



Article

# Rabbit and Human Angiotensin-Converting Enzyme-2: Structure and Electric Properties

Svetlana H. Hristova <sup>1</sup>, Trifon T. Popov <sup>2</sup> and Alexandar M. Zhivkov <sup>3,\*</sup>

<sup>1</sup> Department of Medical Physics and Biophysics, Medical Faculty, Medical University—Sofia, Zdrave Str. 2, 1431 Sofia, Bulgaria

<sup>2</sup> Medical Faculty, Medical University—Sofia, Zdrave Str. 2, 1431 Sofia, Bulgaria

<sup>3</sup> Scientific Research Center, “St. Kliment Ohridski” Sofia University, 8 Dragan Tsankov Blvd., 1164 Sofia, Bulgaria

\* Correspondence: zhivkov\_ipc@doctor.bg

**Abstract:** The angiotensin-converting enzyme-2 (ACE2) is a transmembrane glycoprotein, consisting of two segments: a large carboxypeptidase catalytic domain and a small transmembrane collectrin-like segment. This protein plays an essential role in blood pressure regulation, transforming the peptides angiotensin-I and angiotensin-II (vasoconstrictors) into angiotensin-1-9 and angiotensin-1-7 (vasodilators). During the COVID-19 pandemic, ACE2 became best known as the receptor of the S-protein of SARS-CoV-2 coronavirus. The purpose of the following research is to reconstruct the 3D structure of the catalytic domain of the rabbit enzyme rACE2 using its primary amino acid sequence, and then to compare it with the human analog hACE2. For this purpose, we have calculated the electric properties and thermodynamic stability of the two protein globules employing computer programs for protein electrostatics. The analysis of the amino acid content and sequence demonstrates an 85% identity between the two polypeptide chains. The 3D alignment of the catalytic domains of the two enzymes shows coincidence of the  $\alpha$ -helix segments, and a small difference in two unstructured segments of the chain. The electric charge of the catalytic domain of rACE2, determined by 70 positively chargeable amino acid residues, 114 negatively chargeable ones, and two positive charges of the  $\text{Zn}^{2+}$  atom in the active center exceeds that of hACE2 by one positively and four negatively chargeable groups; however, in 3D conformation, their isoelectric points pI 5.21 coincide. The surface electrostatic potential is similarly distributed on the surface of the two catalytic globules, but it strongly depends on the pH of the extracellular medium: it is almost positive at pH 5.0 but strongly negative at pH 7.4. The pH dependence of the electrostatic component of the free energy discloses that the 3D structure of the two enzymes is maximally stable at pH 6.5. The high similarity in the 3D structure, as well as in the electrostatic and thermodynamic properties, suggests that rabbit can be successfully used as an animal model to study blood pressure regulation and coronavirus infection, and the results can be extrapolated to humans.

**Keywords:** proteins; enzymes; angiotensin-converting enzyme-2 (ACE2); macromolecular models; 3D structural stability; folding energy; protein electrostatics; isoelectric point; surface electrostatic potential



**Citation:** Hristova, S.H.; Popov, T.T.; Zhivkov, A.M. Rabbit and Human Angiotensin-Converting Enzyme-2: Structure and Electric Properties. *Int. J. Mol. Sci.* **2024**, *25*, 12393. <https://doi.org/10.3390/ijms252212393>

Academic Editor: Andreas Beineke

Received: 9 October 2024

Revised: 14 November 2024

Accepted: 16 November 2024

Published: 19 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The angiotensin-converting enzyme-2 (ACE2) is an integral membrane glycoprotein, expressed on the extracellular surface of the cytoplasmic membranes of human and animal epithelial cells. The main function of ACE2 as a carboxypeptidase (exopeptidase) is the reduction of blood pressure by detaching one amino acid residue from the C-end of angiotensin-1 (inactive decapeptide) and its derivative, angiotensin-2 (octapeptide, the strongest vascular constrictor) (Figure 1) [1–4]. ACE2 was identified as an isoform of the ACE enzyme in 2000 [5–8]; before then, researchers did not distinguish these two enzymes,

considering them identical. At present, ACE2 is best known as a receptor for the S-protein of the  $\beta$ -coronavirus SARS-CoV-2 (originator of the COVID-19 pandemic), although this role can also be played by other membrane-integrated proteins [9–12]. ACE2 is highly expressed in epithelial cells of the small intestine and colon, kidney, gallbladder, testis, blood vessels (endothelium), as well as in cardiomyocytes, placental trophoblasts, etc.; in the respiratory tract, ACE2 expression is not so high [13–16].

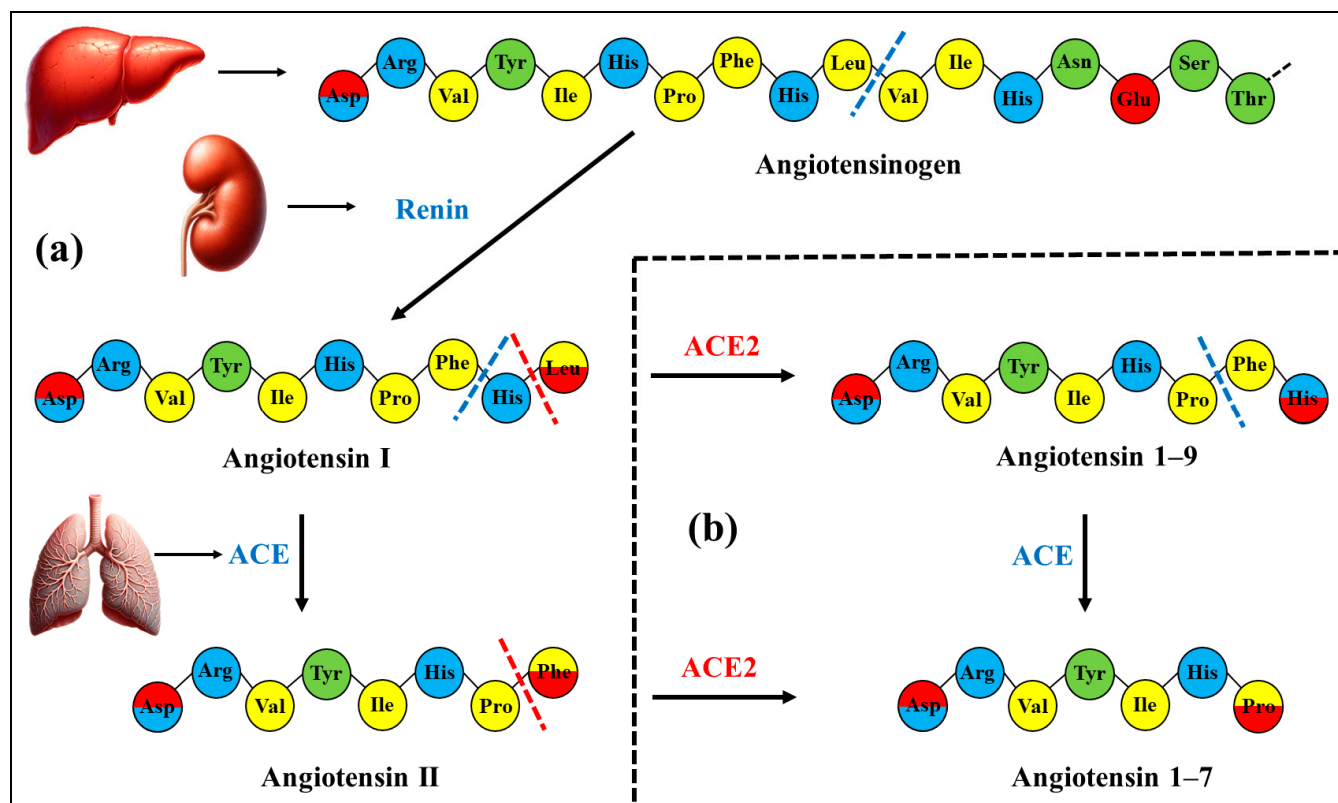


**Figure 1.** Renin-angiotensin system. *Subsection (a)* on the left: Renin and angiotensin-converting enzyme (ACE) action. The precursor  $\alpha$ -2-globulin angiotensinogen is produced by hepatocytes. The renal enzyme renin cleaves the covalent peptide bond after the first 10 amino acids from the N-terminus of angiotensinogen, leading to the formation of angiotensin-I (first reaction). This decapeptide is converted by the pulmonary angiotensin-converting enzyme (ACE) to the octapeptide angiotensin-II by the cleavage of the last two amino acid residues, resulting in the emergence of a carboxylic group on the C-terminus (second reaction). *Subsection (b)* on the right: Angiotensin-converting enzyme-2 (ACE2) action. The angiotensin-converting enzyme-2 (ACE2) cleaves one amino acid residue from the C-terminus of both angiotensin peptides, which leads to the formation of the nonapeptide angiotensin-1-9 and the heptapeptide angiotensin-1-7, respectively. The cleaved peptide bonds (in the two subsections) are shown by dotted lines colored according to the corresponding enzyme: blue (ACE) or red (ACE2). The amino acid residues are colored according to their charge and hydrophilicity: green (uncharged hydrophilic), blue (positively charged hydrophilic), red (negatively charged hydrophilic), and yellow (uncharged hydrophobic); the charges are determined at neutral pH. The end-side residues are marked by double color considering the protonated  $\alpha$ -amino group ( $\text{NH}_3^+$ ) on the N-terminus and the deprotonated carboxylic group ( $\text{COO}^-$ ) on the C-terminus. The N-end aspartic acid (Asp) residue bears one positive ( $\text{NH}_3^+$ ) and one negative ( $\text{COO}^-$ ) charge. The C-end histidine (His) residue of the angiotensin-1-9 bears one negative charge (the deprotonated carboxyl group) and one positive charge at acid pH (the protonated imidazole group), which disappears at basic pH. The C-ends of the remaining three peptides have only negative charge ( $\text{COO}^-$ ).

ACE2 is a conservative protein, since the analysis of its primary amino acid sequence reveals relatively small changes during evolution: the identity is 83%, 81%, 83%, 61%, 60%, and 59% for civet, bat, bird, snake, frog, and fish, respectively, as compared with the human polypeptide chain [14]. The human ACE2 protein (hACE2) is encoded by a gene, located on the short arm of the X chromosome, which consists of 18 exons and 22 introns. The full-length hACE2 polypeptide chain contains 805 amino acids residues, which consists of a signal peptide on the extracellular N-terminus (1–19 residues), a large hydrophilic carboxypeptidase segment (20–615 residues), and a smaller hydrophobic intramembrane collectrin-like segment (616–805 residues). The molecular mass of hACE2 is 92.5 kDa of the whole polypeptide chain, 90.7 kDa without the signal peptide, and 69.1 kDa of the catalytic domain, all calculated according to the amino acid content without the covalently bound saccharides molecules in the real ACE2 glycoprotein. ACE2 in its final form is a membrane-integrated globular protein consisting of a large hydrophilic catalytic domain exposed to the extracellular aqueous medium and a small hydrophobic nonenzymatic domain integrated into the cytoplasmic membrane. In contrast to the ACE isoform, the ACE2 protein has only one catalytic domain (from 147 to 555 amino acid residues, numbered from the N-end), which shares 42% identity with each of the two ACE catalytic domains [6].

The ACE2 catalytic domain has a single metalloproteinase active site, constructed by the zinc-binding motif His-Glu-Met-Gly-His (HEMGH). The amino acid residues, which form the active site of this enzyme, are hydrophilic and charged positively (His-374, His-378, His-505), negatively (Glu-375, Glu-402, Glu-406), and uncharged (Tyr-515). One divalent  $\text{Zn}^{2+}$  cation is coordinately bound to five atoms: nitrogen of the imidazole groups of the two histidine residues from the zinc-binding motif (His-374 and His-378), one oxygen atom of the glutamic acid from the zinc-binding motif (Glu 375), and the two oxygen atoms of the ionized carboxylic group ( $-\text{CO}^{-1/2}\text{O}^{-1/2}$ ) of the amino acid residue of the C-terminus of the substrate (leucine residue in the case of angiotensin-1) [17]. The enzyme splits the peptide bond between the first and second amino acid residues (on the C-end of the peptide chain) due to the shift of the  $\pi$ -electron density in the side carboxylic group  $-\text{C}=\text{O}$  to the zinc cation in the coordination bond  $-\text{C}=\text{O} \cdots \text{Zn}^{2+}$ , which results in the weakening of the peptide bond ( $-\text{NH}-\text{CO}-$ ). The catalytic center is located around a pocket in the extramembrane domain; its size and shape of the pocket determine different enzymatic activity [17–20].

Besides the hydrophobicity of the C-end residue, the enzyme activity probably depends on the electric charges of the angiotensin peptides and the electrostatic potential on the surface of the catalytic domain of the ACE2 enzyme because the adsorption of the peptide is a preliminary step in the enzyme reaction. The decapeptide angiotensin-I has four positively charged groups: one arginine residue, two histidine residues, and  $-\text{NH}_3^+$  group of the N-end aspartic acid, and two negatively charged groups: one of the N-end aspartic acid and the deprotonated carboxylic group of the C-end leucine. The octapeptide angiotensin-II has three positive charges (arginine, histidine, and N-end aspartic acid) and two negative charges (N-end aspartic acid and the  $\text{COO}^-$  group of the C-end phenylalanine) (Figure 1).

In this article, we compare the rabbit (rACE2) and the human (hACE2) angiotensin-converting enzyme-2 by considering their amino acid sequence, 3D structure, pH-dependent thermodynamic stability, hydrophilicity/hydrophobicity, and the local electrostatic potential on the surface of the catalytic domain. The comparison of hACE2 and rACE2 follows the need to extrapolate the experimental results obtained for the rabbit to the human ACE2. Although the catalytic site is conservative—in particular, it is identical in the hACE2 and rACE2—there is some difference in the structural stability and the electric properties because they are determined by the whole polypeptide chain and its 3D conformation. The three-dimensional (3D) structure of the human ACE2 is determined by crystallographic X-ray analysis of the crystallized hydrophilic extramembrane domain. The 1R42 model of hACE2, deposited in the Protein Data Bank (PDB), has 597 amino acid residues (from 19 to 615, numbered from the N-end) with molecular mass 76.77 kDa. In the PDB, there

is no deposited 3D model of a free rabbit ACE2; only its primary amino acid sequence is published. Therefore, before calculation of the thermodynamic and electrostatic properties of rACE2, it is necessary to reconstruct its 3D structure. To do that, we replace the different amino acid residues in the 3D model of the human ACE2 with the corresponding residues in the rabbit ACE2, applying mutation analyses by a computer program as we did for single mutations in our previous research [21].

Our interest in the rACE2 is conditioned by the use of rabbits (mainly European rabbit—*Oryctolagus cuniculus*) as experimental animals, along with rats, mice, guinea pigs, and bulls [22–25]. Compared to rodents, rabbit genes encoding important components of the immune system are phylogenetically more similar to human genes, suggesting a greater similarity of immune response in studies of various infectious diseases between rabbit and human than mouse and human, for example. There are numerous studies in which rabbits are used as model animals for the investigation of both infectious (tuberculosis, syphilis, human immunodeficiency virus—HIV, human papillomaviruses, etc.) and non-infectious (atherosclerosis, oncological diseases, neurodegenerative diseases such as Alzheimer's disease, autoimmune diseases such as systemic lupus erythematosus, arthritis, etc.) diseases, as well as experimental surgery [26]. The data in the literature on the emergence of natural COVID-19 infection in domestic rabbits [27], as well as the successful intranasal infection of New Zealand white rabbits with SARS-CoV-2 [28], suggest that rabbits can be used as a model for the study of coronavirus infection. Since ACE2 is the major receptor through which the coronavirus penetrates host cells, blocking or disrupting the interaction between ACE2 and S protein may result in the reduced entry of viral particles [29].

## 2. Results and Interpretation

The comparison of the electrical properties of rabbit (rACE2) and human (hACE2) angiotensin-converting enzymes requires first the reconstruction of the three-dimensional (3D) structure of rACE2 and then the calculation of the surface electrical potential, which determines, along with the surface relief, the non-covalent binding energy of the peptides, as well as the association energy with the coronavirus S-protein.

### 2.1. Structural Analysis

As a first step, we compared the amino acid sequences of the polypeptide chains of the two ACE2 proteins corresponding to their genetic code. The chain is divided into three segments demarcated by vertical lines in Figure 2 (amino acid residues are numbered from the N-terminus): (a) signal peptide 1–18, (b) catalytic domain 19–615, and (c) transmembrane segment 616–805. The distinctive residues between hACE2 and rACE2 are colored based on their polarity and charge, with negatively charged hydrophilic aspartic acid (D) and glutamic acid (E) in red; positively charged lysine (K), arginine (R), and histidine (H) in blue; and uncharged hydrophilic serine (S), asparagine (N), threonine (T), glutamine (Q), and cysteine (C) in green; and the uncharged hydrophobic tryptophan (W), leucine (L), methionine (M), alanine (A), isoleucine (I), glycine (G), valine (V), phenylalanine (F), tyrosine (Y), and proline (P) in yellow. The identical amino acids are uncolored. Comparison of the primary amino acid sequence of the two proteins shows a high identity of 85% between the two chains (686 of 805 residues are identical, and 119 are different), and no deletions are present. As it can be seen on Figure 2, a great number of point mutations (41 of 197 amino acid residues are different) is concentrated in the transmembrane collectrin-like segment, but only 10 of them (less than one fourth) are associated with a change in hydrophobicity and hydrophilicity. The catalytic segment of both chains is composed of 597 amino acid residues, 520 of which are identical; the different 77 amino acid residues are 12.9%. The full identity of the amino acid residues in the enzyme active center (histidine 374, 378, 505; glutamic acid 375, 402, 406; and tyrosine 515) suggests a complete overlap in the functional aspects. In other words, rabbit ACE2 should work with the same substrates as human ACE2. However, the high compositional identity does not mean that

the two chains are folded identically, i.e., it cannot be said in advance that their 3D structures are analogous.

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|
| h | M | S | S | S | S | W | L | L | L | S | L | V | A | V | T | A | A | Q | S | T | I | E | E | Q | A | K | T | F | L | D | K | F | N | H | E | A | E | D | L | F | 1  |
| r | M | S | G | S | S | W | L | L | L | S | L | V | A | V | T | A | A | Q | S | T | I | E | E | L | A | K | T | F | L | E | K | F | N | Q | E | A | E | D | L | S | 40 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|
| h | Y | Q | S | S | L | A | S | W | N | Y | N | T | N | I | T | E | E | N | V | Q | N | M | N | N | A | G | D | K | W | S | A | F | L | K | E | Q | S | T | L | A | 41 |
| r | Y | Q | S | A | L | A | S | W | D | Y | N | T | N | I | T | E | E | N | V | Q | K | M | N | D | A | E | A | K | W | S | A | F | Y | E | E | Q | S | K | L | A | 80 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | Q | M | Y | P | L | Q | E | I | Q | N | L | T | V | K | L | Q | L | Q | A | L | Q | Q | N | G | S | S | V | L | S | E | D | K | S | K | R | L | N | T | I | L | 81  |
| r | K | T | Y | P | S | Q | E | V | Q | N | L | T | V | K | R | Q | L | Q | A | L | Q | Q | S | G | S | S | A | L | S | A | D | K | S | K | Q | L | N | T | I | L | 120 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | N | T | M | S | T | I | Y | S | T | G | K | V | C | N | P | D | N | P | Q | E | C | L | L | L | E | P | G | L | N | E | I | M | A | N | S | L | D | Y | N | E | 121 |
| r | S | T | M | S | T | I | Y | S | T | G | K | V | C | N | Q | S | N | P | Q | E | C | F | L | L | E | P | G | L | D | E | I | M | A | K | S | T | D | Y | N | E | 160 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | R | L | W | A | W | E | S | W | R | S | E | V | G | K | Q | L | R | P | L | Y | E | E | Y | V | V | L | K | N | E | M | A | R | A | N | H | Y | E | D | Y | G | 161 |
| r | R | L | W | A | W | E | G | W | R | S | V | V | G | K | Q | L | R | P | L | Y | E | E | Y | V | V | L | K | N | E | M | A | R | A | N | N | Y | E | D | Y | G | 200 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | D | Y | W | R | G | D | Y | E | V | N | G | V | D | G | Y | D | Y | S | R | G | Q | L | I | E | D | V | E | H | T | F | E | E | I | K | P | L | Y | E | H | L | 201 |
| r | D | Y | W | R | A | D | Y | E | A | E | G | A | D | G | Y | D | Y | S | R | S | Q | L | I | D | D | V | E | R | T | F | S | E | I | K | P | L | Y | E | Q | L | 240 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | H | A | Y | V | R | A | K | L | M | N | A | Y | P | S | Y | I | S | P | I | G | C | L | P | A | H | L | L | G | D | M | W | G | R | F | W | T | N | L | Y | S | 241 |
| r | H | A | F | V | R | T | K | L | M | D | A | Y | P | S | R | I | S | P | T | G | C | L | P | A | H | L | L | G | D | M | W | G | R | F | W | T | N | L | Y | S | 280 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | L | T | V | P | F | G | Q | K | P | N | I | D | V | T | D | A | M | V | D | Q | A | W | D | A | Q | R | I | F | K | E | A | E | K | F | F | V | S | V | G | L | 281 |
| r | L | T | V | P | F | G | Q | K | P | N | I | D | V | T | D | T | M | V | N | Q | G | W | D | A | E | R | I | F | K | E | A | E | K | F | F | V | S | V | G | L | 320 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | P | N | M | T | Q | G | F | W | E | N | S | M | L | T | D | P | G | N | V | Q | K | A | V | C | H | P | T | A | W | D | L | G | K | G | D | F | R | I | L | M | 321 |
| r | P | S | M | T | Q | G | F | W | E | N | S | M | L | T | E | P | G | D | G | R | K | V | V | C | H | P | T | A | W | D | L | G | K | G | D | F | R | I | K | M | 360 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | C | T | K | V | T | M | D | D | F | L | T | A | H | H | E | M | G | H | I | Q | Y | D | M | A | Y | A | A | Q | P | F | L | L | R | N | G | A | N | E | G | F | 361 |
| r | C | T | K | V | T | M | D | N | F | L | T | A | H | H | E | M | G | H | I | Q | Y | D | M | A | Y | A | T | Q | P | F | L | L | R | N | G | A | N | E | G | F | 400 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | H | E | A | V | G | E | I | M | S | L | S | A | A | T | P | K | H | L | K | S | I | G | L | L | S | P | D | F | Q | E | D | N | E | T | E | I | N | F | L | L | 401 |
| r | H | E | A | V | G | E | I | M | S | L | S | A | A | T | P | E | H | L | K | S | I | G | L | L | P | Y | D | F | H | E | D | N | E | T | E | I | N | F | L | L | 440 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | K | Q | A | L | T | I | V | G | T | L | P | F | T | Y | M | L | E | K | W | R | W | M | V | F | K | G | E | I | P | K | D | Q | W | M | K | K | W | W | E | M | 441 |
| r | K | Q | A | L | T | I | V | G | T | L | P | F | T | Y | M | L | E | K | W | R | W | M | V | F | K | G | E | I | P | K | E | Q | W | M | Q | K | W | W | E | M | 480 |

Figure 2. Cont.

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | K | R | E | I | V | G | V | V | E | P | V | P | H | D | E | T | Y | C | D | P | A | S | L | F | H | V | S | N | D | Y | S | F | I | R | Y | Y | T | R | T | L | 481 |
| r | K | R | E | I | V | G | V | V | E | P | M | P | H | D | E | T | Y | C | D | P | A | A | L | F | H | V | A | N | D | Y | S | F | I | R | Y | Y | T | R | T | I | 520 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | Y | Q | F | Q | F | Q | E | A | L | C | Q | A | A | K | H | E | G | P | L | H | K | C | D | I | S | N | S | T | E | A | G | Q | K | L | F | N | M | L | R | L | 521 |
| r | Y | Q | F | Q | F | Q | E | A | L | C | Q | A | A | Q | H | E | G | P | L | H | K | C | D | I | S | N | S | T | E | A | G | Q | K | L | L | N | M | L | R | L | 560 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | G | K | S | E | P | W | T | L | A | L | E | N | V | V | G | A | K | N | M | N | V | R | P | L | L | N | Y | F | E | P | L | F | T | W | L | K | D | Q | N | K | 561 |
| r | G | R | S | E | P | W | T | L | A | L | E | N | V | V | G | A | K | N | M | D | V | R | P | L | L | N | Y | F | E | P | L | F | T | W | L | K | E | Q | N | R | 600 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | N | S | F | V | G | W | S | T | D | W | S | P | Y | A | D | Q | S | I | K | V | R | I | S | L | K | S | A | L | G | D | K | A | Y | E | W | N | D | N | E | M | 601 |
| r | N | S | F | V | G | W | S | T | E | W | T | P | Y | A | D | Q | S | I | K | V | R | I | S | L | K | T | A | L | G | D | Q | A | Y | E | W | N | D | S | E | M | 640 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | Y | L | F | R | S | S | V | A | Y | A | M | R | Q | Y | F | L | K | V | K | N | Q | M | I | L | F | G | E | E | D | V | R | V | A | N | L | K | P | R | I | S | 641 |
| r | Y | L | F | R | S | S | V | A | Y | A | M | R | K | Y | F | S | E | V | K | N | Q | T | I | L | F | G | E | E | D | V | R | V | S | D | L | K | P | R | I | S | 680 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | F | N | F | F | V | T | A | P | K | N | V | S | D | I | I | P | R | T | E | V | E | K | A | I | R | M | S | R | S | R | I | N | D | A | F | R | L | N | D | N | 681 |
| r | F | N | F | F | V | T | A | P | N | N | V | N | D | I | I | P | R | N | E | V | E | E | A | I | S | M | S | R | S | R | I | N | D | I | F | R | L | D | D | N | 720 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | S | L | E | F | L | G | I | Q | P | T | L | G | P | P | N | Q | P | P | V | S | I | W | L | I | V | F | G | V | V | M | G | V | I | V | V | G | I | V | I | L | 721 |
| r | S | L | E | F | V | G | I | Q | P | T | L | E | P | P | Y | E | S | P | V | P | I | W | L | V | V | F | G | V | V | M | G | M | I | V | I | G | I | V | V | L | 760 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | I | F | T | G | I | R | D | R | K | K | K | N | K | A | R | S | G | E | N | P | Y | A | S | I | D | I | S | K | G | E | N | N | P | G | F | Q | N | T | D | D | 761 |
| r | I | F | T | G | I | K | D | R | R | K | Q | K | Q | A | K | R | E | E | N | P | Y | G | F | V | D | M | S | K | G | E | N | N | S | G | F | Q | N | S | D | D | 800 |

|   |   |   |   |   |   |     |
|---|---|---|---|---|---|-----|
| h | V | Q | T | S | F | 801 |
| r | I | Q | T | S | F | 805 |

**Figure 2.** Primary amino acid sequence of the polypeptide chains of human (the rows beginning with **h**) and rabbit (**r**) ACE2. The numbers at the end of every row indicate the first and the last amino acid residue on the corresponding row. The amino acid residues are denoted according to the standard one-letter code. The cells of the different residues are colored according to the electric charge at pH 7.0 and the hydrophilicity of the given residue: red (negatively charged hydrophilic), blue (positively charged hydrophilic), green (uncharged hydrophilic), and yellow (uncharged hydrophobic). The polypeptide chain is divided (indicated by vertical lines) into three segments: signal peptide (amino acid residues 1–18, colored in bright orange), catalytic domain (19–615, colored in bright gray), and transmembrane segment (616–805 colored in bright purple). The five residues included in the zinc-binding motif HEMGH of the active center of the enzyme are denoted by the red rectangle.

To create a 3D model of the rabbit ACE2 protein, we applied computer mutational analysis, in which the different amino acid residues are sequentially substituted in the catalytic domain of human ACE2, and after each substitution, the new 3D coordinates of all atoms in the protein globule are calculated. The 3D models of the catalytic domains of the X-ray crystallographic hACE2 (PDB: 1r42) and the computer reconstructed rACE2 are shown in Figure 3 (the upper two models), colored in red (human) and green (rabbit). To compare the 3D structure of the rabbit (reconstructed) and human (taken from PDB) catalytic domains of ACE2, we used a computer program to align the 3D atomic coordinates of the two protein globules. The band model of ACE2 (Figure 3, the bottom left model)

shows that the 3D structure of the two proteins is almost identical, there are only two segments where the tertiary structure differs, but this only applies to the unstructured segments of the chain, whereas the  $\alpha$ -helical segments are identical. The bottom right model in Figure 3 shows the surface of the two aligned globules. Each atom is partially colored in red and green (the ratio of the two colors depends on the viewing angle), but the contours of the majority match, indicating that the coordinates of the unsubstituted atoms are virtually identical. The atoms of the different (substituted in rACE2) amino acid residues are colored entirely in red or green. Thus, the alignment (Figure 3, bottom two models) shows that the structure of the rabbit ACE2 is nearly identical to the human one both in the volume of the protein globule and on its surface.

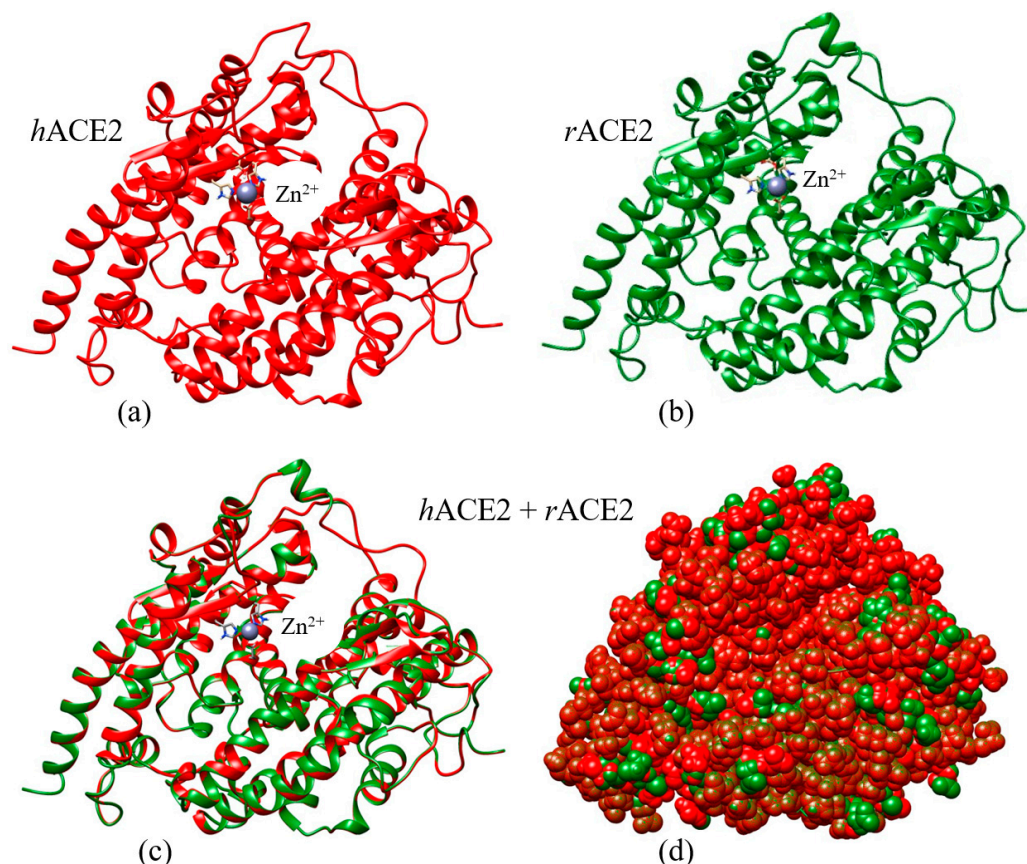


**Figure 3.** (a–d) 3D structure models of the human (red, hACE2) and rabbit (green, rACE2) catalytic domains of the angiotensin-converting enzyme-2 (the upper two models (a,b)); the  $\alpha$ -helix segments are depicted as ribbon spirals. The model of the rabbit ACE2 (b) is reconstructed by replacement of the different amino acid residues in the hACE2 (PDB: 1r42) model (a). The violet spherical object ( $Zn^{2+}$ ) is the zinc atom in the enzyme active center. The low two models (c,d) present the alignment (hACE2+rACE2) of the human and rabbit ACE2 in two different projections: the 3D volume of the protein globules (the ribbon model on the left) (c) and the 2D surface of the globules (the atomic model on the right) (d). The right bottom model (d) presents the atoms exposed on the surface of the aligned two protein globules; the atoms are colored fully in red (hACE2) or green (rACE2) when they are entirely protruded above the others, or in mixed color when their coordinates partially coincide. The brightness, shade, and color nuance of the atomic images give the impression for a quasi 3D surface of the protein globules.

## 2.2. Electric Charge

The electrical properties of the proteins are determined by the ability of 7 of the 20 amino acid residues to associate or dissociate a proton  $H^+$  in aqueous medium, whereby they acquire a positive or negative coulombic electrical charge, respectively (the partial

charges due to polarization of covalent bonds do not contribute significantly, being dipoles whose electric field acts over a very small distance). The ionizable groups have different affinity for the protons, so they join or detach a proton at different concentrations of  $H^+$  in the medium. This dependence on the proton concentration in the medium results in a pH-dependent  $nz$  (pH) of the net charge  $nz$  (the difference between the positive and negative coulombic charges). At a strongly acidic pH, the carboxyl groups are in an undissociated form ( $-COOH$ ) and the charge is determined only by the proton-associating N-containing groups: mainly the amino group ( $-NH_3^+$ ) of lysine, the imidazole group of histidine, and the guanidine group of arginine. At a strongly alkaline pH, only the dissociated carboxyl groups ( $-COO^-$ ) have a charge, since the amino groups ( $-NH_2$ ) have lost the associated proton. At a neutral pH, the net charge is determined by the difference between the positively and negatively charged groups. Only the ionizable groups located on the surface of the protein globule have the ability to associate or dissociate a proton because only they are in contact with the aqueous medium where the proton is associated with a water molecule, forming a hydroxonium ion  $H_3O^+$  which cannot penetrate into the interior of the globule. The local pH near the globule surface is different from the pH within the solution bulk because the surface electrostatic field with a negative or positive sign attracts or repulses protons, respectively. Therefore, each of the ionizable groups has an individual  $pK_a$ , i.e., 50% of the time, it is charged or uncharged at a pH value that is different from the average for these kinds of groups. This means that the net charge of the protein globule is different from that of the polypeptide chain in its unfolded form (random coil conformation), but this difference is manifested in the pH regions where the association or dissociation of protons takes place: a moderate low pH for acidic groups and high pH for basic ones.

Taking the above considerations into account, we calculated the pH dependence of the net charge  $nz$  in the 3D conformation of the catalytic domain, using the crystallographic structure of hACE2 and the reconstructed rACE2, and accounting for the two positive charges of the  $Zn^{2+}$  atom in the active center of the enzyme (Figure 3). The pH-dependent  $nz$  (pH) have two steep sections where the net charge changes, being positive at an acidic pH and negative at an alkaline pH (Figure 4).

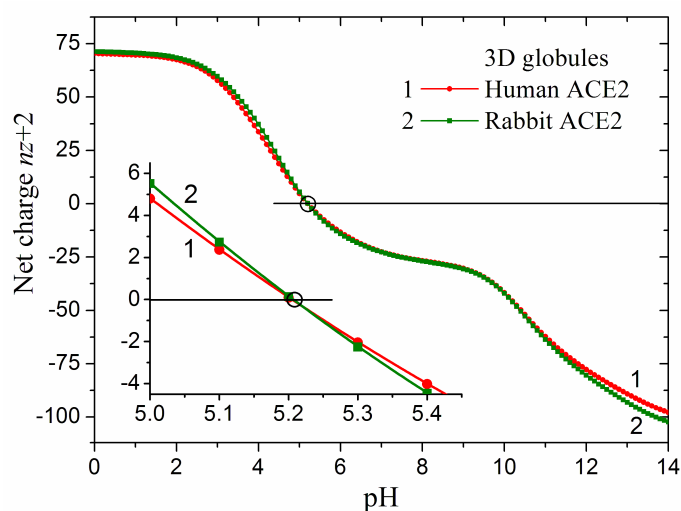


**Figure 4.** pH dependences of the net electric charge  $nz + 2$  of the globular catalytic domain of human (PDB: 1r42, hACE2, red curve 1) and reconstructed rabbit (rACE2, green curve 2) angiotensin-converting enzyme-2 in 3D conformation of the polypeptide chain in aqueous medium. The net charge of the two globular domains is the algebraic sum of the positive and negative coulombic charges of the polypeptide chain with the addition of two positive charges of the  $Zn^{2+}$  atom in the catalytic center. *Insert:* pH dependence of hACE2 (red curve 1) and rACE2 (green curve 2) with denoted isoelectric point ( $nz = 0$ ): pI 5.21 (human) and pI 5.21 (rabbit) ACE2.

The catalytic domain of human ACE2 has 70 positive charges at pH 1 and 98 negative at pH 14, while the rabbit domain has 71 positive and 102 negative charges, respectively. I.e., rACE2 has one positive charge and four negative charges more than hACE2 at the two extremal pH values. However, the difference in the charge number does not shift the isoelectric point, whose value is pI 5.21 for the catalytic domain of both human and rabbit ACE2 in 3D conformation (the insert in Figure 4). At a physiological pH, the number of charges is significantly less because many of the chargeable groups, including the phenol hydroxyl group of tyrosine residues, remain neutral. At pH 7.4, the net charge of the two catalytic domains in 3D conformation is almost equal: 24.6 and 24.7 negative charges (the fractional numbers reflect the different degree of protonation over time). The above numbers of the coulombic charges are actual when the chain retains its 3D conformation (folded state), but at the extremal pH values, the protein is denaturated. In unfolded (random coil) conformation, the number of the charges is bigger because then every chargeable group, depending on its  $pK_a$  value and pH of the aqueous medium, can bear coulombic charge: positive for the proton-binding groups or negative for the proton-dissociating groups (Section 3.3).

### 2.3. Electrostatic Potential

The electric charges of the protein globule create an electrostatic field that is non-uniform due to the irregular distribution of the charges. The sign and value of the local electrostatic potential on the surface of the globule determine the interaction of ACE2 with peptides and other proteins. In Figure 5, one can see the electrostatic potential being unequal in sign and value on the part of the ACE2 surface where the enzyme active center is located. At pH 5.0, this surface is predominantly positively charged, and at pH 7.4, it is almost entirely negatively charged, except for the active center (the blue atoms in the center). This difference should influence the enzyme activity (Section 3.6).

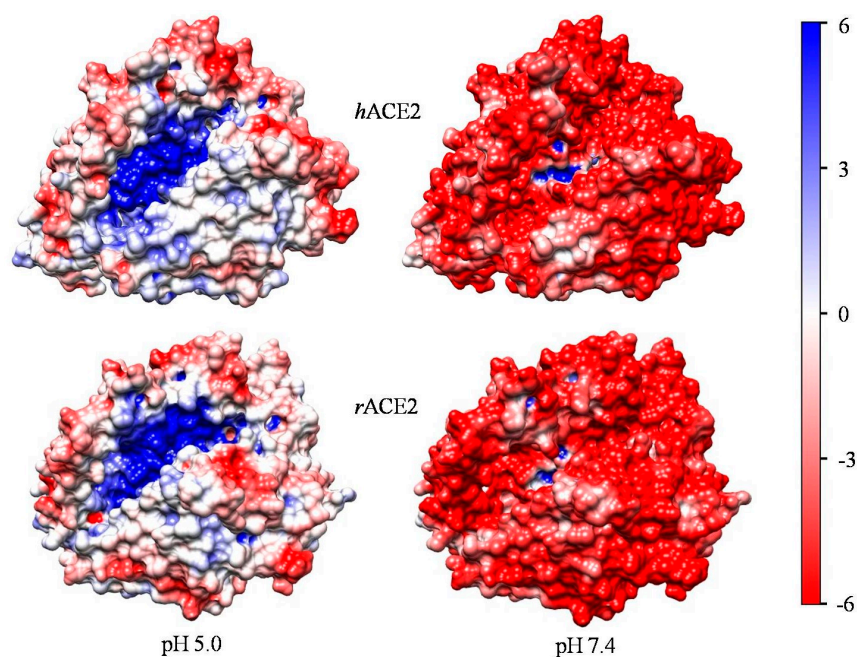


**Figure 5.** Electrostatic potential on the 3D surface of the catalytic domain of human (hACE2, two upper models, PDB: 1r42) and reconstructed rabbit (rACE2, two lower models) of angiotensin-converting enzyme-2 at pH 5.0 (two left models) and at pH 7.0 (two right models). The surfaces of the models are colored according to the electrostatic potential (negative—red, positive—blue), computed at pH 5.0 or pH 7.0, ionic strength 0.0001 mol/L, and temperature 20 °C, and visualized in the range  $kT/e = \pm 6$  J/C (the scale on the right);  $1\text{ }kT/e = 25.26\text{ mV}$  at 20 °C or  $26.73\text{ mV}$  at 37 °C.

2.4. Thermodynamic Stability

The change in the composition and sequence of the polypeptide chain affects the thermodynamic stability of the 3D structure of each globular protein. Any replacement of one amino acid by another can be considered as a point mutation that alters the stability of the globule because of the different hydrophilicity/hydrophobicity and the electric charge of the substituted and substituting amino acid residues [21]. Figure 6 shows the hydrophilic and hydrophobic residues of the entire polypeptide chain of the two enzymes hACE2 and rACE2, including the signal peptide, catalytic domain, and transmembrane segment, which sequentially fold during intracellular synthesis of the polypeptide chain. It can be seen that the number of substitutions of hydrophilic to hydrophobic amino acid residues and vice versa is not large, which means that the difference in primary structure (composition and sequence) cannot dramatically change the thermodynamic stability of the 3D structure of the two proteins.

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|
| h | M | S | S | S | S | W | L | L | L | S | L | V | A | V | T | A | A | Q | S | T | I | E | E | Q | A | K | T | F | L | D | K | F | N | H | E | A | E | D | L | F | 1  |
| r | M | S | G | S | S | W | L | L | L | S | L | V | A | V | T | A | A | Q | S | T | I | E | E | L | A | K | T | F | L | E | K | F | N | Q | E | A | E | D | L | S | 40 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|
| h | Y | Q | S | S | L | A | S | W | N | Y | N | T | N | I | T | E | E | N | V | Q | N | M | N | N | A | G | D | K | W | S | A | F | L | K | E | Q | S | T | L | A | 41 |
| r | Y | Q | S | A | L | A | S | W | D | Y | N | T | N | I | T | E | E | N | V | Q | K | M | N | D | A | E | A | K | W | S | A | F | Y | E | E | Q | S | K | L | A | 80 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | Q | M | Y | P | L | Q | E | I | Q | N | L | T | V | K | L | Q | L | Q | A | L | Q | Q | N | G | S | S | V | L | S | E | D | K | S | K | R | L | N | T | I | L | 81  |
| r | K | T | Y | P | S | Q | E | V | Q | N | L | T | V | K | R | Q | L | Q | A | L | Q | Q | S | G | S | S | A | L | S | A | D | K | S | K | Q | L | N | T | I | L | 120 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | N | T | M | S | T | I | Y | S | T | G | K | V | C | N | P | D | N | P | Q | E | C | L | L | L | E | P | G | L | N | E | I | M | A | N | S | L | D | Y | N | E | 121 |
| r | S | T | M | S | T | I | Y | S | T | G | K | V | C | N | Q | S | N | P | Q | E | C | F | L | L | E | P | G | L | D | E | I | M | A | K | S | T | D | Y | N | E | 160 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | R | L | W | A | W | E | S | W | R | S | E | V | G | K | Q | L | R | P | L | Y | E | E | Y | V | V | L | K | N | E | M | A | R | A | N | H | Y | E | D | Y | G | 161 |
| r | R | L | W | A | W | E | G | W | R | S | V | V | G | K | Q | L | R | P | L | Y | E | E | Y | V | V | L | K | N | E | M | A | R | A | N | N | Y | E | D | Y | G | 200 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | D | Y | W | R | G | D | Y | E | V | N | G | V | D | G | Y | D | Y | S | R | G | Q | L | I | E | D | V | E | H | T | F | E | E | I | K | P | L | Y | E | H | L | 201 |
| r | D | Y | W | R | A | D | Y | E | A | E | G | A | D | G | Y | D | Y | S | R | S | Q | L | I | D | D | V | E | R | T | F | S | E | I | K | P | L | Y | E | Q | L | 240 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | H | A | Y | V | R | A | K | L | M | N | A | Y | P | S | Y | I | S | P | I | G | C | L | P | A | H | L | L | G | D | M | W | G | R | F | W | T | N | L | Y | S | 241 |
| r | H | A | F | V | R | T | K | L | M | D | A | Y | P | S | R | I | S | P | T | G | C | L | P | A | H | L | L | G | D | M | W | G | R | F | W | T | N | L | Y | S | 280 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | L | I | V | P | F | G | Q | K | P | N | I | D | V | T | D | A | M | V | D | Q | A | W | D | A | Q | R | I | F | K | E | A | E | K | F | F | V | S | V | G | L | 281 |
| r | L | T | V | P | F | G | Q | K | P | N | I | D | V | T | D | T | M | V | N | Q | G | W | D | A | E | R | I | F | K | E | A | E | K | F | F | V | S | V | G | L | 320 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | P | N | M | T | Q | G | F | W | E | N | S | M | L | T | D | P | G | N | V | Q | K | A | V | C | H | P | T | A | W | D | L | G | K | G | D | F | R | I | L | M | 321 |
| r | P | S | M | T | Q | G | F | W | E | N | S | M | L | T | E | P | G | D | G | R | K | V | V | C | H | P | T | A | W | D | L | G | K | G | D | F | R | I | K | M | 360 |

Figure 6. Cont.

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | C | T | K | V | T | M | D | D | F | L | T | A | H | H | E | M | G | H | I | Q | Y | D | M | A | Y | A | A | Q | P | F | L | L | R | N | G | A | N | E | G | F | 361 |
| r | C | T | K | V | T | M | D | N | F | L | T | A | H | H | E | M | G | H | I | Q | Y | D | M | A | Y | A | T | Q | P | F | L | L | R | N | G | A | N | E | G | F | 400 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | H | E | A | V | G | E | I | M | S | L | S | A | A | T | P | K | H | L | K | S | I | G | L | L | S | P | D | F | Q | E | D | N | E | T | E | I | N | F | L | L | 401 |
| r | H | E | A | V | G | E | I | M | S | L | S | A | A | T | P | E | H | L | K | S | I | G | L | L | P | Y | D | F | H | E | D | N | E | T | E | I | N | F | L | L | 440 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | K | Q | A | L | T | I | V | G | T | L | P | F | T | Y | M | L | E | K | W | R | W | M | V | F | K | G | E | I | P | K | D | Q | W | M | K | K | W | W | E | M | 441 |
| r | K | Q | A | L | T | I | V | G | T | L | P | F | T | Y | M | L | E | K | W | R | W | M | V | F | K | G | E | I | P | K | E | Q | W | M | Q | K | W | W | E | M | 480 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | K | R | E | I | V | G | V | V | E | P | V | P | H | D | E | T | Y | C | D | P | A | S | L | F | H | V | S | N | D | Y | S | F | I | R | Y | Y | T | R | T | L | 481 |
| r | K | R | E | I | V | G | V | V | E | P | M | P | H | D | E | T | Y | C | D | P | A | A | L | F | H | V | A | N | D | Y | S | F | I | R | Y | Y | T | R | T | I | 520 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | Y | Q | F | Q | F | Q | E | A | L | C | Q | A | A | K | H | E | G | P | L | H | K | C | D | I | S | N | S | T | E | A | G | Q | K | L | F | N | M | L | R | L | 521 |
| r | Y | Q | F | Q | F | Q | E | A | L | C | Q | A | A | Q | H | E | G | P | L | H | K | C | D | I | S | N | S | T | E | A | G | Q | K | L | L | N | M | L | R | L | 560 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | G | K | S | E | P | W | T | L | A | L | E | N | V | V | G | A | K | N | M | N | V | R | P | L | L | N | Y | F | E | P | L | F | T | W | L | K | D | Q | N | K | 561 |
| r | G | R | S | E | P | W | T | L | A | L | E | N | V | V | G | A | K | N | M | D | V | R | P | L | L | N | Y | F | E | P | L | F | T | W | L | K | E | Q | N | R | 600 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | N | S | F | V | G | W | S | T | D | W | S | P | Y | A | D | Q | S | I | K | V | R | I | S | L | K | S | A | L | G | D | K | A | Y | E | W | N | D | N | E | M | 601 |
| r | N | S | F | V | G | W | S | T | E | W | T | P | Y | A | D | Q | S | I | K | V | R | I | S | L | K | T | A | L | G | D | Q | A | Y | E | W | N | D | S | E | M | 640 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | Y | L | F | R | S | S | V | A | Y | A | M | R | Q | Y | F | L | K | V | K | N | Q | M | I | L | F | G | E | E | D | V | R | V | A | N | L | K | P | R | I | S | 641 |
| r | Y | L | F | R | S | S | V | A | Y | A | M | R | K | Y | F | S | E | V | K | N | Q | T | I | L | F | G | E | E | D | V | R | V | S | D | L | K | P | R | I | S | 680 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | F | N | F | F | V | T | A | P | K | N | V | S | D | I | I | P | R | T | E | V | E | K | A | I | R | M | S | R | S | R | I | N | D | A | F | R | L | N | D | N | 681 |
| r | F | N | F | F | V | T | A | P | N | N | V | N | D | I | I | P | R | N | E | V | E | E | A | I | S | M | S | R | S | R | I | N | D | I | F | R | L | D | D | N | 720 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | S | L | E | F | L | G | I | Q | P | T | L | G | P | P | N | Q | P | P | V | S | I | W | L | I | V | F | G | V | V | M | G | V | I | V | V | G | I | V | I | L | 721 |
| r | S | L | E | F | V | G | I | Q | P | T | L | E | P | P | Y | E | S | P | V | P | I | W | L | V | V | F | G | V | V | M | G | M | I | V | I | G | I | V | V | L | 760 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| h | I | F | T | G | I | R | D | R | K | K | K | N | K | A | R | S | G | E | N | P | Y | A | S | I | D | I | S | K | G | E | N | N | P | G | F | Q | N | T | D | D | 761 |
| r | I | F | T | G | I | K | D | R | R | K | Q | K | Q | A | K | R | E | E | N | P | Y | G | F | V | D | M | S | K | G | E | N | N | S | G | F | Q | N | S | D | D | 800 |

|   |   |   |   |   |   |     |
|---|---|---|---|---|---|-----|
| h | V | Q | T | S | F | 801 |
| r | I | Q | T | S | F | 805 |

**Figure 6.** Amino acid sequence of the polypeptide chains of human (rows beginning with **h**) and rabbit (**r**) ACE2. The numbers at the end of every row indicate the first and the last amino acid residue on the corresponding row. The cells of the amino acid residues (denoted by the standard one-letter code) are colored according to their affinity to the water molecules: hydrophilic (green) or hydrophobic (yellow). The vertical lines and the red rectangle denote the beginning and the end of the catalytic domain and the amino acid residues from the zinc-binding motif included in the enzyme active center, respectively.

A quantitative measure of the thermodynamic stability of the 3D structure of the protein globule is the folding energy  $\Delta G_{\text{fold}}$ , which is determined by the Gibbs free energy alteration at the transition from the fully unfolded state (random coil) to the native 3D structure of the protein globule at a given temperature, pH, and solution composition. Negative values of  $\Delta G_{\text{fold}}$  mean that the free energy decreases at the transition; i.e., the 3D structure is thermodynamically advantageous. The folding energy  $\Delta G_{\text{fold}}$  is determined by the contribution of the four types of forces with which the amino acid residues interact with each other and with the water molecules of the medium. The components of  $\Delta G_{\text{fold}}$  determined by the energies of the hydration, hydrogen bonds, and Van der Waals forces are independent of the pH of the medium, but the electrostatic  $\Delta G_{\text{el}}$  component is pH-dependent, since it is determined by the degree of  $\text{H}^+$  association or dissociation of the ionizable groups. Therefore, the pH dependence of  $G_{\text{el}}$  indicates the pH-range of maximal stability of the 3D structure.

Since the number and location of the ionizable groups are determined by the composition and sequence of the amino acid residues in the polypeptide chain, the difference in the primary structure of the human and rabbit ACE2 (Figure 2) means that their thermodynamic 3D stability should be different. Indeed, the pH-dependent  $\Delta G_{\text{el}}(\text{pH})$  of the electrostatic component  $\Delta G_{\text{el}}$  of the folding energy of the catalytic domain of hACE2 and rACE2 are quite different by the absolute values but have almost analogous course (Figure 7). Considering that the isoelectric points pI 5.2 coincide and the net charge is almost equal in the range pH 6–9 (Section 2.2, Figure 4), it can be inferred that the difference in the  $\Delta G_{\text{el}}$  values reflects the different 3D location of the coulomb charges in the two protein globules. The difference in the absolute values of  $\Delta G_{\text{el}}$  (Figure 7) demonstrates the sensitivity of the electrostatic component of the folding energy to the 3D structure of the protein globules (Section 3.2). The negative values of  $\Delta G_{\text{el}}$  mean that the electrostatic energy stabilizes the 3D structure of the hACE2 domain in the pH range 5.5–9.8. In the case of rACE2, this range is slightly wider: pH 5.1–10.0. Both enzymes are most stable at pH 6.5 (the deeper minimum), where the folding energy is maximal by absolute value.

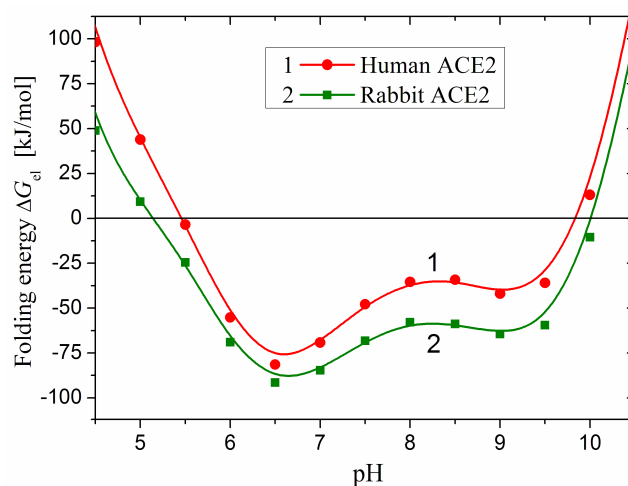


**Figure 7.** pH dependences of the electrostatic component  $\Delta G_{\text{el}}$  of the folding energy  $\Delta G_{\text{fold}}$  of the human (hACE2, PDB: 1r42, curve 1) and reconstructed rabbit (rACE2, curve 2) polypeptide chains of angiotensin-converting enzyme-2 at the transformation of the polypeptide chain from fully unfolded (random coil) to folded (globular 3D structure) conformation. The two 3D models are optimized by the program YASARA.

The positive values of  $\Delta G_{\text{el}}$  at pH < 5.5 and pH > 9.8 for hACE2 and also at pH < 5.1 and pH > 10.0 for rACE2 mean that in these extreme pH ranges, the 3D stability of the protein globule is reduced, since electrostatic repulsion between predominantly positive (at acidic pH) or negative charges (at alkaline pH) dominates there. However, the positive values of  $\Delta G_{\text{el}}$  does not imply that the polypeptide chain necessarily transforms to the

unfolded state (random coil) because the main contribution (about 90%) in the folding energy  $\Delta G_{\text{fold}}$  gives its hydration component, due to which the aqueous medium pushes the hydrophobic amino acid residues into the core of the protein globule, while the hydrophilic ones remain on its surface. The reduction of  $\Delta G_{\text{fold}}$  (at positive values of the electrostatic component  $\Delta G_{\text{el}}$ ) is rather weak due to the relatively small contribution of the electrostatic interactions in 3D conformation of the polypeptide chain. I.e., the predominant electrostatic repulsion (in the two extremal pH ranges where  $\Delta G_{\text{el}}$  has positive values) reduces the 3D structural stability of the protein globule and it can be denaturated in some degree, up to the state of melt globule, when the polypeptide chain retains its globular conformation but loses the secondary structure ( $\alpha$ -helix and  $\beta$ -sheets).

### 3. Discussion

#### 3.1. Collectrin Similarity

In the literature, the transmembrane domain of ACE2 is described as a collectrin-like [30]. To verify this claim, Figure 8 (top rows) shows a part of the hACE2 chain (including the transmembrane segment) as compared to the human collectrin chain (bottom rows). Taking into account the presence of deletions, the alignment was made to maximize the matching of amino acid residues in the two chains. Comparison of this part of the human ACE2 chain shows 82% similarity of the amino acid residues composition, but only 40% identity of their sequence with those of collectrin, which is not sufficient to claim the structural similarity of the two proteins. Because the hydrophobicity of the amino acid residues is particularly important, their cells in Figure 8 are colored yellow, in contrast to those of the hydrophilic residues (green colored). This distinction allows identification of the hydrophobic segment of the two chains (the long, yellow-stained segment). The comparison shows that this part of the hACE2 chain is almost identical in composition and amino acid sequence to the human collectrin [31,32]. This similarity suggests that the structure of the intramembrane segment of the two proteins is also similar, although for now its 3D conformation remains unknown. The other two parts have alternating short segments of mixed hydrophilicity, suggesting that they are extramembrane: the upper segment continues to the catalytic domain of hACE2 (extracellularly located), and the lower one is a small intracellular hydrophilic segment. Collectrin (also called Tmem27) is a transporter protein which plays an important role in neutral amino acid transport in the kidneys, interacting with proteins of the Slc6 family (in particular B<sup>0</sup>AT1, B<sup>0</sup>AT3) [33]. Like collectrin, ACE2 also participates in amino acid transport but in the small intestines, where it interacts with B<sup>0</sup>AT1 [34,35]. Considering the similarity not only in primary amino acid sequence, but also in functional aspect, the claim that the transmembrane segment is collectrin-like can be considered appropriate.

#### 3.2. Stability of 3D Structure

The stability of a three-dimensional (3D) protein globule structure is determined by four types of forces: hydrogen bonds, hydrophilic, electrostatic, and van der Waals. Intramolecular hydrogen bonds determine the secondary structure:  $\alpha$ -helix and  $\beta$ -sheet. The hydrophilicity of the amino acid residues is determined by their ability to form hydrogen bonds with water molecules. The Van der Waals forces are attractive, their main component being the London dispersion forces, while the partial charge interactions (dipole–dipole and dipole-induced dipole) play a minor role. The electrostatic forces are both attractive between the coulombic charges of the opposite sign (positive to negative) and repulsive when the charges have the same sign (positive or negative). The spatial folding of the polypeptide chain is determined primarily by the affinity of the amino acid residues for water molecules: hydrophilic residues remain on the surface of the protein globule, whereas hydrophobic residues are predominantly in the globule core. This distribution is thermodynamically determined: the free energy decreases when the polypeptide chain folds to form a globule with a hydrophobic core and a hydrophilic shell. This ensures a maximum number of contacts of water molecules from the medium with the electronegative atoms (O and N)

with which they form hydrogen bonds (HO–H...O– and HO–H...N–), and no contact in the hydrophobic residues with which hydrogen bond formation is impossible due to the absence of such atoms. The degree of hydrophilicity varies depending on the ratio of hydrophilic to hydrophobic (–CH<sub>2</sub>–) groups in the amino acid residues, which determines the sign and value of the free energy change upon transfer of an amino acid residue from an organic to aqueous solution [36].

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| A | Q | S | I | K | V | R | I | S | L | K | S | A | L | G | D | K | A | Y | E | W | N | D | N | E | M | Y | L | F | R | S | S | V | A | Y | A | M | R | Q | Y | F | 616 |
| C | F | R | S | S | V | A | Y | A | M | R | Q | A | L | G | D | R | A | Y | A | W | D | T | N | E | E | Y | L | F | K | A | M | V | A | F | S | M | R | K | V | P | 655 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| A | L | K | V | K | N | Q | M | I | L | F | G | E | E | D | V | R | V | A | N | L | K | P | R | I | S | F | N | F | F | V | T | A | P | K | N | V | S | D | I | I | 656 |
| C | N | R | E | A | T | E | I | – | – | – | – | – | S | H | V | L | L | C | N | V | T | Q | R | V | S | F | W | F | V | V | T | D | P | S | K | – | N | H | T | L | 695 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| A | P | R | T | E | V | E | K | A | I | R | M | S | R | S | R | I | N | D | A | F | R | L | N | D | N | S | L | E | F | L | G | I | Q | P | T | L | G | P | P | N | 696 |
| C | P | A | V | E | V | Q | S | A | I | R | M | N | K | N | R | I | N | N | A | F | F | L | N | D | Q | T | L | E | F | L | K | I | P | S | T | L | A | P | P | M | 735 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| A | Q | P | P | V | S | I | W | L | I | V | F | G | V | V | M | G | V | I | V | V | G | I | V | I | L | I | F | T | G | I | R | D | R | K | K | K | N | K | A | R | 736 |
| C | D | P | S | V | P | I | W | I | I | I | F | G | V | I | F | C | I | I | I | V | A | I | A | L | L | I | L | S | G | I | W | Q | R | R | R | K | N | K | E | P | 775 |

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| A | S | G | E | N | P | Y | A | S | – | – | – | – | – | I | D | I | S | K | G | – | E | N | N | P | G | F | Q | N | T | D | D | V | Q | T | S | F | 776 |
| C | S | E | V | D | D | A | E | D | K | C | E | N | M | I | T | I | E | N | G | I | P | S | D | P | L | D | M | K | G | G | H | I | N | D | A | F | 805 |

**Figure 8.** Amino acid sequence of the polypeptide chains of transmembrane collectrin-like segment of the human ACE2 (the rows denoted by A) and human collectrin (C). The numbers at the end of every row indicate the first and the last amino acid residue on the corresponding row (the numbering corresponds to that in the hACE2 shown in Figure 6). The cells of the amino acid residues (denoted by the standard one-letter code) are colored according to their affinity to water: hydrophilic (green) or hydrophobic (yellow). The absent amino acid residues are denoted by dashes.

In the Protein Data Bank (PDB), for each deposited 3D structure, usually there are several models obtained with different methods: X-ray crystallography of the protein crystal, nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM). Crystallographic models are the most reliable and have the highest resolution, but they have the disadvantage that in the crystal, the protein globules are tightly packed, and therefore unstructured segments of the polypeptide chain may have a different spatial arrangement in aqueous media. We choose the 3D protein models using as a criterion the pH-dependent  $\Delta G_{el}(\text{pH})$  of the electrostatic component  $\Delta G_{el}$  of the folding energy, which indicates how thermodynamically advantageous the 3D structure of the protein globule is in dilute aqueous solution when the influence of neighboring globules is negligible. These dependences are very sensitive to alterations in the 3D structure of the protein globule because  $\Delta G_{el}$  is the energy of a system of coulombic charges: the shift of every element causes alterations to its value.

Figure 9 shows the  $\Delta G_{el}(\text{pH})$  dependencies of the catalytic domain of rabbit ACE2: the reconstructed (on the base of the original 3D model PDB:1r42 of hACE2) and two optimized rACE2 models. The first model has entirely positive values of  $\Delta G_{el}$  (curve 1), suggesting that it is not the most thermodynamically advantageous 3D structure. To approximate the 3D structure of rACE2 to the real one in aqueous media, we optimized the model, whereby the unstructured segments receive atomic coordinates corresponding to a thermodynamically more favorable conformation with minimal free energy. The optimization performed with two computer programs discloses that better results (a deeper minimum of the pH-dependent  $\Delta G_{el}(\text{pH})$  of the electrostatic component  $\Delta G_{el}$  of the folding

energy) gives the YASARA program, in comparison with Chimera. The pH dependence of the optimized model (curve 3 in Figure 9) shows that the  $\Delta G_{el}$  values become negative in the range pH 5.1–10.0 with minimum at pH 6.5. Consequently, the 3D structure of the catalytic domain of rACE2 is the most thermodynamically stable in this pH range.

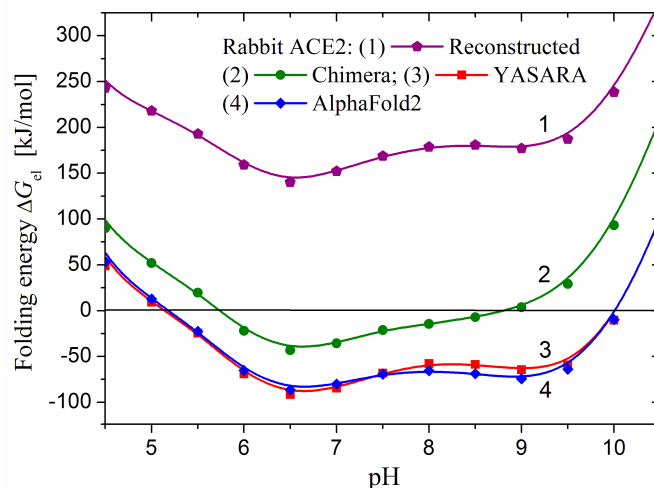


**Figure 9.** pH dependences of the electrostatic component  $\Delta G_{el}$  of the folding energy of the catalytic domain of rabbit ACE2 polypeptide chain at transformation from random coil to 3D structure for four models: reconstructed on the base of hACE2 (PDB: 1r42, curve 1) and optimized by Chimera (curve 2) or YASARA (curve 3), and created by AlphaFold2 on the base of amino acid sequence of rACE2 (curve 4).

The reconstruction of the rACE2 catalytic domain, based on hACE2, is analogous to that used for single point mutations in viral proteins [21]. Since the 3D structure of the catalytic domain of hACE2 is obtained using a recombinant protein synthesized considering the genetic code of only part of the polypeptide chain (indicated by vertical lines in Figure 2), the model of the catalytic domain probably does not include the entire off-membrane hydrophilic part of the enzyme. The models, shown in Figure 3, are colored in red and green according to the human and rabbit specific diet: red meat and green grass. The almost complete coincidence of the 3D structure of the catalytic domains of the human and rabbit ACE2 can be explained by the fact that the substituted amino acids have commensurate side group volumes, which gives them sufficient freedom to form an  $\alpha$ -helical secondary structure. The exception is proline, whose conformational angles are highly constrained and there the helical segments are broken.

Figure 9 presents the electrostatic components  $\Delta G_{el}$  of the folding energy calculated for 3D models of rACE2, which are reconstructed by two different approaches. The first one is based on an experimental 3D model of hACE2, which is modified by mutagenesis according to the primary amino acid sequence of rACE2 (curve 1), after which the model has been optimized to reach the most energetically advantageous atomic coordinates (curves 2 and 3). The second approach is based on the construction of a 3D model beginning directly from the primary amino acid sequence of rACE2, for which purpose we employed the program AlphaFold2. The practical coincidence of curves 3 and 4 shows that 3D models of rACE2, obtained by these two alternative computer techniques, have practically equal thermodynamic stability. The coincidence reflects almost the same coordinates of the coulomb electric charges of the groups ionized at a given pH. The alignment of the two 3D models of rACE2, which compares the atomic coordinates of two protein globules, shows that the root mean square deviation (RMSD) is 0.199 nm, and the template modeling score (TM-score) is 0.95. These values mean that the two rACE2 models practically coincide (values in the ranges  $\text{RMSD} \leq 0.2$  nm and TM-score 0.5–1.0 are designed as a good coincidence).

### 3.3. Electric Charge in Unfolded and 3D-Folded Conformation

The coulombic electric charges of the catalytic domain in the unfolded conformation of the polyelectrolyte chain (random coil, without  $\text{Zn}^{2+}$ ) are 69 and 70 positive at pH 1, or 111 and 115 negative at pH 14 for hACE2 and rACE2, respectively. These numbers correspond to all ionizable (chargeable) groups which can obtain coulombic charges as a result of protonation or deprotonation at a given pH. The amino acid residues of human ACE2 catalytic domain, which can bear positive charges, are 69, including 34 lysine, 18 arginine, 16 histidine, and N-end amino group; the negatively chargeable residues are 111 (33 aspartate, 47 glutamate, 28 tyrosine, two cystine (six cystine form disulfide bridges), and C-end carboxyl group). The chain of the rabbit ACE2 catalytic domain has 70 positively chargeable residues (33 lysine, 23 arginine, 13 histidine, and N-end amino group) and 115 negatively chargeable: 31 aspartate, 53 glutamate, 28 tyrosine, two cystine (six cystine form disulfide bridges), and C-end carboxyl group.

In 3D conformation, the positive charges are 68 and 69 (at pH 1), and the negative charges are 100 and 104 (at pH 14), respectively, for the catalytic domains of hACE2 and rACE2 without  $\text{Zn}^{2+}$  (Section 2.2). So, at these extreme pH values, the numbers of coulombic charges of the folded chain are less with one positive and 10 negative charges, as compared to those in the unfolded conformation for both hACE2 and rACE2. This difference can be explained by the following. Firstly, the number of the chargeable groups is calculated at extreme pH values, but at a physiological pH the number of the charged groups (bearing coulombic electric charges) is significantly less because of a lower degree of ionization, which is determined mainly by the dissociation constant  $\text{pK}_a$  of a given ionizable group (the specific  $\text{pK}_a$  value can obtain extreme values). The groups with a high degree of charging at pH 7.4 are the carboxylic groups of the asparagine and glutamine acid residues ( $\text{pK}_a$  3.0–4.7, negatively charged) and the positively charged amino group of lysine ( $\text{pK}_a$  9.4–10.6) and the guanidine group of arginine residues ( $\text{pK}_a$  11.6–12.6). The phenol group of tyrosine ( $\text{pK}_a$  9.8–10.4) is deprotonated at pH 14, but protonated (uncharged) at neutral pH. In addition, only about half of the imidazole groups of the histidine residues ( $\text{pK}_a$  5.6–7.0) are protonated; the rest are uncharged. Secondly, some of the dissociable groups catch a proton and become uncharged when the chain folds into a 3D globule. In the hydrophobic core of the protein globule, the dielectric permittivity  $\epsilon \approx 4$  is very low, and consequently the electrostatic attraction between  $\text{COO}^-$  and  $\text{H}^+$  is about 20 times stronger than those in the aqueous medium with  $\epsilon = 80$ . Therefore, the deeply located carboxylic groups usually are uncharged even at pH 14.

The disappearance of one positive charge (from 69 to 68 for hACE2, and from 70 to 69 for rACE2) at pH 1 when the polypeptide chain folds in the 3D globule is surprising because at such high proton concentration (pH 0 at 1 mol/L  $\text{H}_3\text{O}^+$ ), all positively chargeable groups, both located on the surface and in the hydrophobic core (where they more strongly retain the caught proton due to the low dielectric permittivity), must be protonated. We supposed that this disappearance can be explained by the emergence of one negative charge because of deprotonation of a surface-located carboxylic group having an extremely low dissociation constant  $\text{pK}_a < 1$ , instead the usual  $\text{pK}_a$  3.0–4.7. Such a low dissociation constant can appear as a result of a strong positive electrostatic potential around these groups in 3D conformation of the polypeptide chain. This leads to a very low concentration of  $\text{H}_3\text{O}^+$  cations in the vicinity of this carboxylic group, and therefore it loses its proton and becomes negatively charged even at  $\text{pH} \leq 1$ . We have found that the suspect is the carboxylic group of Glu-402 residue with calculated  $\text{pK}_a \approx 0$ , positioned in the surface pocket of the enzyme active center, where it is surrounded by positively charged groups. Due to its negative charge, this carboxylic group forms a coordinative bond with the  $\text{Zn}^{2+}$  atom when the polypeptide chain is in 3D conformation.

### 3.4. Isoelectric Point in Unfolded and 3D-Folded Conformation

In unfolded (random coil) conformation, all chargeable groups of the polypeptide chain that are in contact with the aqueous medium can bear a positive or negative coulomb-

bic charge due to proton association or dissociation in the dependence on pH and the dissociation constant  $pK_a$  (pH at 50% protonation). The intrinsic dissociation constant  $pK_a$  is then determined by the average degree of protonation for all groups of a given type. When the chain is folded into a 3D globule each chargeable group obtains an individual (apparent)  $pK_a$  value which can be defined by the equality of the times when the group is protonated or deprotonated. The difference between the intrinsic and apparent  $pK_a$  constants is due to the different local proton concentrations caused by the surface electrostatic potential in the vicinity of the group, which repels or attracts the  $H_3O^+$  cations dependent on its sign. I.e., the charged groups create an electrostatic potential which influences the degree of their protonation in an acidic pH range for carboxylic ( $H^+$ -dissociable) or in an alkaline pH range for amino ( $H^+$ -associable) groups. The isoelectric point  $pI$  (pH at zero net charge, equality of the positive and negative coulombic charges) of the polypeptide chain is different in its unfolded (random coil) and folded (3D globule) conformation because of two reasons: (a) some of the chargeable groups can be located not on the surface of the globule, but in its hydrophobic core; and (b) the degree of protonation depends on the local electrostatic potential. In the case of ACE2, a third reason appears: upon folding into the 3D structure, seven amino acid residues form an active center (Section 2.1), which bind coordinately a  $Zn^{2+}$  cation, and respectively, the net charge of the protein globule increases by two coulombic charges.

The different number of the chargeable amino acid residues of the catalytic domains of the human and rabbit ACE2 (Section 3.3) leads to different  $pI$  values in the unfolded conformation of their polypeptide chains:  $pI$  5.03 and  $pI$  4.97, respectively, for hACE2 and rACE2; the lower rabbit  $pI$  is due to the bigger number of negatively charged groups. At the folding of the chain into 3D conformation, the isoelectric points become closer:  $pI$  5.12 and  $pI$  5.13, and shift to higher pH values owing to the disappearance of 11 negative charges. When the  $Zn^{2+}$  cation is bound in the active center of the 3D globule, the isoelectric points coincide:  $pI$  5.21 (Figure 10).

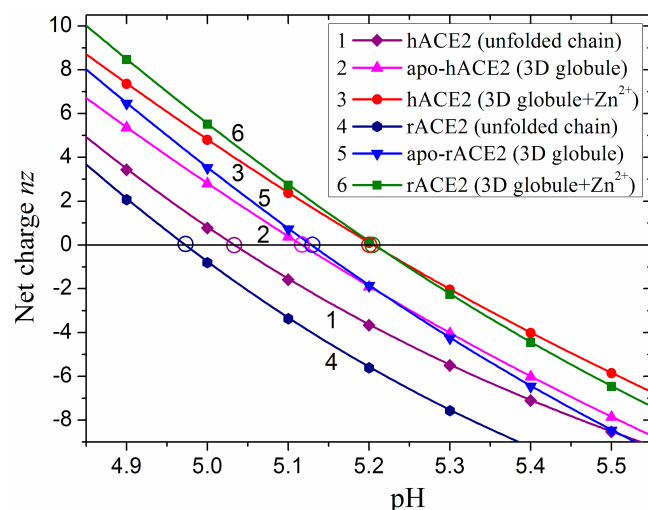


**Figure 10.** pH dependences of the net electric charge  $nz$  of the catalytic domain of human (hACE2, curves 1, 2, and 3) and rabbit (rACE2, curves 4, 5, and 6) angiotensin-converting enzyme-2 in aqueous medium when the polypeptide chain is in unfolded conformation (random coil, curves 1 and 4), folded in 3D globule without  $Zn^{2+}$  (curves 2 and 5) and when the  $Zn^{2+}$  cation is bound in the enzyme active center (curves 3 and 6). The folded 3D conformations correspond to the crystallographic model of hACE2 (PDB:1r42, red curve 3) and to the reconstructed model of rACE2 (green curve 6). The isoelectric points are denoted by open circles. The net charge  $nz$  is the algebraic sum of the positive and negative coulombic charges of the polypeptide chain without (curves 1, 2, 4, 5) or with the attached  $Zn^{2+}$  cation (curves 3 and 6).

On the contrary, at the unfolding of the polypeptide chain of the catalytic domain from 3D conformation to random coil (full denaturation), the isoelectric points shift to lower pH values with 0.17 and 0.23 pH units for human and rabbit ACE2, respectively. This phenomenon is caused by the disappearance of the surface electrostatic potential of the former ACE2 globule; consequently, the concentration of  $\text{H}_3\text{O}^+$  cations (local pH) becomes almost equal to that in the bulk of the protein solution and the apparent  $\text{pK}_a$ , especially those with extreme values, shifts to the intrinsic  $\text{pK}_a$  values of the chargeable groups.

### 3.5. Isoelectric Point: Calculated and Experimental

In our previous study [37,38], the experimentally determined isoelectric mean point pI 4.0 (two bands with pI 3.9 and pI 4.1 in isoelectric focusing gel) of the recombinant human ACE2-Fc (with Fc-tag attached to the C-end of the hACE2 polypeptide chain) was found to be two pH units lower than the calculated pI 5.99 (taking into account the charges of the Fc-tag). There are two reasons for this. The first is that the experiment was performed by isoelectric focusing in a medium with 8 mol/L urea, in which a complete denaturation occurs: the irreversible transition from the 3D structure state to the fully unfolded conformation (random coil). The influence of the polypeptide chain conformation on the isoelectric point (Section 3.4) is relatively small and quite insufficient to explain the difference of two pH units between the calculated pI 6.0 and experimentally measured pI 4.0 values of the isoelectric point.

The second reason for the two pH unit discrepancy that we found between the calculated and measured isoelectric points is that in the isoelectric focusing experiment, we have used recombinant hACE2 expressed in embryonic human cell culture with a high glucose concentration in the medium. In these conditions, an enzymatic glycosylation and non-enzymatic glycation occur (Section 3.8). Both processes shift the isoelectric point pI of proteins to lower pH values due to a reduction of the positively charged amino groups and emergence of additional negatively charged carboxylic groups.

### 3.6. Electrostatic Potential and Enzyme Activity

The surface electrostatic potential is dependent on the pH of the surrounding aqueous medium because of protonation or deprotonation of the chargeable groups of the amino acid residues located on the surface of the protein globule. The pH-dependent potential of ACE2 (Figure 5) influences its enzyme activity by strengthen or weakening the electrostatic interactions with the positively charged angiotensin peptides, since the enzymatic reaction is preceded by electrostatic attraction of them to, and adsorption onto, the negatively charged surface of the protein globule, as well as of the negatively charged C-terminal carboxyl group to the positive charges in the active center, where it forms a complex with the positively charged  $\text{Zn}^{2+}$  atom. The angiotensin I, II, and 1–9 (Figure 1) carry four, three, and four positive charges, respectively. The following suppositions can be made.

Firstly, the enzyme is differentially active depending on the localization of the epithelial cells to whose cytoplasmic membrane it is integrated and oriented extracellularly. ACE2 is most active in blood vessels, where pH 7.4 determines the highly negatively charged surface of the enzyme (Figure 5), adsorbing electrostatically the positively charged peptide. In the lumen of the upper respiratory vessels ACE2 is weakly active or even inactive, since there the pH of the secretion is lowered to pH 5.5–6.5 [39–42] because the  $\text{CO}_2$  (dissolved from the air) forms the carbonic acid  $\text{H}_2\text{CO}_3$  whose dissociation to  $\text{HCO}_3^-$  reduces the pH.

Secondly, the ACE2 enzyme must be differentially active toward the three peptides (Figure 1). Considering only the surface electrostatic potential of the catalytic domain (Figure 5, pH 7.4), the enzyme should be most active with angiotensin-I due to its three positive coulombic charges in the chain. This peptide binds most strongly to the negatively charged (at pH 7.4) surface and to the positively charged  $\text{Zn}^{2+}$  atom in the active center by the negatively charged carboxylic group of the C-terminal amino acid residue. Angiotensin-II is less readily adsorbed because it has only two positively charged groups in the chain. Angiotensin-1–9 also has two positive charges in the chain, but its C-terminal amino acid

residue is neutral because it carries both a positive and negative charge and therefore binds weakly to the  $\text{Zn}^{2+}$  atom in the active center.

### 3.7. Verification of the 3D Reconstruction Procedure

In the present work, we have developed a procedure for the reconstruction of a 3D model of a globular protein whose amino acid sequence is known, and for which there exists an analogical protein with a determined 3D structure (obtained by X-ray crystallographic, NMR, or cryoEM techniques), which we refer to as the “basic protein”. For this purpose, we compare the polypeptide chains of the two proteins, called, respectively, basic and mutant, and then substitute the different amino acid residue in the basic 3D model, assuming that they are point mutations [21]. The approach is applicable when the two polypeptide chains are similar and there are no deletions and insertions. Then, we use a computer program for mutational analysis which generates the atomic coordinates of the substituting amino acid residues in the basic 3D structure. In this process, the atomic coordinates of the surrounding amino acid residues can be slightly shifted. Repeating this procedure step-by-step, we substitute all the different (mutant) residues. The individual shifts are summed up, which can lead to the appearance of a methodological artifact when the number of mutant residues is large. Additionally, we optimize the 3D structure with another program which minimizes the free energy of the protein globule by shifting the atomic coordinates. This raises the question of how different the reconstructed 3D structure is from the real one.

To validate the 3D reconstruction procedure, we have to find a couple of two analogical proteins (a basic and a mutant) with an experimentally determined 3D structure, substitute the different (mutant) amino acid residues in the basic 3D model, and compare the reconstructed 3D model of the mutant protein with its experimentally determined 3D structure. Finding such a couple of proteins with a reliably determined 3D structure is not a trivial task, because different 3D models have been deposited in the Protein Data Bank for the majority of proteins and it is not clear which one is correct. To select a model, we calculate the electrostatic component  $\Delta G_{\text{el}}$  of the folding energy  $\Delta G_{\text{fold}}$ , using it as a criterion for correctness of the models. This criterion is based on the conception that  $\Delta G_{\text{el}}$  is determined by the free energy of a 3D system of coulombic charges; the shift of their coordinates alters the  $\Delta G_{\text{el}}$  value. The absolute value of  $\Delta G_{\text{el}}$  and its pH-dependent  $\Delta G_{\text{el}}(\text{pH})$  are very sensitive to small changes in the 3D structure of the protein globule, as these emerge by the optimization of the 3D structure (Figure 9). Then, the model with maximal 3D stability is the one that has the minimal  $\Delta G_{\text{el}}$  (maximal in absolute value with negative sign, a deeper minimum of the  $\Delta G_{\text{el}}(\text{pH})$  dependence), which is determined by the most energetically optimal location of the coulomb charges of the protein globule.

Assuming that the crystallographic models of well-known proteins are the most reliable, we have chosen the horse myoglobin which has a well-known human analogue. For these two human–animal proteins, we have reconstructed the 3D model of the horse analog by the substitution of 18 mutant residues in the human 3D model, according to the difference in the amino acid sequences of the two proteins. The root mean square deviation (RMSD) of the atomic positions of the backbone atoms of original and reconstructed horse myoglobin is 0.076 nm (values under 0.1 nm indicates very high degree of identity). The coincidence of the isoelectric points  $\text{pI}$  6.7 and the pH-dependent  $\text{nz}(\text{pH})$  of the net charge  $\text{nz}$  of the original and reconstructed 3D models of the horse myoglobin (Figure 11, Insert) are evidence that all mutations are included correctly. Figure 11 shows very close courses of the pH-dependent  $\Delta G_{\text{el}}(\text{pH})$  of the electrostatic component  $\Delta G_{\text{el}}$  of the folding energy of the original and the reconstructed 3D models of horse myoglobine. Taking into consideration the high sensitivity of  $\Delta G_{\text{el}}(\text{pH})$  dependence to a small shift in the atomic coordinates, and also that the optimization procedure is not performed, the nearness of the two pH curves shows that the 3D models (original and reconstructed) can be assumed as practically identical. This verification confirms the procedure for the reconstruction of the 3D model of the rabbit (mutant) protein by substitution of the different amino acid residues in the 3D model of the human protein (Section 2.1).

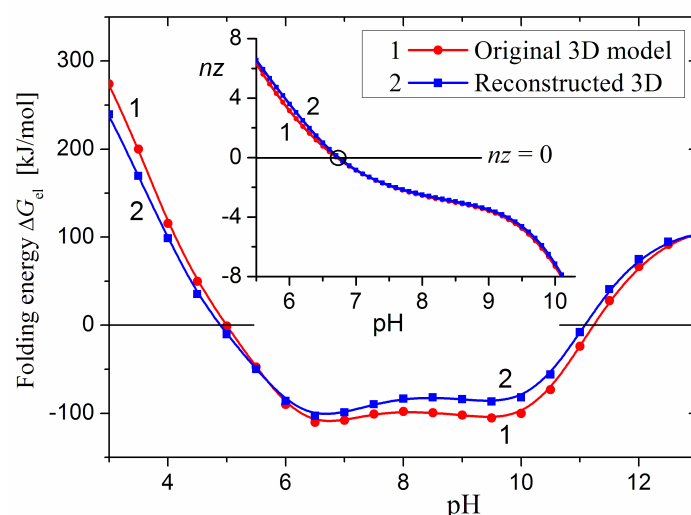


**Figure 11.** pH dependences of the electrostatic component  $\Delta G_{el}$  of the folding energy (the main figure) and the net electric charge  $nz$  (the inserted figure) of horse myoglobin according to the original 3D model (PDB: 1AZI, curves 1) and its reconstructed analog (curves 2). The reconstruction of the 3D model of the horse myoglobin is performed on the base of human myoglobin (PDB: 3RGK) considering the difference in the amino acid sequences of the human and horse myoglobin without optimization.

### 3.8. Glycosylation and Glycation

Glycosylation is a post-translational modification of the proteins which takes place in the Golgi apparatus. This process is enzymatically mediated by glycosyltransferases and results in the formation of N- or O-glycans [43]. The tree-branched sugar structures formed by glycosylation terminate in mammals with one to four sialic acid residues whose carboxyl groups are deprotonated ( $\text{COOH} \rightarrow \text{COO}^-$ ) at physiological pH values, and this results in the appearance of up to four negative charges for each modified amino acid residue which shifts the isoelectric point to a low pH [44–48]. Conversely, the enzymatic removal of sugar residues results in a decrease in molecular mass and an increase in isoelectric point [49–52].

In contrast to glycosylation, glycation is a non-enzymatic, spontaneous reaction, in which the free amino groups ( $-\text{NH}_2$ ) of lysine and asparagine amino acid residues interact with the carbonyl group of monosaccharides (such as glucose, fructose, etc.) [53,54]. This leads to the formation of a Schiff base which undergoes rearrangement and finally an Amadori product is obtained. The consequence of the glycation is that as the number of the charged amino groups ( $-\text{NH}_3^+$ ) in the protein decreases, the positive coulombic charges also respectively reduce, and the isoelectric point shifts to a low pH.

In our previous experiment with isoelectric focusing (Section 3.5), we have used recombinant hACE2 expressed in embryonic human cell culture (line HEK 293). There is evidence in the literature indicating the presence of various sugars (fucose, mannose, galactose, N-acetylglucosamine) bound to proteins synthesized by this recombinant technology [55]. Because of the high glucose concentration in the medium of the cell culture, both processes of glycosylation and glycation are reinforced, leading to an increase in the molecular mass and a change in the electric charge. The used recombinant hACE2 catalytic domain contains 723 amino acid residues (from 18 to 740, numbered from N-end, Figure 2) and covalently connected Fc-tag, totaling 952 residues. The molecular mass (calculated according to the amino acid content) was 83.68 kDa and 109.4 kDa with and without Fc-tag, respectively. The molecular mass of this recombinant protein (hACE2+Fc-tag), experimentally measured by gel electrophoreses, is 145–150 kDa, according to the data of the producer. The increment 36–41 kDa is caused by the attached sugars.

#### 4. Methods

The analysis of the analogy in amino acid content and sequence of the polypeptide chain of human (hACE2) and rabbit (rACE2) angiotensin-converting enzyme-2 was performed using their published primary sequences obtained by analysis of the DNA sequence of the corresponding gene [56]. The crystallographic 3D model of the catalytic domain of hACE2 with code 1r42 in the Protein Data Bank (PDB) was used for the reconstruction of its rabbit analog. The 1r42 model is a recombinant protein expressed in the system based on the butterfly *Spodoptera frugiperda* and crystallized in aqueous solution with high ionic strength 100 mM Tris-HCl (pH 8.5), 200 mM MgCl<sub>2</sub>, and 14% polyethylene glycol 8000 [57]. The atomic coordinates of this model were optimized by the computer programs Chimera [58] and YASARA [59] to find the energetically most advantageous conformation of the flexible side-chain amino acid residues under physiological conditions in an aqueous medium: pH 7.4 and ionic strength 0.15 mol/L NaCl, at which the free energy of the protein globule is minimal. The 3D atomic coordinates of rabbit ACE2 were obtained by the replacement of the different amino acid residues using the program for in silico mutagenesis DUET, combining two programs: the Site Directed Mutator (SDM) and a program for missense mutations, called mCSM [60–62], which finds the 3D atomic coordinates in the protein globule after the replacement of an amino acid residue in hACE2 structure with the corresponding residue of rACE2 in its polypeptide chain. The mutation procedure was repeated step-by-step to replace all different residues. By this, three 3D models of rabbit ACE2 were obtained: one based on the original hACE2 (PDB: 1r42), and two optimized. To select the most appropriate one for further calculations of the electric and thermodynamic properties, the six models (three human and three rabbit) of the catalytic domain of ACE2 were examined by calculation of the electrostatic component  $\Delta G_{el}$  of their folding energy  $\Delta G_{fold}$ , using as a criterion the maximal absolute value of the decrease in the Gibbs free energy at the transition from fully unfolded (random coil) to the 3D structure (protein globule) of the corresponding model. The results of this examination disclosed that the most energetically advantageous (with minimal free energy) are the two models obtained by optimization using the program YASARA. The results shown in this article, including 3D models, the net charge, the surface electrostatic potential, and the electrostatic folding energy  $\Delta G_{el}$ , correspond to these two models. The program AlphaFold2 [63] was used for the independent construction of a 3D model of rACE2 using only its primary amino acid sequence.

The computer programs for macromolecular alignment SuperPose [64] and TM-score [65] were used to compare the 3D structure of the catalytic domains of the hACE2 and the reconstructed rACE2. The computer program for protein electrostatics Blueues [66] was employed to calculate the pH-dependent electrostatic component  $\Delta G_{el}$  of the free energy of folding. The program Propka [67,68] for protein electrostatics was employed to compute the electric parameters of the catalytic domains of hACE2 and rACE2: pH dependence of the net charge, the isoelectric point (zero net charge), and the local surface electrostatic potential at a given pH. The programs Chimera [58], PBEQ Solver [69,70], and VMD: Visual molecular dynamics 1.9.2 [71] were used to visualize the 3D models and the electrostatic potential on the surface of the protein globules.

#### 5. Conclusions

Comparison of the primary amino acid sequence of the full polypeptide chains (including the signal peptide, catalytic domain, and the transmembrane segment) of human ACE2 (hACE2) and rabbit ACE2 (rACE2) shows that 85% of the residues are identical (686 of 805 amino acids). A high degree of similarity is also observed by 3D macromolecule alignment at the tertiary structure: the  $\alpha$ -helical segments are identical; only two unstructured segments have noticeably different coordinates. The pH dependences of the electrostatic component  $\Delta G_{el}$  of the folding energy of the two catalytic domains disclose that the 3D structure of the two proteins is maximally stable at pH 6.5. The negative values of  $\Delta G_{el}$  indicate that the intramolecular electrostatic interactions contribute to the 3D structural

stability (determined mainly by the hydration energy) in the range pH 5.5–9.8 for hACE2 and 5.1–9.8 for rACE2.

The number of positively chargeable groups of the unfolded polypeptide chains of the catalytic domains of hACE2 and rACE2 are 69 and 70, and the negatively chargeable ones are 111 and 115, as calculated at pH 1 and pH 14, respectively. The charge number depends on the conformation of the polypeptide chain: the 3D catalytic globules have 68 and 69 (pH 1) positive charges, and 100 and 104 (pH 14) negative charges, respectively, for hACE2 and rACE2. I.e., both catalytic domains have one positive and 10 negative charges less in 3D-folded conformation, as compared to the fully unfolded chain (random coil). The difference is due to the different degree of ionization of the same type of chargeable groups on the surface (in contact with aqueous medium) and being submerged in the hydrophobic core of the protein globule. In contrast to the different number of the chargeable groups, in 3D conformation the isoelectric points pI 5.21 of hACE2 and rACE2 coincide, and their net charges are almost equal: about 25 negative coulombic charges at pH 7.4. The electrostatic potential is likewise distributed on the surface of the two catalytic globules, but it is unhomogeneous and strongly dependent on the pH of the aqueous medium: at a low pH, the potential is almost positive, but strongly negative at pH 7.4. Considering the positive charges of the angiotensin peptides, the pH-determined difference of the surface potential suggests that the ACE2 enzyme is most active at pH 7.4.

The significant similarity in 3D structures, thermodynamic, and electric properties of the catalytic domains of hACE2 and rACE2 suggests that rabbits can be successfully used as an animal model for the investigation of both infectious and non-infectious diseases with involved ACE2, including the coronavirus infection, allowing the results to be extrapolated to humans, and also for the in vivo investigation of various agents that disrupt the interaction of ACE2 with the coronavirus S-protein, such as potential antiviral drugs.

**Author Contributions:** S.H.H.: computer methodology, computer experiments, molecular models, literature search. T.T.P.: computer calculations, schemes, literature search. A.M.Z.: conceptualization, interpretation, graphics, writing and revision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Bulgaria Sciences Fund, Grant Nr. KPI-06-H53/2 (11 November 2021).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Eriksson, U.; Danilczyk, U.; Penninger, J.M. Just the beginning: Novel functions for angiotensin-converting enzymes. *Curr. Biol.* **2002**, *12*, R745–R752. [[CrossRef](#)] [[PubMed](#)]
2. Oudit, G.Y.; Crackower, M.A.; Backx, P.H.; Penninger, J.M. The role of ACE2 in cardiovascular physiology. *Trends Cardiovasc. Med.* **2003**, *13*, 93–101. [[CrossRef](#)] [[PubMed](#)]
3. Yagil, Y.; Yagil, C. Hypothesis: ACE2 modulates blood pressure in the mammalian organism. *Hypertension* **2003**, *41*, 871–873. [[CrossRef](#)] [[PubMed](#)]
4. Raizada, M.K.; Ferreira, A.J. ACE2: A new target for cardiovascular disease therapeutics. *J. Cardiovasc. Pharmacol.* **2007**, *50*, 112–119. [[CrossRef](#)]
5. Tipnis, S.R.; Hooper, N.M.; Hyde, R.; Karran, E.; Christie, G.; Turner, A.J. A human homolog of angiotensin-converting enzyme: Cloning and functional expression as a captopril-insensitive carboxypeptidase. *J. Biol. Chem.* **2000**, *275*, 33238–33243. [[CrossRef](#)]
6. Donoghue, M.; Hsieh, F.; Baronas, E.; Godbout, K.; Gosselin, M.; Stagliano, N.; Acton, S. A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ. Res.* **2000**, *87*, e1–e9. [[CrossRef](#)]
7. Crackower, M.A.; Sarao, R.; Oudit, G.Y.; Yagil, C.; Kozieradzki, I.; Scanga, S.E.; Olivera-dos-Santos, A.J.; da Costa, J.; Zhang, L.; Pei, Y.; et al. Angiotensin-converting enzyme 2 is an essential regulator of heart function. *Nature* **2002**, *417*, 822–828. [[CrossRef](#)]
8. Danilczyk, U.; Eriksson, U.; Crackower, M.A.; Penninger, J.M. A story of two ACEs. *J. Mol. Med.* **2003**, *81*, 227–234. [[CrossRef](#)]

9. Avdonin, P.P.; Rybakova, E.Y.; Trufanov, S.K.; Avdonin, P.V. SARS-CoV-2 receptors and their involvement in cell infection. *Biochem. (Mosc.) Suppl. Ser. A Membr. Cell Biol.* **2023**, *17*, 1–11. [\[CrossRef\]](#)
10. Shirbhate, E.; Pandey, J.; Patel, V.K.; Kamal, M.; Jawaid, T.; Gorain, B.; Kesharwani, P.; Rajak, H. Understanding the role of ACE-2 receptor in pathogenesis of COVID-19 disease: A potential approach for therapeutic intervention. *Pharmacol. Rep.* **2021**, *73*, 1539–1550. [\[CrossRef\]](#)
11. Wan, Y.; Shang, J.; Graham, R.; Baric, R.S.; Li, F. Receptor recognition by the novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* **2020**, *94*, 1110–1128. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Wu, J.; Deng, W.; Li, S.; Yang, X. Advances in research on ACE2 as a receptor for 2019-nCoV. *Cell. Mol. Life Sci.* **2021**, *78*, 531–544. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Hikmet, F.; Méar, L.; Edvinsson, Å.; Micke, P.; Uhlén, M.; Lindskog, C. The protein expression profile of ACE2 in human tissues. *Mol. Syst. Biol.* **2020**, *16*, e9610. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Chen, Y.; Guo, Y.; Pan, Y.; Zhao, Z.J. Structure analysis of the receptor binding of 2019-nCoV. *Biochem. Biophys. Res. Commun.* **2020**, *525*, 135–140. [\[CrossRef\]](#)
15. Verma, S.; Saksena, S.; Sadri-Ardekani, H. ACE2 receptor expression in testes: Implications in coronavirus disease 2019 pathogenesis. *Biol. Reprod.* **2020**, *103*, 449–451. [\[CrossRef\]](#)
16. Hamming, I.; Timens, W.; Bulthuis, M.L.C.; Lely, A.T.; Navis, G.V.; van Goor, H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol. A J. Pathol. Soc. Great Br. Irel.* **2004**, *203*, 631–637. [\[CrossRef\]](#)
17. Guy, J.L.; Jackson, R.M.; Acharya, K.R.; Sturrock, E.D.; Hooper, N.M.; Turner, A.J. Angiotensin-converting enzyme-2 (ACE2): Comparative modeling of the active site, specificity requirements, and chloride dependence. *Biochemistry* **2003**, *42*, 13185–13192. [\[CrossRef\]](#)
18. Lubbe, L.; Cozier, G.E.; Oosthuizen, D.; Acharya, K.R.; Sturrock, E.D. ACE2 and ACE: Structure-based insights into mechanism, regulation and receptor recognition by SARS-CoV. *Clin. Sci.* **2020**, *134*, 2851–2871. [\[CrossRef\]](#)
19. Warner, F.J.; Smith, A.I.; Hooper, N.M.; Turner, A.J. Angiotensin-converting enzyme-2: A molecular and cellular perspective. *Cell. Mol. Life Sci. CMLS* **2004**, *61*, 2704–2713. [\[CrossRef\]](#)
20. Huang, L.; Sexton, D.J.; Skogerson, K.; Devlin, M.; Smith, R.; Sanyal, I.; Parry, T.; Kent, R.; Enright, J.; Wu, Q.L.; et al. Novel peptide inhibitors of angiotensin-converting enzyme 2. *J. Biol. Chem.* **2003**, *278*, 15532–15540. [\[CrossRef\]](#)
21. Hristova, S.H.; Zhivkov, A.M. Three-dimensional structural stability and local electrostatic potential at point mutations in spike protein of SARS-CoV-2 coronavirus. *Int. J. Mol. Sci.* **2024**, *25*, 2174. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Huang, C.; Jiang, Y.; Yan, J. Comparative analyses of ACE2 and TMPRSS2 gene: Implications for the risk to which vertebrate animals are susceptible to SARS-CoV-2. *J. Med. Virol.* **2021**, *93*, 5487–5504. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Preziuso, S. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) exhibits high predicted binding affinity to ACE2 from lagomorphs (rabbits and pikas). *Animals* **2020**, *10*, 1460. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Li, Y.; Wang, H.; Tang, X.; Fang, S.; Ma, D.; Du, C.; Wang, Y.; Pan, H.; Yao, W.; Zhang, R.; et al. SARS-CoV-2 and three related coronaviruses utilize multiple ACE2 orthologs and are potently blocked by an improved ACE2-Ig. *J. Virol.* **2020**, *94*, e01283-20. [\[CrossRef\]](#)
25. Zaharieva, M.M.; Foka, P.; Karamichali, E.; Kroumov, A.D.; Philipov, S.; Ilieva, Y.; Kim, T.C.; Podlesniy, P.; Manasiev, Y.; Kussovski, V.; et al. Photodynamic Inactivation of Bovine Coronavirus with the Photosensitizer Toluidine Blue O. *Viruses* **2024**, *16*, 48. [\[CrossRef\]](#)
26. Esteves, P.J.; Abrantes, J.; Baldauf, H.M.; BenMohamed, L.; Chen, Y.; Christensen, N.; González-Gallego, J.; Giacani, L.; Hu, J.; Kaplan, G.; et al. The wide utility of rabbits as models of human diseases. *Exp. Mol. Med.* **2018**, *50*, 1–10. [\[CrossRef\]](#)
27. Fritz, M.; de Riols de Fonclare, D.; Garcia, D.; Beurlet, S.; Becquart, P.; Rosolen, S.G.; Briend-Marchal, A.; Leroy, E.M. First Evidence of Natural SARS-CoV-2 infection in domestic rabbits. *Vet. Sci.* **2022**, *9*, 49. [\[CrossRef\]](#)
28. Mykityn, A.Z.; Lamers, M.M.; Okba, N.M.; Breugem, T.I.; Schipper, D.; van den Doel, P.B.; van Run, P.; van Amerongen, G.; de Waal, L.; Koopmans, M.P.; et al. Susceptibility of rabbits to SARS-CoV-2. *Emerg. Microbes Infect.* **2021**, *10*, 1–7. [\[CrossRef\]](#)
29. Jia, H.; Neptune, E.; Cui, H. Targeting ACE2 for COVID-19 therapy: Opportunities and challenges. *Am. J. Respir. Cell Mol. Biol.* **2021**, *64*, 416–425. [\[CrossRef\]](#)
30. Zhang, H.; Wada, J.; Hida, K.; Tsuchiyama, Y.; Hiragushi, K.; Shikata, K.; Wang, H.; Lin, S.; Kanwar, S.Y.; Makino, H. Collectrin, a collecting duct-specific transmembrane glycoprotein, is a novel homolog of ACE2 and is developmentally regulated in embryonic kidneys. *J. Biol. Chem.* **2001**, *276*, 17132–17139. [\[CrossRef\]](#)
31. Zhang, Y.; Wada, J. Collectrin, a homologue of ACE2, its transcriptional control and functional perspectives. *Biochem. Biophys. Res. Commun.* **2007**, *363*, 1–5. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Chu, P.L.; Le, T.H. Role of collectrin, an ACE2 homologue, in blood pressure homeostasis. *Curr. Hypertens. Rep.* **2014**, *16*, 1–7. [\[CrossRef\]](#)
33. Danilczyk, U.; Sarao, R.; Remy, C.; Benabbas, C.; Stange, G.; Richter, A.; Arya, S.; Pospisilik, J.A.; Singer, D.D.; Camargo, S.M.; et al. Essential role for collectrin in renal amino acid transport. *Nature* **2006**, *444*, 1088–1091. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Singer, D.; Camargo, S.M. Collectrin and ACE2 in renal and intestinal amino acid transport. *Channels* **2011**, *5*, 410–423. [\[CrossRef\]](#) [\[PubMed\]](#)

35. Turner, A.J.; Nalivaeva, N.N. Angiotensin-converting enzyme 2 (ACE2): Two decades of revelations and re-evaluation. *Peptides* **2022**, *151*, 170766. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Fauchere, I.I.; Pliska, N. Hydrophobic parameters II of amino acid side-chains from the partitioning of N-acetyl-amino acid amides. *Eur. J. Med. Chem-Chim.* **1983**, *18*, 369–375.
37. Hristova, S.H.; Zhivkov, A.M. Omicron Coronavirus: pH-Dependent Electrostatic Potential and Energy of Association of Spike Protein to ACE2 Receptor. *Viruses* **2023**, *15*, 1752. [\[CrossRef\]](#)
38. Hristova, S.H.; Zhivkov, A.M. Wild Type Coronavirus and Omicron Variant: pH-Dependent Energy of Association of the Spike Protein to ACE2, Receptor. In *Prime Archives in Virology*; Raad, H., Ed.; Vide Leaf: Hyderabad, India, 2023.
39. Maier, C.; Schadock, I.; Haber, P.K.; Wysocki, J.; Ye, M.; Kanwar, Y.; Falsk, C.A.; Yu, X.; Hoit, B.D.; Adams, N.G.; et al. Prolylcarboxypeptidase deficiency is associated with increased blood pressure, glomerular lesions, and cardiac dysfunction independent of altered circulating and cardiac angiotensin II. *J. Mol. Med.* **2017**, *95*, 473–486. [\[CrossRef\]](#)
40. Vickers, C.; Hales, P.; Kaushik, V.; Dick, L.; Gavin, J.; Tang, J.; Godbout, K.; Parsons, T.; Baronas, E.; Hsieh, F.; et al. Hydrolysis of biological peptides by human angiotensin-converting enzyme-related carboxypeptidase. *J. Biol. Chem.* **2002**, *277*, 14838–14843. [\[CrossRef\]](#)
41. Houston, M.C. Pleural Fluid PH: Diagnostic, therapeutic, and prognostic value. *Am. J. Surg.* **1987**, *154*, 333–337. [\[CrossRef\]](#)
42. Fischer, H.; Widdicombe, J.H. Mechanisms of acid and base secretion by the airway epithelium. *J. Membr. Biol.* **2006**, *211*, 139. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Moremen, K.W.; Tiemeyer, M.; Nairn, A.V. Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 448–462. [\[PubMed\]](#)
44. Steen, P.V.D.; Rudd, P.M.; Dwek, R.A.; Opdenakker, G. Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 151–208. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Ramirez Hernandez, E.; Hernandez-Zimbron, L.F.; Martinez Zuniga, N.; Leal-Garcia, J.J.; Ignacio Hernandez, V.; Ucharima-Corona, L.E.; Campos, E.P.; Zenteno, E. The role of the SARS-CoV-2 S-protein glycosylation in the interaction of SARS-CoV-2/ACE2 and immunological responses. *Viral Immunol.* **2021**, *34*, 165–173. [\[CrossRef\]](#)
46. Ernst, J.F.; Prill, S.H. O-glycosylation. *Med. Mycol.* **2001**, *39*, 67–74. [\[CrossRef\]](#)
47. Mitra, N.; Sinha, S.; Ramya, T.N.; Surolia, A. N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends Biochem. Sci.* **2006**, *31*, 156–163. [\[CrossRef\]](#)
48. Wandall, H.H.; Nielsen, M.A.; King-Smith, S.; de Haan, N.; Bagdonaite, I. Global functions of O-glycosylation: Promises and challenges in O-glycobiology. *FEBS J.* **2021**, *288*, 7183–7212. [\[CrossRef\]](#)
49. Grueninger-Leitch, F.; D'Arcy, A.; D'Arcy, B.; Chène, C. Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases. *Protein Sci.* **1996**, *5*, 2617–2622. [\[CrossRef\]](#)
50. Bünning, P.; Budek, W.; Escher, R.; Schönherr, E. Characteristics of angiotensin converting enzyme and its role in the metabolism of angiotensin I by endothelium. *J. Cardiovasc. Pharmacol.* **1986**, *8*, S52–S57. [\[CrossRef\]](#)
51. Lanzillo, J.J.; Stevens, J.; Dasarathy, Y.; Yotsumoto, H.; Fanburg, B.L. Angiotensin-converting enzyme from human tissues. Physicochemical, catalytic, and immunological properties. *J. Biol. Chem.* **1985**, *260*, 14938–14944. [\[CrossRef\]](#)
52. Lanzillo, J.J.; Joanne, S.; Tumas, J.; Fanburg, B.L. Spontaneous change of human plasma angiotensin I converting enzyme isoelectric point. *Arch. Biochem. Biophys.* **1983**, *227*, 434–439. [\[CrossRef\]](#)
53. Reily, C.; Stewart, T.J.; Renfrow, M.B.; Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **2019**, *15*, 346–366. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Blakytty, R.; Harding, J.J. Glycation (non-enzymic glycosylation) inactivates glutathione reductase. *Biochem. J.* **1992**, *288*, 303–307. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Conroy, J.M.; Hartley, J.L.; Soffer, R.L. Canine pulmonary angiotensin-converting enzyme. Physicochemical, catalytic and immunological properties. *Biochim. Biophys. Acta (BBA)-Enzymol.* **1978**, *524*, 403–412. [\[CrossRef\]](#)
56. Ross, M.T.; Grafham, D.V.; Coffey, A.J.; Scherer, S.; McLay, K.; Muzny, D.; Platzer, M.; Howell, G.R.; Burrows, C.; Bird, C.P.; et al. The DNA sequence of the human X chromosome. *Nature* **2005**, *434*, 325–337. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Towler, P.; Staker, B.; Prasad, S.G.; Menon, S.; Tang, J.; Parsons, T.; Ryan, D.; Fisher, M.; Williams, D.; Dales, N.A.; et al. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *J. Biol. Chem.* **2004**, *279*, 17996–18007. [\[CrossRef\]](#)
58. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [\[CrossRef\]](#)
59. Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins Struct. Funct. Bioinform.* **2009**, *77*, 114–122. [\[CrossRef\]](#)
60. Pandurangan, A.P.; Ochoa-Montano, B.; Ascher, D.B.; Blundell, T.L. SDM: A server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* **2017**, *45*, W229–W235. [\[CrossRef\]](#)
61. Worth, C.L.; Preissner, R.; Blundell, T.L. SDM—A server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* **2011**, *39* (Suppl. S2), W215–W222. [\[CrossRef\]](#)
62. Pires, D.E.; Ascher, D.B.; Blundell, T.L. DUET: A server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.* **2014**, *42*, W314–W319. [\[CrossRef\]](#) [\[PubMed\]](#)

63. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)] [[PubMed](#)]
64. Available online: <http://wishart.biology.ualberta.ca/SuperPose> (accessed on 30 August 2024).
65. Zhang, Y.; Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **2004**, *57*, 702–710. [[CrossRef](#)] [[PubMed](#)]
66. Walsh, I.; Minervini, G.; Corazza, A.; Esposito, G.; Tosatto, S.C.E.; Fogolari, F. Bluues server: Electrostatic properties of wild-type and mutated protein structures. *Bioinformatics* **2012**, *28*, 2189–2190. [[CrossRef](#)]
67. Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **2004**, *32*, W665–W667. [[CrossRef](#)]
68. Li, H.; Robertson, A.D.; Jensen, J.H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct. Funct. Bioinform.* **2005**, *61*, 704–721. [[CrossRef](#)]
69. Vargyas, M.; Vasko-Szedlar, J.; Roux, B.; Im, W. PBEQ-Solver for online visualization of electrostatic potential of Biomolecules. *Nucleic Acids Res.* **2008**, *36*, W270–W275.
70. Available online: <https://charmm-gui.org/?doc=tutorial&project=pbeqsolver> (accessed on 30 August 2024).
71. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.