



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 31, 2020 – 06:33 AM BST

PDB ID : 6FJJ  
Title : Joint neutron and x-ray crystal structure of human carbonic anhydrase IX mimic (saccharin).  
Authors : Fisher, S.Z.  
Deposited on : 2018-01-22  
Resolution : 1.50 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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : **FAILED**  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.13

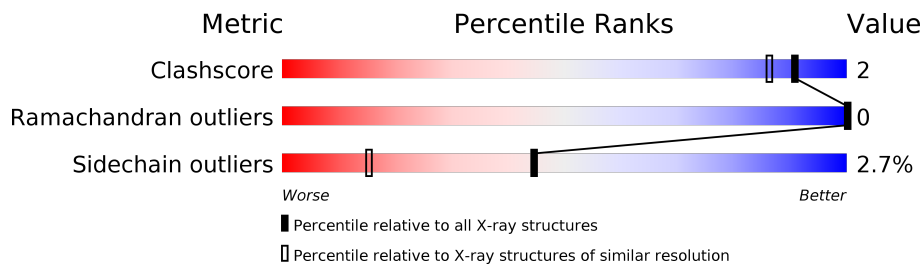
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION, NEUTRON DIFFRACTION*

The reported resolution of this entry is 1.50 Å.

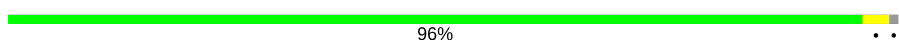
Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)

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 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  $\leq 5\%$ .

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	260	 96% ..

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4970 atoms, of which 1907 are hydrogens and 776 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 Carbonic anhydrase 2.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
			Total	C	D	H	N	O	S			
1	A	258	4373	1332	390	1903	360	386	2	94	218	0

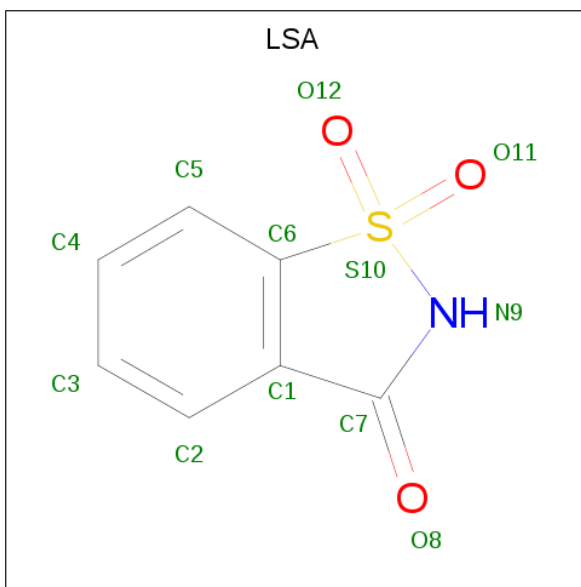
There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	65	SER	ALA	engineered mutation	UNP P00918
A	67	GLN	ASN	engineered mutation	UNP P00918
A	69	THR	GLU	engineered mutation	UNP P00918
A	91	LEU	ILE	engineered mutation	UNP P00918
A	130	VAL	PHE	engineered mutation	UNP P00918
A	169	GLU	LYS	engineered mutation	UNP P00918
A	203	ALA	LEU	engineered mutation	UNP P00918

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Zn	0	0
			1	1		

- Molecule 3 is 1,2-BENZISOTHIAZOL-3(2H)-ONE 1,1-DIOXIDE (three-letter code: LSA) (formula: C<sub>7</sub>H<sub>5</sub>NO<sub>3</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
			Total	C	H	N	O			S
3	A	1	16	7	4	1	3	1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	D	O		
4	A	194	580	386	194	0	0

### 3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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.

Note EDS failed to run properly.

- Molecule 1: Carbonic anhydrase 2

Chain A:  96%



## 4 Data and refinement statistics i

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	42.53Å 41.83Å 72.70Å 90.00° 104.04° 90.00°	Depositor
Resolution (Å)	25.59 – 1.50	Depositor
% Data completeness (in resolution range)	96.5 (25.59-1.50)	Depositor
$R_{merge}$	0.22	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.56 (at 1.20Å)	Xtrriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.135 , 0.166	Depositor
Wilson B-factor (Å <sup>2</sup> )	9.7	Xtrriage
Anisotropy	0.220	Xtrriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.026 for h,-k,-h-l	Xtrriage
Total number of atoms	4970	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.51% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, DOD, LSA

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	0.85	4/3959 (0.1%)	0.88	0/5366

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	224[A]	LYS	C-N	7.82	1.52	1.34
1	A	224[B]	LYS	C-N	7.82	1.52	1.34
1	A	224[C]	LYS	C-N	7.82	1.52	1.34
1	A	224[D]	LYS	C-N	7.82	1.52	1.34

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	2470	1903	383	4	0
2	A	1	0	0	0	0
3	A	12	4	4	0	0
4	A	580	0	0	4	0
All	All	3063	1907	387	6	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 2.

All (6) 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:254[B]:GLN:NE2	4:A:409:DOD:O	2.49	0.45
1:A:64[B]:HIS:CE1	4:A:428:DOD:O	2.70	0.44

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	480/260 (185%)	467 (97%)	13 (3%)	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	414/225 (184%)	400 (97%)	14 (3%)	37 9

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	60[A]	LEU

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	60[B]	LEU
1	A	74[A]	GLN
1	A	74[B]	GLN
1	A	120[A]	LEU
1	A	120[B]	LEU
1	A	224[A]	LYS
1	A	224[B]	LYS
1	A	224[C]	LYS
1	A	224[D]	LYS
1	A	227[A]	LYS
1	A	227[B]	LYS
1	A	260[A]	LYS
1	A	260[B]	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

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
3	LSA	A	302	2	13,13,13	2.22	3 (23%)	20,20,20	2.97	11 (55%)

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
3	LSA	A	302	2	-	-	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	302	LSA	C1-C7	5.26	1.56	1.48
3	A	302	LSA	C7-N9	3.77	1.44	1.37
3	A	302	LSA	O11-S10	3.21	1.47	1.43

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	302	LSA	C5-C6-C1	-6.64	117.15	122.73
3	A	302	LSA	C6-C1-C7	-5.11	107.22	112.08
3	A	302	LSA	O12-S10-N9	-4.74	106.93	110.39
3	A	302	LSA	O11-S10-N9	4.25	113.48	110.39
3	A	302	LSA	C2-C1-C6	3.69	123.74	119.91
3	A	302	LSA	C7-N9-S10	-3.67	108.46	114.81
3	A	302	LSA	C6-S10-N9	2.91	98.14	93.26
3	A	302	LSA	C3-C4-C5	2.44	123.90	120.19
3	A	302	LSA	C1-C6-S10	2.30	112.44	109.63
3	A	302	LSA	O8-C7-C1	-2.28	121.69	127.67
3	A	302	LSA	C5-C6-S10	2.24	130.40	127.09

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

There are no chain breaks in this entry.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

EDS failed to run properly - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

EDS failed to run properly - this section is therefore empty.

### 6.3 Carbohydrates

EDS failed to run properly - this section is therefore empty.

### 6.4 Ligands

EDS failed to run properly - this section is therefore empty.

### 6.5 Other polymers

EDS failed to run properly - this section is therefore empty.