ASP	CB-CG-OD2	-5.24	113.59	118.30
2	B	73	ASP	CB-CG-OD1	-5.15	113.66	118.30
2	B	196	TYR	CA-CB-CG	-5.13	103.64	113.40
2	B	479	ARG	NE-CZ-NH2	5.09	122.84	120.30
2	B	291	ARG	NE-CZ-NH1	5.04	122.82	120.30
2	B	272	GLU	CA-CB-CG	-5.02	102.35	113.40

There are no chirality outliers.

All (8) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	203	GLN	Mainchain
1	A	206	SER	Mainchain
1	A	44	GLU	Mainchain

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
2	B	144	ASP	Mainchain
2	B	403	ASP	Mainchain
2	B	444	ASN	Mainchain
2	B	495[B]	ARG	Mainchain
2	B	85	GLU	Mainchain

## 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	1662	0	1609	20	1
2	B	4423	0	4254	47	1
3	B	1	0	0	0	0
4	B	24	0	20	1	0
5	A	182	0	0	3	0
5	B	462	0	0	3	0
All	All	6754	0	5883	59	1

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 5.

All (59) 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:214:LYS:H	2:B:214:LYS:HE2	1.18	1.06
2:B:214:LYS:H	2:B:214:LYS:CE	1.91	0.83
1:A:181:LEU:HD21	1:A:201:PHE:HB2	1.63	0.81
2:B:479:ARG:HH21	2:B:483:ASN:HD22	1.31	0.79
2:B:317:ARG:HH11	2:B:317:ARG:HG3	1.48	0.77
2:B:214:LYS:HE2	2:B:214:LYS:N	1.97	0.76
1:A:199:LEU:HD11	2:B:225:MET:HE3	1.68	0.76
1:A:164:ASP:HB3	5:A:2137:HOH:O	1.88	0.72
2:B:388:ASN:HD22	2:B:389:ILE:H	1.38	0.71
2:B:520:HIS:HE1	2:B:548:GLU:OE2	1.74	0.70
2:B:120:ARG:HH11	2:B:125:ASN:ND2	1.90	0.69
1:A:71:ARG:NH2	5:A:2055:HOH:O	2.27	0.64

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:203:HIS:HD2	5:B:2173:HOH:O	1.81	0.62
2:B:520:HIS:HD2	2:B:523:ASP:OD2	1.82	0.61
2:B:444:ASN:C	2:B:444:ASN:HD22	2.04	0.60
1:A:56:GLU:HG2	2:B:107:THR:HB	1.85	0.59
2:B:44:VAL:HG11	2:B:158:MET:HB3	1.83	0.59
2:B:471:ARG:NH2	2:B:522:GLU:OE1	2.31	0.58
1:A:199:LEU:HD21	2:B:225:MET:HE1	1.87	0.57
2:B:458:ASN:HD21	2:B:460:PHE:HB2	1.71	0.55
2:B:458:ASN:C	2:B:458:ASN:HD22	2.09	0.55
1:A:183:ASN:HD22	1:A:185:SER:H	1.54	0.55
2:B:388:ASN:ND2	2:B:389:ILE:H	2.03	0.54
5:A:2129:HOH:O	2:B:203:HIS:HE1	1.90	0.54
2:B:214:LYS:HG2	2:B:215:TRP:CD1	2.43	0.53
1:A:72:ASN:HD21	2:B:139:LYS:NZ	2.06	0.53
2:B:378:THR:HG22	2:B:383:PRO:HG3	1.91	0.52
1:A:183:ASN:ND2	1:A:185:SER:H	2.08	0.51
1:A:183:ASN:HD22	1:A:183:ASN:C	2.15	0.51
2:B:504:ALA:HA	2:B:505:PRO:C	2.31	0.50
1:A:139:VAL:HG22	2:B:147:GLU:HB3	1.94	0.49
2:B:479:ARG:HH21	2:B:483:ASN:ND2	2.06	0.49
2:B:468:GLU:H	2:B:468:GLU:CD	2.16	0.49
1:A:181:LEU:HD21	1:A:201:PHE:CB	2.38	0.47
2:B:388:ASN:HD22	2:B:389:ILE:N	2.08	0.47
1:A:193:GLN:NE2	1:A:193:GLN:H	2.12	0.47
1:A:112:GLU:CD	1:A:112:GLU:H	2.17	0.46
1:A:166:TYR:HB3	1:A:170:GLN:HG2	1.97	0.46
1:A:29:TYR:HA	1:A:95:GLY:O	2.15	0.45
2:B:35:ILE:HG13	2:B:36:GLY:N	2.31	0.45
2:B:274:LYS:HE2	5:B:2226:HOH:O	2.16	0.45
2:B:12:GLN:O	2:B:13:ASP:HB2	2.17	0.45
2:B:71:PHE:CZ	4:B:1559:SOX:H5	2.52	0.44
2:B:378:THR:CG2	2:B:383:PRO:HG3	2.48	0.44
1:A:159:LEU:HD11	1:A:163:LYS:HE3	2.01	0.43
2:B:22:PRO:HB2	2:B:24:PHE:CE2	2.53	0.42
2:B:112:GLN:NE2	5:B:2095:HOH:O	2.46	0.42
2:B:59:HIS:HB2	2:B:63:ILE:O	2.20	0.41
2:B:303:LEU:HD21	2:B:350:TRP:CE2	2.55	0.41
2:B:414:PHE:CD1	2:B:419:GLN:HG2	2.55	0.41
2:B:350:TRP:CZ2	2:B:354:MET:HG3	2.56	0.41
2:B:526:LYS:HD3	2:B:526:LYS:HA	1.86	0.41
1:A:154:ASP:HB3	2:B:370:TRP:CH2	2.56	0.41

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:207:GLN:HA	2:B:215:TRP:CZ2	2.56	0.41
2:B:326:TRP:CZ3	2:B:343:GLY:HA3	2.56	0.41
2:B:414:PHE:CG	2:B:419:GLN:HG2	2.56	0.41
1:A:74:TRP:CZ2	2:B:100:LEU:HG	2.57	0.40
2:B:31:TYR:CE2	2:B:49:PRO:HB3	2.56	0.40
2:B:65:TRP:HA	2:B:180:TYR:O	2.22	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:208:THR:OG1	2:B:494:ASP:OD1[2_645]	2.15	0.05

## 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	205/209 (98%)	199 (97%)	6 (3%)	0	100	100
2	B	556/557 (100%)	545 (98%)	11 (2%)	0	100	100
All	All	761/766 (99%)	744 (98%)	17 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	178/180 (99%)	171 (96%)	7 (4%)	37	34
2	B	460/459 (100%)	433 (94%)	27 (6%)	23	17
All	All	638/639 (100%)	604 (95%)	34 (5%)	26	21

All (34) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	4	SER
1	A	31	TYR
1	A	73	TYR
1	A	112	GLU
1	A	183	ASN
1	A	193	GLN
1	A	204	GLN
2	B	20	ASN
2	B	81	ARG
2	B	86	LYS
2	B	95	LYS
2	B	98	LYS
2	B	114	GLU
2	B	131	GLN
2	B	132	THR
2	B	154	TRP
2	B	173	GLN
2	B	203	HIS
2	B	209	VAL
2	B	214	LYS
2	B	224	GLU
2	B	274	LYS
2	B	312	GLN
2	B	317	ARG
2	B	337	LYS
2	B	348	ASN
2	B	388	ASN
2	B	413	LEU
2	B	432	GLU
2	B	436	LYS
2	B	444	ASN
2	B	458	ASN
2	B	519	LYS
2	B	550	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (20) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	22	ASN
1	A	72	ASN
1	A	120	ASN
1	A	183	ASN
1	A	193	GLN
1	A	204	GLN
2	B	93	ASN
2	B	110	ASN
2	B	112	GLN
2	B	125	ASN
2	B	203	HIS
2	B	239	ASN
2	B	245	GLN
2	B	348	ASN
2	B	388	ASN
2	B	444	ASN
2	B	458	ASN
2	B	473	GLN
2	B	483	ASN
2	B	520	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 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	$\# Z  > 2$
4	SOX	B	1559	-	19,26,26	1.39	1 (5%)	26,40,40	1.35	2 (7%)

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
4	SOX	B	1559	-	-	0/8/48/48	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	1559	SOX	C3-N4	-5.15	1.43	1.47

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	1559	SOX	O16-C15-C17	-2.24	116.32	122.00
4	B	1559	SOX	O16-C15-N14	4.12	130.10	122.97

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
4	B	1559	SOX	1	0

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

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	207/209 (99%)	-0.87	1 (0%) 90 90	9, 16, 27, 46	0
2	B	557/557 (100%)	-0.87	2 (0%) 92 92	7, 15, 32, 72	0
All	All	764/766 (99%)	-0.87	3 (0%) 92 92	7, 15, 31, 72	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	3	SER	3.0
2	B	131	GLN	2.9
2	B	132	THR	2.2

### 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
4	SOX	B	1559	24/24	0.89	0.16	4.75	17,26,35,38	0
3	CA	B	1558	1/1	1.00	0.08	3.64	14,14,14,14	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.