

# Supplementary Materials: Mechanism of action of non-synonymous single nucleotide variations associated with $\alpha$ -carbonic anhydrase II deficiency

Taremekedzwa Allan Sanyanga <sup>1</sup>, Bilal Nizami <sup>1,2</sup> and Özlem Tastan Bishop <sup>1\*</sup>

**Table S1.** Identified CA-II residues important for structure, function and stability.

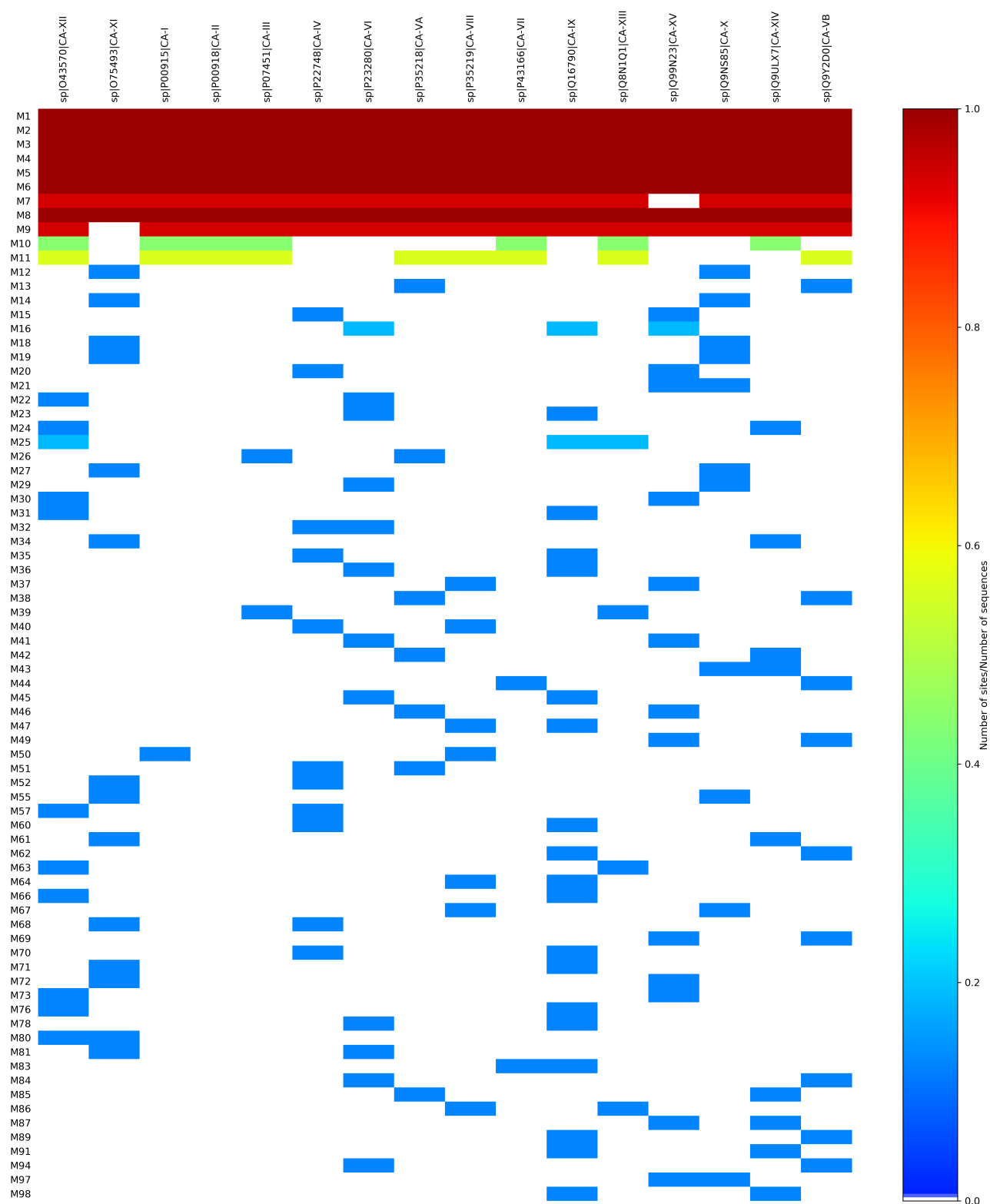
Residue	Functional role	Reference <sup>a</sup>
Trp5	Primary aromatic cluster residue, tertiary CO <sub>2</sub> binding pocket formation, His64 stabilisation when in the "out" conformation	[1–3]
Tyr7	Primary aromatic cluster residue, active site water network coordination	[1–9]
Trp16	Primary aromatic cluster residue	[1–3]
Phe20	Primary aromatic cluster residue	[1–3]
Ser29	Enzymatic stability	[10,11]
Asn62	Active site water network coordination	[4]
His64	Proton shuttling residue, tertiary CO <sub>2</sub> binding pocket formation	[1,4–8]
Phe66	Secondary aromatic cluster residue, secondary CO <sub>2</sub> binding pocket formation	[1,5–8]
Asn67	Active site water network coordination	[12]
Phe70	Secondary aromatic cluster residue	[1–3]
Phe93	Secondary aromatic cluster residue	[1–3]
Gln92	Secondary Zn <sup>2+</sup> ligand	[1,4]
His94	Zn <sup>2+</sup> coordination	[1,4]
Phe95	Secondary aromatic cluster residue, secondary CO <sub>2</sub> binding pocket formation	[1,5–8]
His96	Zn <sup>2+</sup> coordination	[1,4]
Trp97	Secondary aromatic cluster residue, secondary CO <sub>2</sub> binding pocket formation	[1,5–8]
Glu106	Orientation of Zn <sup>2+</sup> water ligand molecule for catalysis	[1]
Glu117	Zn <sup>2+</sup> affinity and catalytic efficiency, secondary Zn <sup>2+</sup> ligand	[13]
His119	Zn <sup>2+</sup> coordination	[1,4]
Val121	Primary CO <sub>2</sub> binding pocket formation	[1,5–8]
Val142	Primary CO <sub>2</sub> binding pocket formation	[1,5–8]
Phe175	Secondary aromatic cluster residue	[1–3]
Phe178	Secondary aromatic cluster residue	[1–3]
Leu197	Primary CO <sub>2</sub> binding pocket formation	[1,5–8]
Thr198	Deep water molecule stabilisation, catalytic orientation of Zn <sup>2+</sup> water ligand molecule	[4]
Thr199	Active site water coordination, tertiary CO <sub>2</sub> binding pocket formation	[4]
Pro200	Tertiary CO <sub>2</sub> binding pocket formation	[1,5–8]
Trp208	Primary CO <sub>2</sub> binding pocket formation	[1,5–8]
Phe225	Secondary aromatic cluster residue, secondary CO <sub>2</sub> binding pocket formation	[1,5–8]
Asn243	Tertiary CO <sub>2</sub> binding pocket formation, secondary Zn <sup>2+</sup> ligand	[1,5–8]
Arg245*	Enzyme stability	[14]

\* associated with stability reduction in CA-I

<sup>a</sup> references at end of supplementary data

**Table S2.** Table of UniProt CA accession numbers for sequences used in motif discovery.

Sequence	UniProt accession
CA-I	P00915
CA-II	P00918
CA-III	P07451
CA-IV	P22748
CA-VA	P35218
CA-VB	Q9Y2D0
CA-VI	P23280
CA-VII	P43166
CA-VIII	P35219
CA-IX	Q16790
CA-X	Q9NS85
CA-XI	O75493
CA-XII	O43570
CA-XIII	Q8N1Q1
CA-XIV	Q9ULX7
CA-XV	Q99N23



**Figure S1.** Heat map of all conserved motifs within human  $\alpha$ -CA family and associated UniProt accession. Motif conservation is represented as number of motif sites per total protein sequences. A value of zero shows that motifs are not conserved in any sequence, whereas a value of 1 shows that motif is conserved in all sequences. The prefix 'M' represents the word motif.

**Table S3.** Zn<sup>2+</sup> non-bonded, bonds, angles and dihedral parameters derived within this study. **K<sub>b</sub>**: bond force constant; **K<sub>θ</sub>**: angle force constant; **R<sub>min</sub>**: vdW radius; **ε**: LJ potential well energy.

**Non-bonded:**

Atom	R <sub>min</sub> (Å)	ε (kcal/mol)
M1	1.40	0.02
Y1	1.82	0.17
Y2	1.82	0.17
Y3	1.82	0.17
Y4	1.77	0.15

**Bonds:**

Bond type	K <sub>b</sub> (kcal/mol/Å <sup>2</sup> )	Bond length (Å)
M1-Y4	41.20	2.12
Y1-M1	91.70	1.98
Y2-M1	94.30	1.98
Y3-M1	93.00	1.98

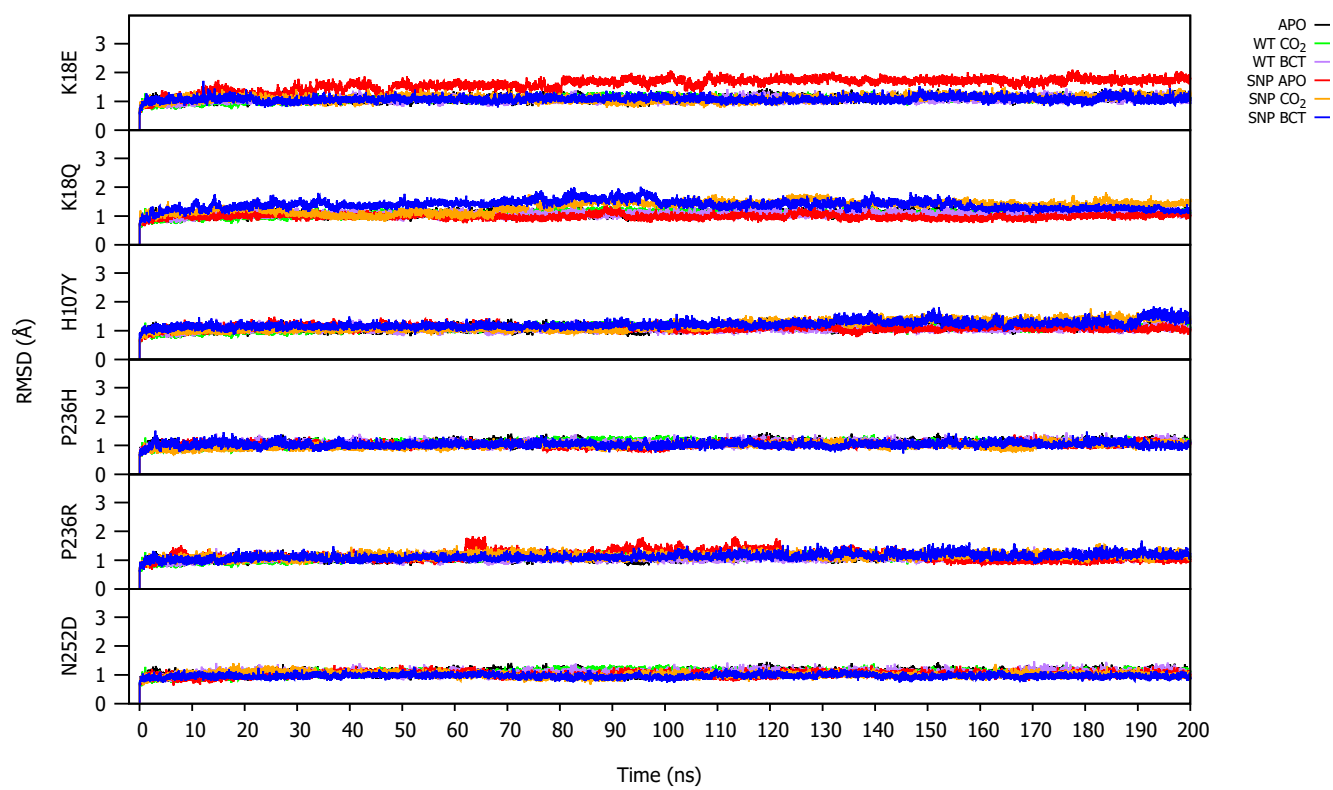
**Angles:**

Angle type	K <sub>θ</sub> (kcal/mol/radian <sup>2</sup> )	Equilibrium angle degrees (θ)
CC-Y3-M1	56.44	127.30
CR-Y1-M1	38.97	125.70
CR-Y2-M1	53.67	127.46
M1-Y1-CV	39.62	128.05
M1-Y2-CV	54.89	126.22
M1-Y3-CR	54.18	125.48
M1-Y4-HW	44.55	122.59
Y1-M1-Y2	39.89	116.07
Y1-M1-Y3	37.44	115.90
Y1-M1-Y4	31.02	101.10
Y2-M1-Y3	36.29	114.63
Y2-M1-Y4	36.83	105.42
Y3-M1-Y4	30.04	100.62

**Dihedral:**

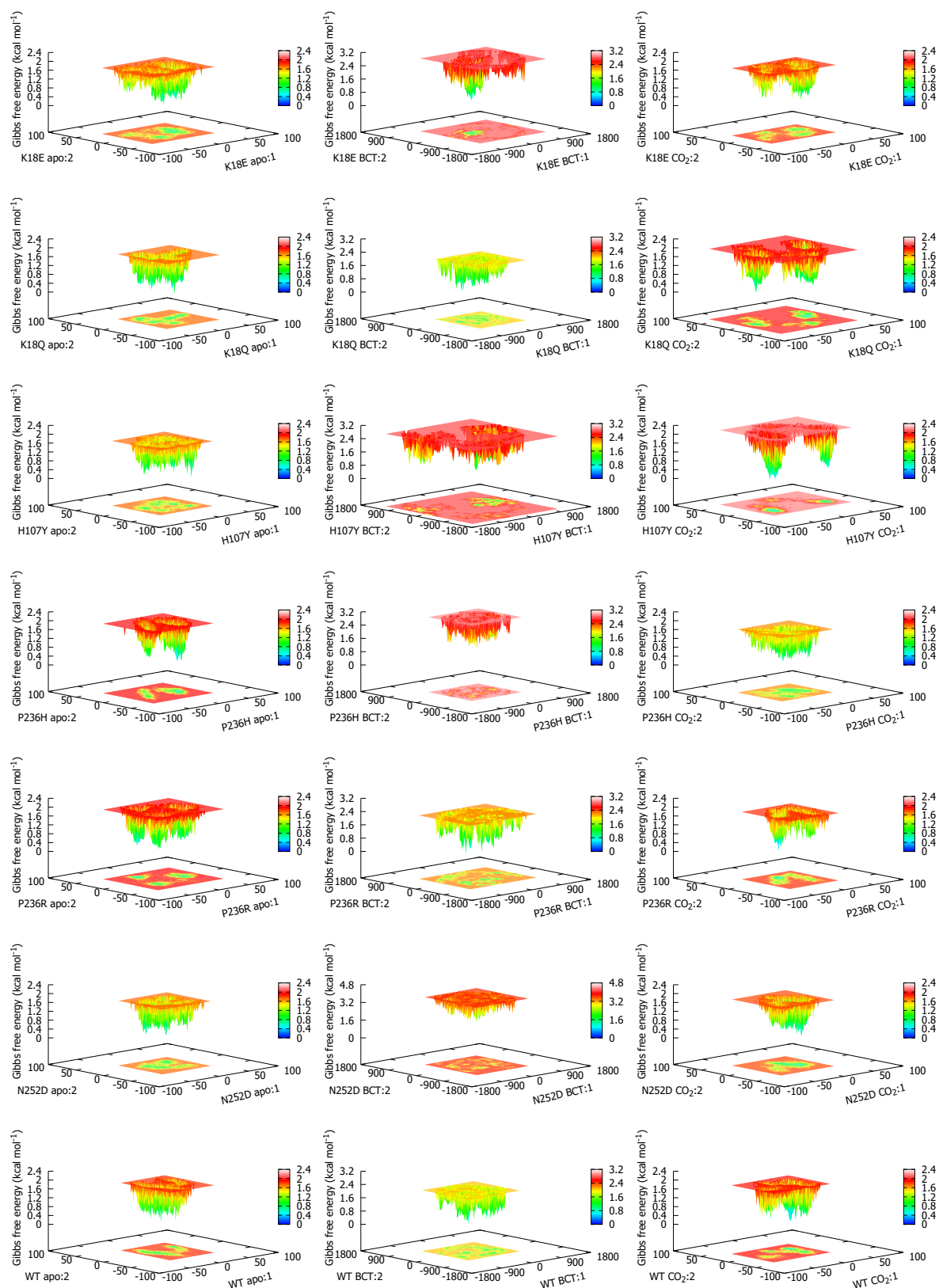
Definition	Divider	Barrier (kcal/mol)	Phase degrees (θ)	Periodicity
X -CC-Y3-X	2	4.80	180.0	2.0
X -CR-Y1-X	2	10.00	180.0	2.0
X -CR-Y2-X	2	10.00	180.0	2.0
X -CV-Y1-X	2	4.80	180.0	2.0
X -CV-Y2-X	2	4.80	180.0	2.0
X -Y3-CR-X	2	10.00	180.0	2.0
CX-CT-CC-Y3	1	0.05	180.0	-4.0
CX-CT-CC-Y3	1	0.74	0.0	-3.0
CX-CT-CC-Y3	1	0.20	0.0	-2.0
CX-CT-CC-Y3	1	0.69	0.0	1.0

M1: Zn; Y1: His94 NE2 (epsilon nitrogen); Y2: His96 NE2 ((epsilon nitrogen)) Y3: His199 ND1 (delta hydrogen); Y4: O (H<sub>2</sub>O); CC: CG (gamma carbon); CR: CE1 (epsilon carbon); CV: CD2 (delta carbon)



**Figure S2.** RMSD comparison between the WT and variant proteins during MD.



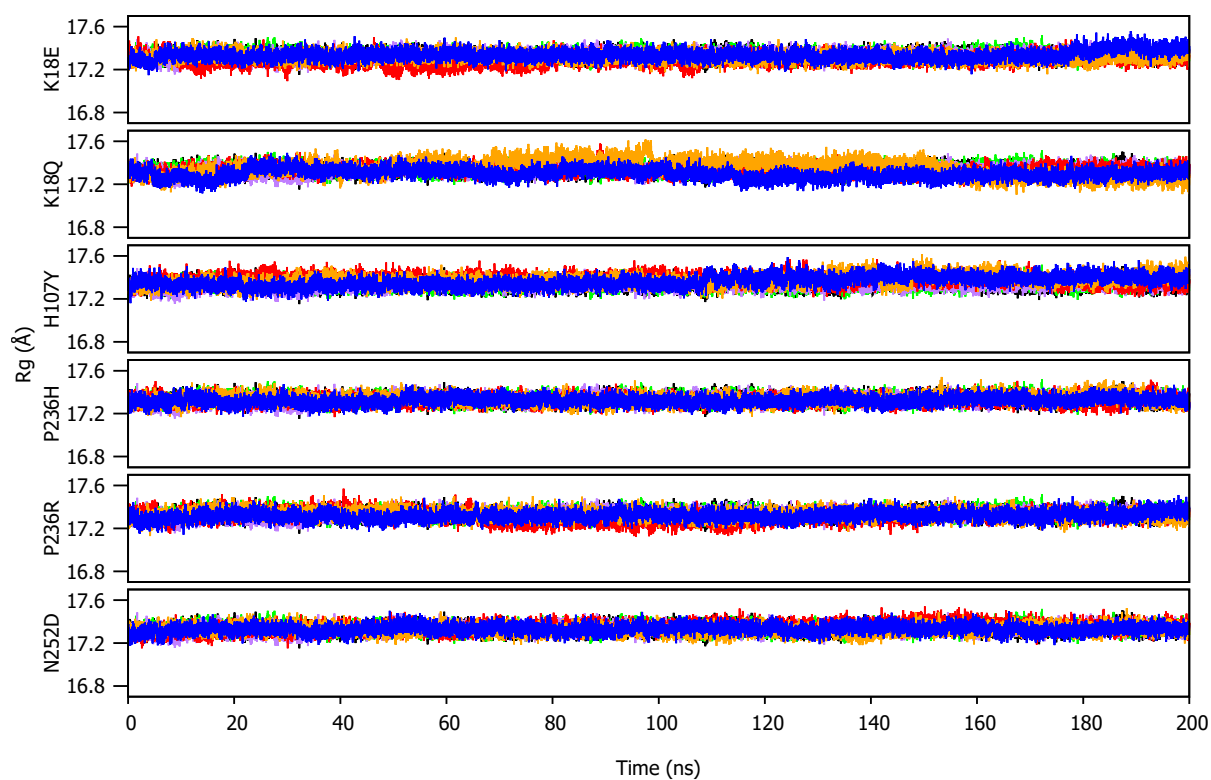


**Figure S3.** 3-dimensional (3D) plot of PC1 vs PC2 of the WT and variants proteins as a function of free energy. 2D PCA is projected onto the x-axis and y-axis. Free energy is represented in kcal mol<sup>-1</sup>.

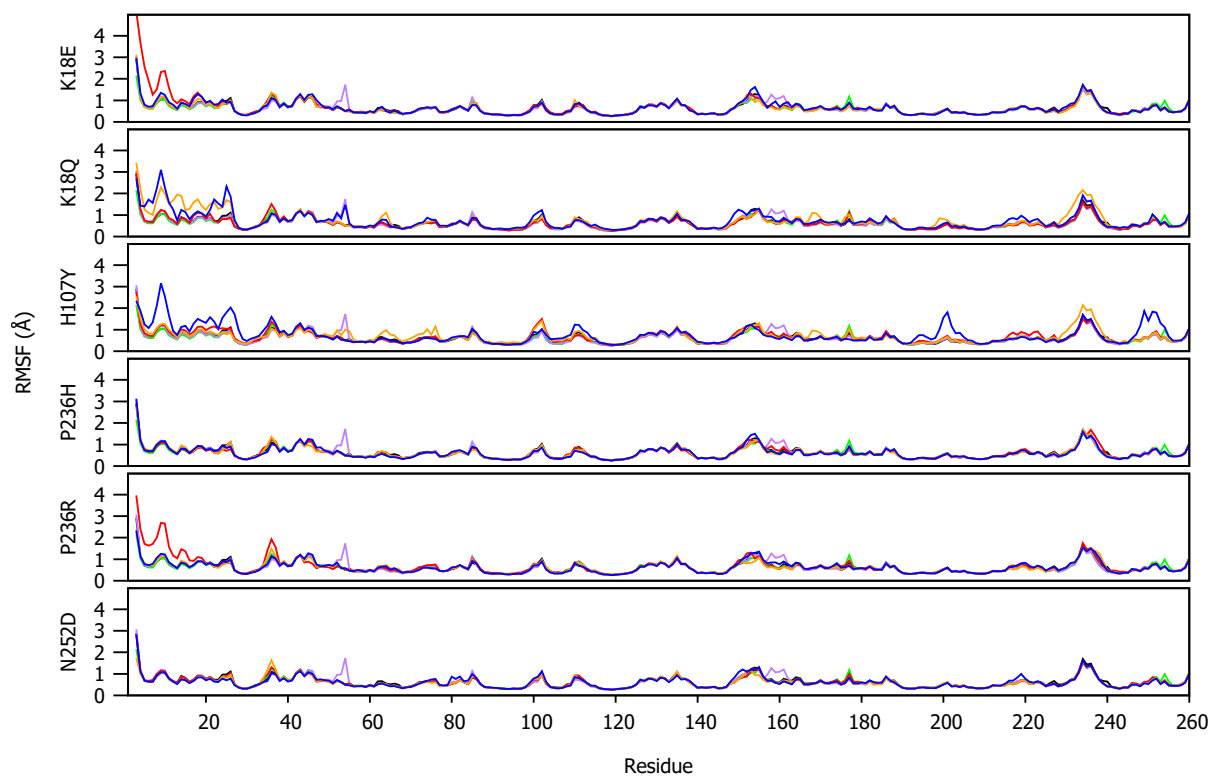
**Table S5.** Eigenvalue fraction of each principal component for the active site residues within the wild-type and variant proteins in the absence (apo) and presence (non apo) of CO<sub>2</sub>.

<b>Variant</b>	<b>Protein</b>	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>
<b>K18E</b>	Apo	0.5503	0.2500	0.1997
	BCT	0.4836	0.3165	0.1999
	CO <sub>2</sub>	0.5772	0.2218	0.2010
<b>K18Q</b>	Apo	0.5260	0.2732	0.2009
	BCT	0.4497	0.3598	0.1905
	CO <sub>2</sub>	0.6091	0.2647	0.1261
<b>H107Y</b>	Apo	0.4439	0.3322	0.2239
	BCT	0.5688	0.2764	0.1548
	CO <sub>2</sub>	0.7201	0.1670	0.1129
<b>P236H</b>	Apo	0.5535	0.2628	0.1837
	BCT	0.5195	0.2741	0.2064
	CO <sub>2</sub>	0.4846	0.2977	0.2177
<b>P236R</b>	Apo	0.4693	0.3107	0.2200
	BCT	0.4772	0.3415	0.1813
	CO <sub>2</sub>	0.4218	0.3208	0.2574
<b>N252D</b>	Apo	0.4415	0.3358	0.2227
	BCT	0.4635	0.3607	0.1758
	CO <sub>2</sub>	0.4348	0.3235	0.2417
<b>WT</b>	Apo	0.4673	0.3253	0.2074
	BCT	0.4256	0.3550	0.2194
	CO <sub>2</sub>	0.5399	0.2589	0.2012





**Figure S4.** Rg comparison between the WT and variant proteins during MD.



**Figure S5.** Residue RMSF comparison between the WT and variant proteins.

**Table S6.** Variant residues showing decreases and increases to  $\Delta L$  during MD simulation. Residues from Table reftab:residues are underlined and highlighted in bold. SNV positions are underlined, italicised and highlighted in bold red.

Variant	Apo	BCT	CO <sub>2</sub>
<b><math>\Delta L</math> decrease (Residue accessibility increase)</b>			
<b>K18E</b>	His4 His10 <b><u>Trp16</u></b> Ile22 Gly170 Phe230 Asn231 Glu235 Pro236 Glu237 Glu238	Lys111 Gly155 Leu156	Ala54 Lys158 Val159 Val160 Asp161 Leu163 Asp164 Ser165 Ile166 Asn177 Phe178
<b>K18Q</b>	His3 Ile22 Ala152	His3 His4 <b><u>Trp5</u></b> Gly6 Gly8 Lys9 His10 Asn11 Lys24 Lys111	His3 His4 Lys9 Lys24 Lys158 Val159 Val160 Asp161 Leu163 Asp164 Ser165 Ile166
<b>H107Y</b>	Lys24	Val49 Ser50 Asp52 Gln53 Lys111 Leu184	Asp52 Gln53 Gly155 Lys158 Val159 Val160 Asp161 Val162 Leu163 Ser165 Ile166 Asp179 Pro180
<b>P236H</b>	Ile22 Asp101 Gly155 Thr199 His236 Glu237 Glu238	His3 Gly155 Leu156 Phe178 Asp179 Pro180 Arg181 Leu183 Pro201	His3 Ala54 Lys158 Val159 Val160 Asp161 Val162 Leu163 Asp164 Ser165 Ile166 Ala173 Phe175 Asn177 Glu237
<b>P236R</b>	Pro21 Ile22 Phe230 Asn231 Glu235 Arg236 Glu237 Glu238	Val49 Asp52 Gln53 Pro180	Asp52 Ile59 Lys158 Val159 Val160 Asp161 Val162 Leu163 Ile166 Pro180 Glu237
<b>N252D</b>	Leu163 Asp164	Lys111 Gly155 Leu156 Gln157 Leu183	Ala54 Lys158 Val159 Val160 Asp161 Val162 Leu163 Asp164 Ser165 Ile166 Phe175
<b><math>\Delta L</math> increase (Residue accessibility decrease)</b>			
<b>K18E</b>	<b><u>Trp5</u></b> Lys111 Lys153 Leu188	Ala152 Lys153	Gly155 Leu156 Leu188
<b>K18Q</b>	Thr35 Lys111 Pro154 Gly155 Leu156 Lys158 Val159 Pro180 Gly182 Leu183 Glu220	Gly12 Pro13 Glu14 Asp19 Leu163 Glu233	Ile22 Ala54 Ala152
<b>H107Y</b>	<b><u>Ser29</u></b> Val31 Asp101 Lys111 Leu140 Ala152 Lys153 Val241	Val31 Ala54 Lys158	Gly25 Glu26 Arg27 Ala54 Lys111 Gln248 Lys251
<b>P236H</b>	His3 Thr35 Ala152 Lys158 Val159 Val160	His36 Thr37 Leu163 Asp164 Gly170	Leu156 <b><u>Phe225</u></b> Arg226
<b>P236R</b>	His3 Gly6 <b><u>Tyr7</u></b> Thr35 Gly98 Ser99 Leu100 Lys111 Asp242	Ala54 Leu163 Asp164 Gly170 <b><u>Thr199</u></b> Gly232 Gly234	Ala54 Leu156 Gly170
<b>N252D</b>	His3 Thr35 Gly63 Pro154 Gly155 Leu156 Gln157 Lys158 Pro180 Leu183	His3 Thr35 Ser43 Ala152 Lys153	Gly155 Leu156 Arg226

**Table S7.** Variant residues showing decreases and increases to  $\Delta BC$  during MD simulation. Residues from Table S1 are underlined and highlighted in bold. SNV positions are underlined, italicised and highlighted in bold red.

Variant	Apo	BCT	CO <sub>2</sub>
<b><math>\Delta BC</math> decrease (reduction in residue communication)</b>			
K18E	<u>Trp5</u> Gly6 Pro200 Met240 <u>Asp242</u> <u>Asn243</u>	<u>His64</u> His107 Lys113 <u>Glu117</u> Ile145 Leu147 Lys153 <u>Thr199</u> Ile215 Val222	Asn61 Ala77 Leu90 Val162 Gly182 Leu183 Val217 <b>Phe225</b>
K18Q	Gly63 <u>His94</u> <u>Glu117</u> <u>His119</u> Leu147 Pro180 Leu183 Asp242 Trp244	<u>Trp5</u> Gly12 Pro13 <u>Trp16</u> <u>Phe66</u> <u>Asn67</u> His107 <u>Glu117</u> Thr168 Asp242 <u>Asn243</u> Pro246	<u>Trp5</u> Ala54 Leu60 Asn61 <u>His64</u> Phe70 Gly182 <u>Phe225</u> Asp242
H107Y	Val31 Gly63 <u>His94</u> <u>Glu117</u> <u>His119</u> Leu140 <u>Val142</u> Gly144	Val31 Ala54 <u>Phe66</u> <u>Tyr107</u> <u>Glu117</u> Gly144 <u>Thr199</u>	Ile33 Ala54 <u>His64</u> Glu69 <u>Glu117</u> Leu147 <u>Thr199</u> Val222 <b>Phe225</b>
P236H	Gly63 Val68 <u>Gln92</u> Val121 <u>Val142</u> Leu147 Val159 Asp242	<u>Phe66</u> <u>Phe70</u> <u>Gln92</u> <u>Phe95</u> <u>Glu117</u> Leu163	Asn61 Ala77 Leu90 Leu147 Gly182 Leu183 Val222 <b>Phe225</b>
P236R	Gly6 <u>His94</u> Ser105 <u>His119</u> <u>Val142</u> Asp242 <u>Asn243</u>	Tyr51 Ala54 <u>Asn67</u> <u>Phe95</u> <u>Glu117</u> Leu118 Leu147 <u>Thr199</u>	Ala54 Leu60 Asn61 Glu69 Gly182 Leu183 <u>Phe225</u>
N252D	Gly63 <u>His64</u> <u>His94</u> <u>His119</u> Leu156 Pro180 Leu183 <u>Thr199</u> <b>Phe225</b>	<u>His64</u> His107 <u>Glu117</u> Leu147 Lys153 <u>Thr199</u> Ile215	Leu60 Asn61 <u>Phe70</u> Ala77 Gly182 <b>Phe225</b>
<b><math>\Delta BC</math> increase (residue communication increase)</b>			
K18E	His4 <u>Trp16</u> <u>His96</u> <u>Thr199</u> Phe230	Ser105 Ala116 Leu118 Gly155 Leu156 Leu183 Val210 Val217	Ala54 Ala65 Val68 <b>Phe93</b> <b>Phe95</b> Ala116 Leu118 Val159 <b>Thr199</b>
K18Q	<u>Phe66</u> <u>His96</u> Lys113 Ala116 Ile145 Ile215 Val217	Gly8 Asn11 <u>His94</u> Ser105 Lys169 Phe230 Trp244 Gln248	<u>Tyr7</u> Gly8 Asn11 Ala65 <b>Phe66</b> Val68 <u>His96</u> Val160 Leu163 <u>Thr199</u> Phe230
H107Y	<u>Phe66</u> <u>Asn67</u> <u>Phe95</u> Glu106 Leu118 Val217	Thr55 Ala77 Leu90 <u>His94</u> Ser105 Val210	Ala65 <u>Phe66</u> <u>Phe93</u> <u>Phe95</u> <u>His96</u> Glu106 Ala116 Asp179
P236H	<u>His64</u> <u>Phe66</u> <u>His96</u> Ala116 Gly155 <u>Thr199</u> Val217	Gly63 <u>Phe93</u> <u>His94</u> <u>Trp97</u> Val134	Ala54 Ala65 <u>Phe66</u> <u>Phe93</u> <b>Phe95</b> Ala116 Leu163
P236R	<u>Tyr7</u> <u>Phe66</u> <u>Phe95</u> <u>His96</u> Glu106 <u>Thr199</u> Phe230 Arg245	Ala77 <u>His94</u> Ala116 <u>His119</u> Val217	Ala65 <u>Phe66</u> <u>His94</u> Ala116 Leu163 Phe175 Pro180
N252D	<u>Phe66</u> <u>Asn67</u> <u>Gln92</u> <u>Phe95</u> <u>His96</u> Ile215 <u>Asn243</u>	Ser105 Ala116 Gly155 Leu156 Leu183 Val217 <u>Asn243</u> Trp244	Tyr51 Ala54 Ala65 <b>Phe66</b> Val68 Ile91 <u>Gln92</u> <b>Phe93</b> <u>Val142</u> Val160 Leu163 Ile215

**References (Table S1)**

1. Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacology & therapeutics* **1997**, *74*, 1–20.
2. Eriksson, A.E.; Jones, T.A.; Liljas, A. Refined structure of human carbonic anhydrase II at 2.0 Å resolution. *Proteins: Structure, Function, and Bioinformatics* **1988**, *4*, 274–282.
3. Hunt, J.A.; Fierke, C.A. Selection of carbonic anhydrase variants displayed on phage aromatic residues in zinc binding site enhance metal affinity and equilibration kinetics. *Journal of Biological Chemistry* **1997**, *272*, 20364–20372.
4. Silverman, D.N.; McKenna, R. Solvent-mediated proton transfer in catalysis by carbonic anhydrase. *Accounts of chemical research* **2007**, *40*, 669–675.
5. Merz Jr, K.M. Carbon dioxide binding to human carbonic anhydrase II. *Journal of the American Chemical Society* **1991**, *113*, 406–411.
6. Liang, J.Y.; Lipscomb, W.N. Binding of substrate CO<sub>2</sub> to the active site of human carbonic anhydrase II: a molecular dynamics study. *Proceedings of the National Academy of Sciences* **1990**, *87*, 3675–3679.
7. Domsic, J.F.; Avvaru, B.S.; Kim, C.U.; Gruner, S.M.; Agbandje-McKenna, M.; Silverman, D.N.; McKenna, R. Entrapment of carbon dioxide in the active site of carbonic anhydrase II. *Journal of Biological Chemistry* **2008**, *283*, 30766–30771.
8. Alexander, R.S.; Nair, S.K.; Christianson, D.W. Engineering the hydrophobic pocket of carbonic anhydrase II. *Biochemistry* **1991**, *30*, 11064–11072.
9. Mikulski, R.L.; Silverman, D.N. Proton transfer in catalysis and the role of proton shuttles in carbonic anhydrase. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2010**, *1804*, 422–426.
10. Mårtensson, L.G.; Jonsson, B.H.; Andersson, M.; Kihlgren, A.; Bergenheim, N.; Carlsson, U. Role of an evolutionarily invariant serine for the stability of human carbonic anhydrase II. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1992**, *1118*, 179–186.
11. Almstedt, K.; Lundqvist, M.; Carlsson, J.; Karlsson, M.; Persson, B.; Jonsson, B.H.; Carlsson, U.; Hammarström, P. Unfolding a folding disease: folding, misfolding and aggregation of the marble brain syndrome-associated mutant H107Y of human carbonic anhydrase II. *Journal of molecular biology* **2004**, *342*, 619–633.
12. Silverman, D.N.; Lindskog, S. The catalytic mechanism of carbonic anhydrase: implications of a rate-limiting protolysis of water. *Accounts of chemical research* **1988**, *21*, 30–36.
13. Kiefer, L.L.; Paterno, S.A.; Fierke, C.A. Hydrogen bond network in the metal binding site of carbonic anhydrase enhances zinc affinity and catalytic efficiency. *Journal of the american chemical society* **1995**, *117*, 6831–6837.
14. Tashian, R.E. Genetics of the mammalian carbonic anhydrases. In *Advances in genetics*; Elsevier, 1992; Vol. 30, pp. 321–356.