JOURNAL OF BIOCHEMICAL SCIENCES
Jan-Mar 2024, VOL. 1, ISSUE 1, pp. 3-7
https://doi.org/10.61577/jbs.2024.100002

RESEAPRO JOURNALS

SHORT COMMUNICATION

OPEN ACCESS

Amino acid interaction networks: application to biomolecules

Wai-Yim Ching[1], Puja Adhikari[1] and Bahaa Jawad[1,2]

[1]Department of Physics and Astronomy, University of Missouri-Kansas City, Missouri, USA
[2]Department of Applied Sciences, University of Technology, Baghdad, Iraq

ABSTRACT

The formation of the three-dimensional (3D) protein structure depends essentially on the amino acid-amino acid (AA-AA) interaction networks. Structural and computational approaches have been developed to predict these amino acid networks. However, there is a pressing need to estimate these interaction networks based on more accurate approaches, such as quantum-mechanical calculations. In this context, we calculated a parameter known as amino acid bond pairs (AABP), which we successfully applied to a variety of biological systems. AABP is applied to biomolecular systems, including SARS-CoV-2 spike protein, and the interface between spike protein and angiotensin enzyme 2 (ACE2). This new concept can be applied to protein design, understanding the mutation process leading to vaccine development.

KEYWORDS

Amino acid interaction network; Protein structure; Biomolecular system; AA bond pair unit; DFT calculation

ARTICLE HISTORY

Received 12 October 2023;
Revised 21 December 2023;
Accepted 26 December 2023

Introduction

Proteins are the main cellular machinery that perform a wide range of functions necessary for life. They are composed of amino acids (AAs) that are linked together by peptide bonds and naturally appear on a compact 3D structure. The AA interaction networks play an important role in the process of generating 3D protein structures, also known as protein folding [1,2]. This process is caused by the covalent and noncovalent interactions between AAs to produce 3D protein structures. Noncovalent interactions, specifically the hydrophobic effect, traditional hydrogen bonding, Coulombic interactions, and van der Waals interactions, play a predominant role in this phenomenon [3]. In terms of structure, AAs share a common foundational structure, with distinctions arising solely from the sidechain or R-group unique to each AA. Each AA comprises a central alpha carbon atom bonded to hydrogen, an acidic carboxyl group (COOH), an amino group (NH_2), and an individualized side chain. These 20 canonical AAs exhibit varying properties, including acidity, basicity, polarity, and non-polarity. Consequently, the interactions among AAs exhibit distinct attributes governed by the nature of their respective side chains. AA sequence is important for determining the structure, functions, and interactions of proteins. In particular, the conserved sequence remains basically unmodified along the so-called phylogenetic tree. Conventional analysis of conserved sequences relies on the comparative study of the sequence homology of numerous proteins from closely related species. The impact of these interactions solely relies on the linear sequence of AA residues. The current approach focuses solely on interactions between adjacent or nearest neighbor (NN) AAS, neglecting the involvement of non-local (NL) AAs in the three-dimensional (3D) space that constitutes the secondary or tertiary protein structures. Recent efforts aim to extend beyond this linear sequence model, primarily relying on statistical techniques and probability theory. These methods include multiple sequence alignment (MSA) [4], statistical coupling analysis (SCA) [5], and direct coupling analysis (DCA) [6], which are all based on similarities in the AA sequences within proteins. In contrast, more efficient algorithms and data analysis techniques have shown promise in elucidating the directional dependencies among AAs [4]. Nevertheless, a common limitation persists across these approaches: they do not adequately quantify interactions between AAs that are not nearest neighbors. Unlike paired atoms or small molecules, the precise distance separating two nearest neighbor amino acids remains undefined. AAs are diverse biomolecules with variations in structure, size, orientation, and composition, making it challenging to establish a standardized parameter for routine use in the scientific literature [7]. Hence, it becomes imperative to employ atomic-level first-principles calculations for accurate quantification.

One of the most daunting challenges in biological research lies in the prediction of the structure of proteins that remain uncharted. Numerous strategies and techniques have emerged to address this challenge, encompassing the utilization of protein databases, sequence analysis, crystallography, and NMR spectroscopy. Of particular note is the introduction of cryogenic electron microscopy (cryo-EM), a revolutionary method capable of discerning biomolecular structures with nearly atomic precision [8-11]. Several databases, such as PDB, CNBI, and GenBank, offer access to protein sequences of interest. Recent reports suggest the potential for further enhancing the resolution of cryo-EM from its current 3.5 Å to less than 1.3 Å through the visualization of hydrogen atoms [12,13]. The new objective is to accurately determine the structures of novel proteins and validate numerous untested hypotheses. In this pursuit, the vital step is the refinement of protein structures using

***Correspondence:** Dr. Wai-Yim Ching, Department of Physics and Astronomy, University of Missouri-Kansas City, MO, 64110, USA, e-mail: ChingW@umkc.edu

RESEAPRO JOURNALS

extensive computational modeling. This approach provides intricate insights into both intra- and inter-protein interactions at the atomic level, shedding light on potential mechanisms of biological interactions in diverse environments [14]. The emergence of the COVID-19 pandemic in late 2019, caused by the newly identified severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), highlights the urgency of comprehending the structural intricacies and functionality of this virus [15-18]. Computational research has spurred atomic-level investigations in the pursuit of combating this virus effectively [14,19-21].

Methods

In this communication, we have provided a concise overview of advancements that transcend statistical methods for analyzing the interactions among all AAs in three-dimensional space. This progress involves the application of a quantitative descriptor known as amino acid bond pair (AABP) [22]. The capacity to conduct precise AABP calculations using *ab initio* quantum chemical techniques empowers the exploration of interactions within AAs and proteins at the atomic scale.

The key computational technique we used is the in-house developed OLCAO method [23]. In the OLCAO method, the basis expansion of the Bloch function is atomic orbitals. One crucial parameter in this context is the bond order (BO), denoted as $\rho\alpha\beta$. BO is computed for every pair of atoms (α, β) using a cutoff distance of 4.5 Å. The positions of individual atoms are well-defined quantities. However, AAs lack consistent positions due to their diverse atomic compositions, structures, and orientations. Consequently, attempting to establish a fixed separation distance between different amino acids in a protein for the purpose of characterizing their interactions is a conceptually flawed approach. With the utilization of the OLCAO method, interatomic interactions between all pairs of atoms are determined through quantum mechanics. This methodology enables the unambiguous definition of bonding between two amino acids (u, v), which we coined as an AABP:

$$AABP(u,v) = \sum_{\alpha \epsilon u} \sum_{\beta \epsilon v} \rho_{\alpha i, \beta j}$$

AABP stands out as a meticulously defined entity, as it comprehensively accounts for all conceivable bonds between two AAs. This single quantitative parameter, derived from electronic structure calculations, effectively quantifies the internal bonding strength among AAs. It can be further dissected into two significant components: nearest neighbor (NN) and non-local (NL) bonding, offering essential insights into the nature of inter-amino acid bonding within various biomolecules. From a geometric perspective, AABP can be employed to define a structural unit known as AABPU, which encompasses all the interacting AAs at a designated site or sequence number. This concept is valuable for characterizing the structural relationships within a biomolecule.

Based on the clearly defined stages of methodology and debate on many elements of the implications of the presence of the non-local AA contact network, we focused on this significant subject that has not been addressed in the research community. The novel AABP idea for AA interaction obviously goes beyond the present consideration of interatomic interactions in proteins [14,22]. The method we have developed has found practical applications in extensive computational efforts, including protein design, elucidating the mutation process, and therapeutic drug design, specifically in the context of addressing the COVID-19 pandemic [24-30].

Application to Biological Research

The current study necessitates thorough information on the interaction in various contexts involving single AAs or clusters of numerous AAs. These computations are presently underway and will be issued in the near future. To further clarify, we show an example of AABP and AABPU, a sequence number 493, within the interface of receptor binding domain (RBD) with angiotensin enzyme-2 (ACE2) from our published work [28]. In Figure 1, the RBD-ACE2 interface is depicted, with 15 Omicron variant (OV) mutations marked in blue. This interface holds particular significance as it shows the initial interaction between SARS-CoV-2 and the human receptor ACE2. In Figure 2, we can see the mutational changes in site 493 comparing Wild type (WT) and Omicron variant (OV). Their AABP values are shown in Table 1, incorporated from our published paper [28]. The total AABP value has increased after mutation due to the stronger interaction of NN AAs, NL AAs, and hydrogen bonding (HB). There is a significant increase in the number of interacting NL AAs from eight to ten. This also leads to an increase in the volume and area of OV mutation. Such detailed results for all 15 mutations are shown in our past publication [28]. AABP analysis can be conducted for all AAs in the protein. To investigate the impact of the OV mutations within the interface model, Figure 3 presents the AABP values for all interacting AAs between RBD and ACE2 from our published paper [28]. Figure 3(a) shows AABP values for mutated AAs of RBD, while Figure 3(b) illustrates the AABP values for the unmutated AAs within the RBD. Let us again discuss the site 493. OV R493 has one new interaction with K31 of ACE2 and a stronger interaction with E35 in comparison to WT Q493.

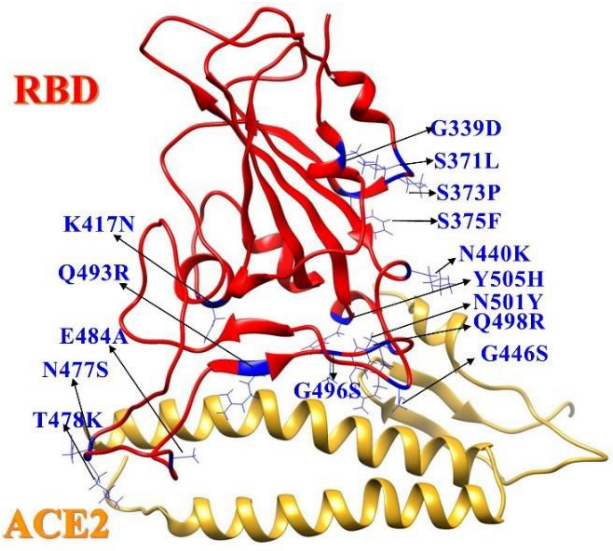

Figure 1. Ribbon structure of RBD-ACE2 interface of SARS-CoV-2 Omicron variant. Mutated 15 amino acids in RBD are marked in blue wire structure.
*RBD: Receptor binding domain, ACE2: Angiotensin enzyme-2.

https://doi.org/10.61577/jbs.2024.100002

Table 1. AABPU comparison for site 493 between WT and OV mutation in the RBD-ACE2 interface. AABP is measured in the electron units (e-). Incorporated from reference [28].

Site	Total AABP	NN AABP	NL AABP	AABP from HB	No. of NL AAs	Volume (Å^3)	Area (Å^2)
WT Q493	1.213	0.966	0.248	0.256	8	1459.0	955.1
OV R493	1.348	1.071	0.277	0.324	11	2021.0	1217.0

*AABPU: Amino acid bond pair units, AABP: Amino acid bond pairs, WT: Wild type, OV: Omicron variant, RBD: Receptor binding domain, ACE2: Angiotensin enzyme-2.

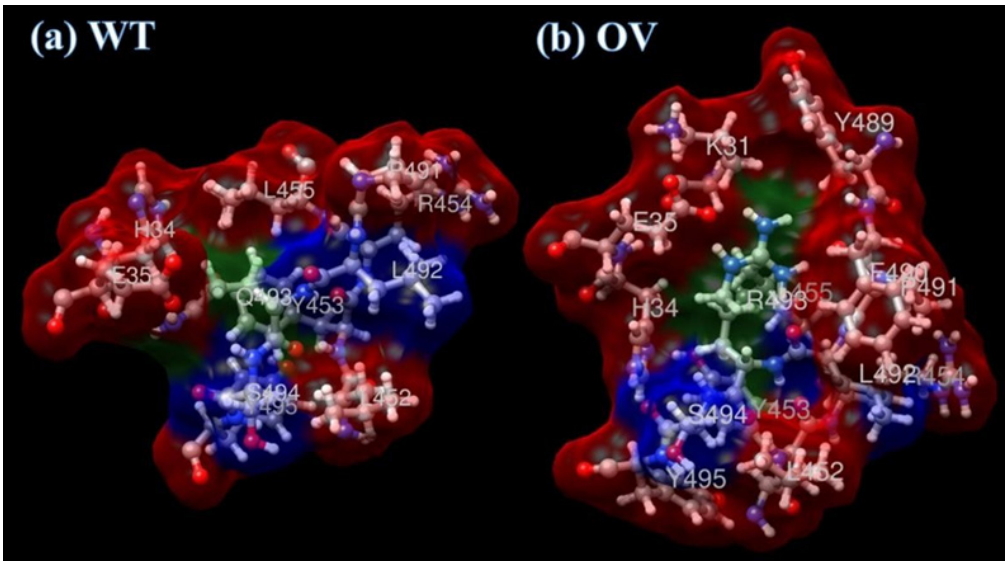

Figure 2. AABPU comparison between WT Q493 with its counterpart and OV R493 located in the RBD. The surface representation of site 493 is depicted in green, while its NN AAs are highlighted in blue, and NL AAs are in red. AABPU reveals the disparities in interatomic dynamics and the number of involved NL AAs.
*AABPU: Amino acid bond pair units, WT: Wild type, OV: Omicron variant, RBD: Receptor binding domain, NN AAS: Nearest neighbor amino acids, NL AAs: Non-local amino acids.

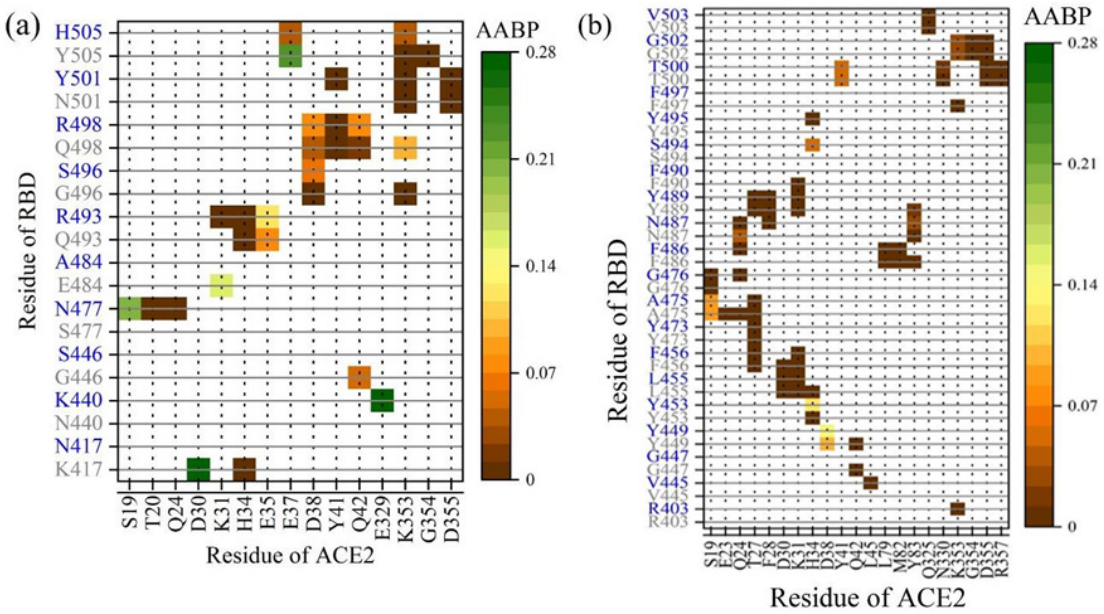

Figure 3. Visualization of AABP interaction: (a) highlights the interactions of mutated AAs in the RBD, while (b) illustrates interactions involving unmutated AAs in the RBD. AAs within RBD are the in y- axis labeled in gray for WT and blue for OV, respectively. The AAs from ACE2 involved in the interaction are shown on the x-axis. Incorporated from reference [28].
*AABP: Amino acid bond pairs, AAs: Amino acids, RBD: Receptor binding domain, WT: Wild type, OV: Omicron variant, ACE2: Angiotensin enzyme-2.

https://doi.org/10.61577/jbs.2024.100002

In addition, we studied mini proteins, LCB1 and LCB3, as SARS-CoV-2 therapeutic inhibitors using the combination of Molecular dynamics (MD) and *ab initio* methodology [24]. Mini proteins LCB1 and LCB3 are known to have higher binding affinity to the RBD with high neutralizing ability [31]. However, they are considered as large-size inhibitors [32]. A small size mini protein could penetrate easily into the tissues and cells, as well as allowing for a lower manufacturing price. Since LCB1–RBD has a higher binding affinity than LCB3–RBD, we chose LCB1 for further structural modification by truncating one of the alpha-helices followed by single and double amino acid substitutions. This resulted in enhanced binding of truncated LCB1 with RBD. From AABP analysis, replacing residue D17 with R showed stronger binding of truncated LCB1 with RBD [24]. In addition, the *ab initio* study provided a detailed role of HB and partial charge distribution in stabilizing the truncated LCB1 with RBD, complementing the MD analysis. While a priori force field MD uses fixed partial charge and cannot describe forming or breaking covalent bonding between atoms during the chemical reaction, our methodology allows for an accurate determination of the partial charge in RBD-ACE2 interface complex from *ab initio* computations, which were then fed into the MD force field, allowing for an accurate prediction of the electrostatic interactions [26].

Challenges and Future Directions

The most prominent challenge in *ab initio* calculation is the constraint imposed by size. MD can compute the behavior of hundreds of thousands of atoms. In contrast, the *ab initio* method has thus far been limited to 5000 atoms. Despite this size limitation, the accuracy of results obtained via *ab initio* calculations is notably higher.

To overcome this challenge, a divide-and-conquer strategy can be employed [33]. This involves conducting separate *ab initio* calculations on distinct portions of a large biomolecule. In doing so, it becomes feasible to study the system in a more detailed manner, thereby expanding the applicability of *ab initio* methods to larger biomolecules.

Looking forward, we believe that such an *ab initio* study extends to the realm of protein–protein interactions. It opens up the possibility of establishing *ab initio* interaction metrics through single-point calculations. This encompasses the exploration of mutations in certain sites, with certain AAs changing the behavior of the protein. Furthermore, there is promising potential for application in drug design. Such a study provides a rigorous and fundamental theoretical foundation, addressing the noticeable gap in prior research efforts.

Conclusions

Ab initio calculation, while being a well-known topic in material science, is still unconventional in biophysics. The detailed results based upon *ab initio* approaches have the capability to complement MD and enhance its quantitative predictions, providing atom-level information on bonding and identifying key interacting AAs as well as their PC. Such information in biological systems is certainly helpful in designing protein inhibitors and drugs, thus specifically contributing to the field of biomolecule science.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

1. EO'Rourke KF, Gorman SD, Boehr DD. Biophysical and computational methods to analyze amino acid interaction networks in proteins. Comput Struct Biotechnol J. 2016;14:245-251. https://doi.org/10.1016/j.csbj.2016.06.002
2. Kuhlman B, Bradley P. Advances in protein structure prediction and design. Nat Rev Mol Cell Biol. 2019;20(11):681-697. https://doi.org/10.1038/s41580-019-0163-x
3. Newberry RW, Raines RT. Secondary forces in protein folding. ACS Chem Biol. 2019;14(8):1677-1686. https://doi.org/10.1021/acschembio.9b00339
4. Sruthi CK, Prakash M. Amino acid impact factor. PLoS One. 2018;13(6):e0198645. https://doi.org/10.1371/journal.pone.0198645
5. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science. 1999;286(5438):295-299. https://doi.org/10.1126/science.286.5438.295
6. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein–protein interaction by message passing. Proc Natl Acad Sci USA. 2009;106(1):67-72. https://doi.org/10.1073/pnas.0805923106
7. Anishchenko I, Ovchinnikov S, Kamisetty H, Baker D. Origins of coevolution between residues distant in protein 3D structures. Proc Natl Acad Sci USA. 2017;114(34):9122-9127. https://doi.org/10.1073/pnas.1702664114
8. Cheng Y, Grigorieff N, Penczek PA, Walz T. A primer to single-particle cryo-electron microscopy. Cell. 2015;161(3):438-449. https://doi.org/10.1016%2Fj.cell.2015.03.050
9. Doerr A. Cryo-electron tomography. Nat Methods. 2017;14:34. https://doi.org/10.1038/nmeth.4115
10. Dubochet J, McDowall AW. Vitrification of pure water for electron microscopy. J Microsc. 1981;124(3):3-4. https://doi.org/10.1111/j.1365-2818.1981.tb02483.x
11. Adrian M, Dubochet J, Lepault J, McDowall AW. Cryo-electron microscopy of viruses. Nature. 1984;308(5954):32-36. https://doi.org/10.1038/308032a0
12. Yip KM, Fischer N, Paknia E, Chari A, Stark H. Breaking the next Cryo-EM resolution barrier–atomic resolution determination of proteins!. BioRxiv. 2020:106740. https://doi.org/10.1101/2020.05.21.106740
13. Callaway E. 'It opens up a whole new universe': revolutionary microscopy technique sees individual atoms for first time. Nature. 2020;582(7811):156-157. https://doi.org/10.1038/d41586-020-01658-1
14. Adhikari P, Li N, Shin M, Steinmetz NF, Twarock R, Podgornik R, et al. Intra-and intermolecular atomic-scale interactions in the receptor binding domain of SARS-CoV-2 spike protein: implication for ACE2 receptor binding. Phys Chem Chem Phys. 2020;22(33):18272-18283. https://doi.org/10.1039/D0CP03145C
15. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367(6483):1260-1263. https://doi.org/10.1126/science.abb2507
16. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cell Mol Immunol. 2020;17(6):613-620. https://doi.org/10.1038/s41423-020-0400-4
17. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265-269. https://doi.org/10.1038/s41586-020-2008-3
18. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020;395(10224):565-574. https://doi.org/10.1016/s0140-6736(20)30251-8
19. Wang J. Fast identification of possible drug treatment of

https://doi.org/10.61577/jbs.2024.100002

coronavirus disease-19 (COVID-19) through computational drug repurposing study. J Chem Inf Model. 2020;60(6):3277-3286. https://doi.org/10.1021/acs.jcim.0c00179
20. Zou J, Yin J, Fang L, Yang M, Wang T, Wu W, et al. Computational prediction of mutational effects on SARS-CoV-2 binding by relative free energy calculations. J Chem Inf Model. 2020;60(12):5794-5802. https://doi.org/10.1021/acs.jcim.0c00679
21. Smith M, Smith JC. Repurposing therapeutics for COVID-19: supercomputer-based docking to the SARS-CoV-2 viral spike protein and viral spike protein-human ACE2 interface. ChemRxiv. 2020;1-29. https://doi.org/10.26434/chemrxiv.11871402.v4
22. Adhikari P, Ching WY. Amino acid interacting network in the receptor-binding domain of SARS-CoV-2 spike protein. RSC Adv. 2020;10(65):39831-39841. https://doi.org/10.1039/D0RA08222H
23. Ching WY, Rulis P. Electronic structure methods for complex materials: the orthogonalized linear combination of atomic orbitals. Oxford: OUP Oxford; 2012. https://doi.org/10.1093/acprof:oso/9780199575800.001.0001
24. Jawad B, Adhikari P, Cheng K, Podgornik R, Ching WY. Computational design of miniproteins as SARS-CoV-2 therapeutic inhibitors. Int J Mol Sci. 2022;23(2):838. https://doi.org/10.3390%2Fijms23020838
25. Adhikari P, Jawad B, Rao P, Podgornik R, Ching WY. Delta variant with P681R critical mutation revealed by ultra-large atomic-scale ab initio simulation: implications for the fundamentals of biomolecular interactions. Viruses. 2022;14(3):465. https://doi.org/10.3390/v14030465
26. Jawad B, Adhikari P, Podgornik R, Ching WY. Key interacting residues between RBD of SARS-CoV-2 and ACE2 receptor: combination of molecular dynamics simulation and density functional calculation. J Chem Inf Model. 2021;61(9):4425-4441. https://doi.org/10.1021/acs.jcim.1c00560
27. Adhikari P, Jawad B, Podgornik R, Ching WY. Mutations of omicron variant at the interface of the receptor domain motif and human angiotensin-converting enzyme-2. Int J Mol Sci. 2022;23(5):2870. https://doi.org/10.3390%2Fijms23052870
28. Ching WY, Adhikari P, Jawad B, Podgornik R. Effect of delta and omicron mutations on the RBD-SD1 domain of the spike protein in SARS-CoV-2 and the omicron mutations on RBD-ACE2 interface complex. Int J Mol Sci. 2022;23(17):10091. https://doi.org/10.3390/ijms231710091
29. Adhikari P, Jawad B, Podgornik R, Ching WY. Quantum chemical computation of omicron mutations near cleavage sites of the spike protein. Microorganisms. 2022;10(10):1999. https://doi.org/10.3390%2Fmicroorganisms10101999
30. Ching WY, Adhikari P, Jawad B, Podgornik R. Towards quantum-chemical level calculations of SARS-CoV-2 spike protein variants of concern by first principles density functional theory. Biomedicines. 2023;11(2):517. https://doi.org/10.3390/biomedicines11020517
31. Cao L, Goreshnik I, Coventry B, Case JB, Miller L, Kozodoy L, et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. Science. 2020;370(6515):426-431. https://doi.org/10.1126/science.abd9909
32. Schütz D, Ruiz-Blanco YB, Münch J, Kirchhoff F, Sanchez-Garcia E, Müller JA. Peptide and peptide-based inhibitors of SARS-CoV-2 entry. Adv Drug Deliv Rev. 2020;167:47-65. https://doi.org/10.1016/j.addr.2020.11.007
33. Ching WY, Adhikari P, Jawad B, Podgornik R. Ultra-large-scale ab initio quantum chemical computation of bio-molecular systems: the case of spike protein of SARS-CoV-2 virus. Comput Struct Biotechnol J. 2021;19:1288-1301. https://doi.org/10.1016%2Fj.csbj.2021.02.004

https://doi.org/10.61577/jbs.2024.100002