Full wwPDB X-ray Structure Validation Report (i)

Dec 16, 2023 – 08:24 am GMT

PDB ID	:	4CNX
Title	:	Surface residue engineering of bovine carbonic anhydrase to an extreme halophilic enzyme for potential application in postcombustion CO2 capture
Authors	:	Warden, A.; Newman, J.; Peat, T.S.; Seabrook, S.; Williams, M.; Dojchinov, G.; Haritos, V.
Deposited on	:	2014-01-25
Resolution	:	1.23 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

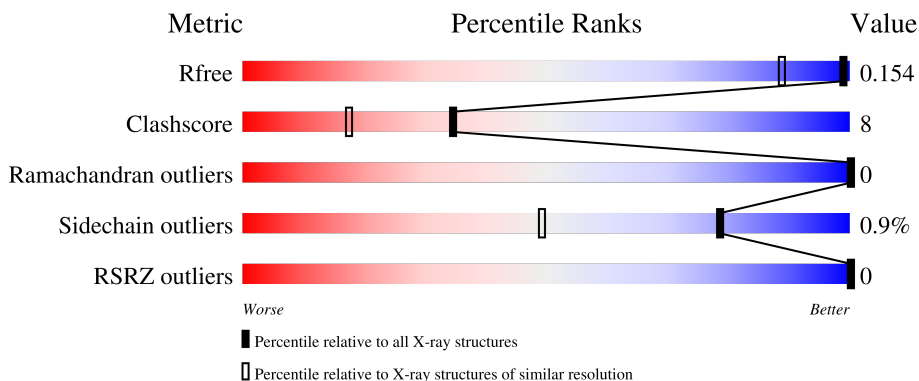|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.23 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		0.154
Clashscore		8
Ramachandran outliers		0
Sidechain outliers		0.9%
RSRZ outliers		0

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	2024 (1.28-1.20)
Clashscore	141614	1007 (1.26-1.22)
Ramachandran outliers	138981	2053 (1.28-1.20)
Sidechain outliers	138945	2051 (1.28-1.20)
RSRZ outliers	127900	1987 (1.28-1.20)

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 of 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	262	83% 13% · ·

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 2641 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 CARBONIC ANHYDRASE 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	257	2228	1405	373	447	3	0	22	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	expression tag	UNP P00921
A	0	GLY	-	expression tag	UNP P00921
A	8	ASP	GLY	engineered mutation	UNP P00921
A	18	GLU	LYS	engineered mutation	UNP P00921
A	24	ASP	ASN	engineered mutation	UNP P00921
A	36	ASP	LYS	engineered mutation	UNP P00921
A	39	ASP	VAL	engineered mutation	UNP P00921
A	50	ASP	VAL	engineered mutation	UNP P00921
A	57	ASP	ARG	engineered mutation	UNP P00921
A	62	ASP	ASN	engineered mutation	UNP P00921
A	74	GLU	GLN	engineered mutation	UNP P00921
A	85	ASP	THR	engineered mutation	UNP P00921
A	136	GLU	GLN	engineered mutation	UNP P00921
A	169	GLU	LYS	engineered mutation	UNP P00921
A	177	ASP	ASN	engineered mutation	UNP P00921
A	186	GLU	ASN	engineered mutation	UNP P00921
A	220	GLU	GLN	engineered mutation	UNP P00921
A	238	GLU	LEU	engineered mutation	UNP P00921
A	252	ASP	ASN	engineered mutation	UNP P00921
A	254	GLU	GLN	engineered mutation	UNP P00921

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

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

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			7	4	3		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	405	Total	O	0	0
			405	405		

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 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: CARBONIC ANHYDRASE 2

Chain A:
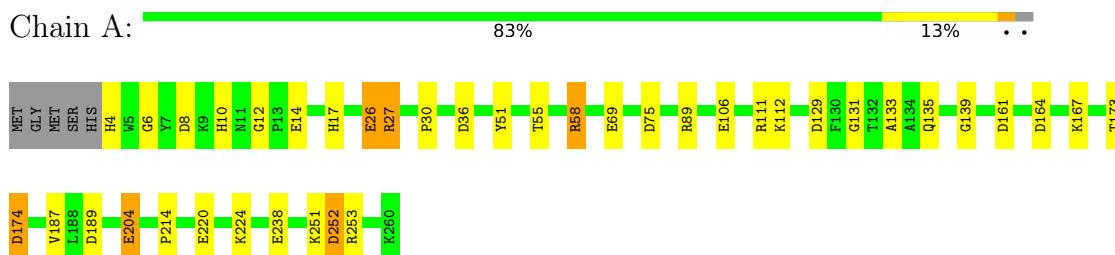
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	41.59Å　67.94Å　43.13Å 90.00°　103.07°　90.00°	Depositor
Resolution (Å)	35.76　–　1.23 35.73　–　1.23	Depositor EDS
% Data completeness (in resolution range)	98.6 (35.76-1.23) 98.6 (35.73-1.23)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.03 (at 1.24Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.111　,　0.147 0.121　,　0.154	Depositor DCC
R_{free} test set	3344 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	8.7	Xtriage
Anisotropy	0.073	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 43.0	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.034 for l,-k,h	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	2641	wwPDB-VP
Average B, all atoms (Å^2)	12.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.42% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[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)

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

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	1.15	14/2292 (0.6%)	1.14	18/3120 (0.6%)

All (14) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	26[A]	GLU	CG-CD	8.92	1.65	1.51
1	A	26[B]	GLU	CG-CD	8.92	1.65	1.51
1	A	26[A]	GLU	CD-OE1	8.92	1.35	1.25
1	A	26[B]	GLU	CD-OE1	8.92	1.35	1.25
1	A	204	GLU	CD-OE1	7.13	1.33	1.25
1	A	129	ASP	CB-CG	-6.37	1.38	1.51
1	A	27[A]	ARG	NE-CZ	-6.26	1.25	1.33
1	A	27[B]	ARG	NE-CZ	-6.26	1.25	1.33
1	A	220	GLU	CD-OE1	-6.09	1.19	1.25
1	A	27[A]	ARG	CD-NE	5.96	1.56	1.46
1	A	27[B]	ARG	CD-NE	5.96	1.56	1.46
1	A	238[A]	GLU	CD-OE1	-5.55	1.19	1.25
1	A	238[B]	GLU	CD-OE1	-5.55	1.19	1.25
1	A	112	LYS	CD-CE	5.38	1.64	1.51

All (18) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	129	ASP	CB-CG-OD1	12.18	129.27	118.30
1	A	89	ARG	NE-CZ-NH1	-10.01	115.30	120.30
1	A	129	ASP	CB-CG-OD2	-9.20	110.02	118.30
1	A	51	TYR	CB-CG-CD1	6.88	125.12	121.00
1	A	252[A]	ASP	CB-CG-OD2	-6.85	112.14	118.30
1	A	252[B]	ASP	CB-CG-OD2	-6.85	112.14	118.30
1	A	112	LYS	CD-CE-NZ	-6.36	97.08	111.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	189	ASP	CB-CG-OD2	-6.24	112.69	118.30
1	A	75	ASP	CB-CG-OD2	-6.24	112.69	118.30
1	A	58[A]	ARG	NE-CZ-NH1	5.70	123.15	120.30
1	A	58[B]	ARG	NE-CZ-NH1	5.70	123.15	120.30
1	A	26[A]	GLU	CG-CD-OE1	5.68	129.65	118.30
1	A	26[B]	GLU	CG-CD-OE1	5.68	129.65	118.30
1	A	26[A]	GLU	CG-CD-OE2	-5.66	106.99	118.30
1	A	26[B]	GLU	CG-CD-OE2	-5.66	106.99	118.30
1	A	89	ARG	NH1-CZ-NH2	5.49	125.43	119.40
1	A	174	ASP	CB-CG-OD2	-5.24	113.58	118.30
1	A	253	ARG	NE-CZ-NH1	5.08	122.84	120.30

There are no chirality outliers.

There are no planarity outliers.

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	2228	0	2067	36	0
2	A	1	0	0	0	0
3	A	7	0	10	2	0
4	A	405	0	0	21	2
All	All	2641	0	2077	36	2

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 8.

All (36) 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:58[A]:ARG:CZ	1:A:69[A]:GLU:OE1	1.69	1.38
1:A:167[B]:LYS:HD3	4:A:2298:HOH:O	1.14	1.26
1:A:26[B]:GLU:CD	4:A:2069:HOH:O	1.85	1.16
1:A:58[A]:ARG:NH2	1:A:69[A]:GLU:OE1	1.82	1.10
1:A:8[B]:ASP:OD1	4:A:2014:HOH:O	1.68	1.10

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:36[A]:ASP:OD1	4:A:2102:HOH:O	1.84	0.94
1:A:173:THR:OG1	3:A:1261:PEG:H32	1.72	0.90
1:A:174:ASP:O	3:A:1261:PEG:H31	1.76	0.85
1:A:251:LYS:O	1:A:252[B]:ASP:OD1	1.98	0.80
1:A:224[A]:LYS:HE3	4:A:2289:HOH:O	1.81	0.80
1:A:26[B]:GLU:OE2	4:A:2069:HOH:O	1.96	0.76
1:A:26[B]:GLU:CG	4:A:2069:HOH:O	2.24	0.75
1:A:26[A]:GLU:OE2	1:A:204:GLU:OE1	2.06	0.73
1:A:58[A]:ARG:NE	1:A:69[A]:GLU:OE1	2.25	0.69
1:A:131:GLY:O	1:A:135[B]:GLN:NE2	2.28	0.66
1:A:167[A]:LYS:NZ	4:A:2308:HOH:O	2.19	0.65
1:A:187[A]:VAL:HG22	4:A:2338:HOH:O	1.95	0.65
1:A:36[A]:ASP:HA	4:A:2100:HOH:O	1.96	0.65
1:A:4:HIS:N	4:A:2001:HOH:O	2.29	0.64
1:A:161:ASP:HA	4:A:2287:HOH:O	1.98	0.63
1:A:26[A]:GLU:CD	4:A:2067:HOH:O	2.39	0.60
1:A:58[A]:ARG:NH1	1:A:69[A]:GLU:OE1	2.32	0.59
1:A:14[A]:GLU:CD	4:A:2037:HOH:O	2.44	0.56
1:A:55[A]:THR:HG21	4:A:2150:HOH:O	2.08	0.54
1:A:164:ASP:OD1	1:A:167[B]:LYS:NZ	2.40	0.54
1:A:17:HIS:HD2	4:A:2033:HOH:O	1.93	0.52
1:A:167[B]:LYS:CD	4:A:2298:HOH:O	1.99	0.49
1:A:187[A]:VAL:CG2	4:A:2359:HOH:O	2.62	0.47
1:A:27[A]:ARG:HG3	1:A:204:GLU:HB3	1.98	0.45
1:A:133:ALA:O	1:A:139:GLY:HA3	2.17	0.44
1:A:214[B]:PRO:HG2	4:A:2360:HOH:O	2.17	0.43
1:A:26[A]:GLU:HG3	4:A:2067:HOH:O	2.19	0.42
1:A:6:GLY:O	1:A:12:GLY:HA2	2.20	0.42
1:A:30:PRO:HG3	1:A:106:GLU:HB3	2.02	0.41
1:A:26[A]:GLU:CD	1:A:204:GLU:OE1	2.57	0.41
1:A:14[A]:GLU:HG2	4:A:2037:HOH:O	2.22	0.40

All (2) 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)
4:A:2069:HOH:O	4:A:2280:HOH:O[1_455]	1.86	0.34
4:A:2066:HOH:O	4:A:2284:HOH:O[1_455]	1.99	0.21

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	277/262 (106%)	268 (97%)	9 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	243/225 (108%)	240 (99%)	3 (1%)	71	36

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	10[A]	HIS
1	A	10[B]	HIS
1	A	111	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	4	HIS
1	A	17	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 monosaccharides 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	$\#|Z| > 2$
3	PEG	A	1261	-	6,6,6	1.56	1 (16%)	5,5,5	1.49	2 (40%)

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	PEG	A	1261	-	-	2/4/4/4	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1261	PEG	O2-C3	2.62	1.53	1.42

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	1261	PEG	O1-C1-C2	2.36	125.50	111.81
3	A	1261	PEG	C3-O2-C2	-2.03	104.47	113.29

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1261	PEG	O1-C1-C2-O2
3	A	1261	PEG	C4-C3-O2-C2

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1261	PEG	2	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 ⓘ

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(Å²)	Q<0.9
1	A	257/262 (98%)	0.05	0	100	100	5, 8, 21, 45	0

There are no RSRZ outliers to report.

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 monosaccharides in this entry.

6.4 Ligands ⓘ

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. 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	B-factors(Å²)	Q<0.9
3	PEG	A	1261	7/7	0.92	0.12	8,20,29,29	0
2	ZN	A	268	1/1	1.00	0.05	6,6,6,6	1

6.5 Other polymers ⓘ

There are no such residues in this entry.