Article @ Virology

In silico Analysis of Interaction between SARS-CoV-2 Membrane Protein and Human AKR1C2 Protein Revealed Universally Conserved Binding Site

Vivek Darapaneni[1]; Anusha Jaldani[2]*

1. *Department of virology and computational biochemistry, Anvek Institute of Biomolecular Research, Visakhapatnam, India.*
2. *Department of computational pharmacology, Anvek Institute of Biomolecular Research, Visakhapatnam, India.*

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for an ongoing COVID-19 pandemic that has devastated mankind. During this pandemic, it has been observed that the mortality rate in men is markedly higher than that in women. The SARS-CoV-2 membrane protein plays a decisive role in the viral life cycle. In infected people, membrane protein impedes the conversion of testosterone from active form to its inactive form via its interaction with human Aldo-keto reductase family 1 member C2 protein. This leads to the high availability of active testosterone which in turn promotes the SARS-CoV-2 entry into the host cell. In the present study, in silico analysis of interaction between membrane protein and Aldo-keto reductase family 1 member C2 protein and conservation analysis of 16,39,480 SARS-CoV-2 membrane protein revealed novel universally conserved binding site. Targeting this conserved binding site with small drug-like molecules would inhibit the interaction which leads to inhibition of SARS-CoV-2 entry into the host cell.

Article history: Submitted: 04/02/2022; Revised: 01/03/2022; Accepted: 18/03/2022

DOI:10.21092/jav.v11i1.96

Key Words: SARS-CoV-2, Membrane protein, AKR1C2, Conserved, Testosterone, Binding site.

Abbreviations: SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2; AKR1C2, Aldo-keto reductase family 1 member C2; M, Membrane; NSP, non-structural protein

* Corresponding author, e-mail: vivek110385@gmail.com

Introduction

Coronaviruses are enveloped viruses with a large positive strand RNA genome. Coronaviruses have been known to infect humans, bats, civets, cattle, horses, swine, dogs, cats, turkeys, rabbits, chickens, rats and mice. Coronaviruses cause gastroenteritis and respiratory tract diseases in their hosts [1]. Previously HKU1, NL63, 229E and OC43 have been known to cause mild respiratory disease in humans. In the year 2002 a novel coronavirus has crossed over from bats to humans through palm civet cats [intermediary host] [2]. Again a decade later in 2012 another novel coronavirus has crossed over from bats to humans through dromedary camels [intermediary host] [3]. The former virus was named as severe acute respiratory syndrome coronaviruses (SARS--CoV), while the later virus was named as middle east respiratory syndrome coronaviruses (MERS-CoV). In the year 2019 yet again a novel coronavirus had crossed over from bat to humans via malaya pangolin [intermediary host] [4]. This virus was designated as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). SARS-CoV-2 causes severe respiratory tract infection and causes a global pandemic resulting in deaths worldwide. As of 7 January, 2022, approximately 299 million confirmed cases of SARS-CoV-2 infection were reported globally. Of which approxi--mately 5.5 million deaths were reported globally [5].

SARS-CoV-2 genome encodes four main structural proteins and sixteen non-structural proteins [6]. Coronavirus protein expression in the infected cell occurs in two different and distinguishable phases. In the first phase, the 5'-terminal region (≈ two thirds) of the viral genome is translated into sixteen non-structural proteins, NSP1–NSP16. In the second phase, the 3'-terminal region (≈ one third) of the viral genome is translated into four structural proteins namely, spike, envelope, membrane and nucleocapsid proteins [6].

In this ongoing pandemic, it has been established that mortality rate in males is over and above that of females due to COVID-19 infection. This has led the scientists to speculate that testosterone might play a crucial role in disease severity. SARS-CoV-2 has previously been shown to reduce serum testosterone levels [7]. Recently it has been shown that the membrane (M) protein of SARS-CoV-2 plays a critical role in availability of active testosterone via interaction with the enzyme Aldo-keto reductase family 1 member C2 (AKR1C2) [8].

Till date, there are no studies which characterize the interaction between SARS-CoV-2 M protein and human AKR1C2 enzyme. In the present study, by analyzing the structural interaction between M protein and AKR1C2 enzyme, we identified the residues involved in the M-ARK1C2 interaction. Targeting the

M-ARK1C2 interaction region with small drug-like molecules would inhibit the M-AKR1C2 interaction.

Methods

1. Protein structures of M protein and AKR1C2

The experimental protein structure of AKR1C2 protein was retrieved from protein data bank (PDB) (https://www.rcsb.org/) [9] with the PDB ID 4XO6 [10]. The PDB file consists of two chains namely, chain A and chain B, of which chain B was removed. To date the experimental structure of M protein of SARS-CoV-2 is not elucidated. Therefore the structure of M protein was predicted by the Robetta server using the RoseTTAFold method (https://robetta.bakerlab.org/submit.php) [11]. RoseTTAFold uses a three-track network, that is successful transformation and integration of information at sequence level (one dimensional), distance map level (two dimensional) and coordinate level (three dimensional) [11]. The confidence score given by the server ranges from 0 to 1. "0" means a bad model whereas "1" means a good model [11].

2. M protein - AKR1C2 protein docking

Protein-protein docking of M and AKR1C2 proteins was carried out using HDOCK server (http://hdock.phys.hust.edu.cn/) [12] which is based on a hybrid algorithm of template based modeling and ab initio free docking. Protein structure (.pdb) files were given as input to the HDOCK server [12].

3. M protein Interface residues evolutionary analysis

Full length sequences of SARS-CoV-2 M protein were downloaded from National Centre for Biotechnology Information (NCBI) SARS-CoV-2 resource (https://www.ncbi.nlm.nih.gov/sars-cov-2/). Sequences with ambiguous characters were removed. Redundant sequences were removed using CD-HIT (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi) [13]. Multiple sequence alignment was performed using MAFFT version 7 (https://mafft.cbrc.jp/alignment/software/) [14]. By giving multiple sequence alignment file as an input, conserved and variable regions in M protein were identified using the ConSurf server. (https://consurf.tau.ac.il/)[15]. ConSurf algorithm produces meaningful conservation scores by taking evolutionary relationships among protein sequences into account. The conservation score given by the ConSurf server is divided into nine grades. Most variable positions in the protein are placed in grade 1, intermediately conserved positions are placed in grade 5, and most conserved positions are placed in grade 9 [15].

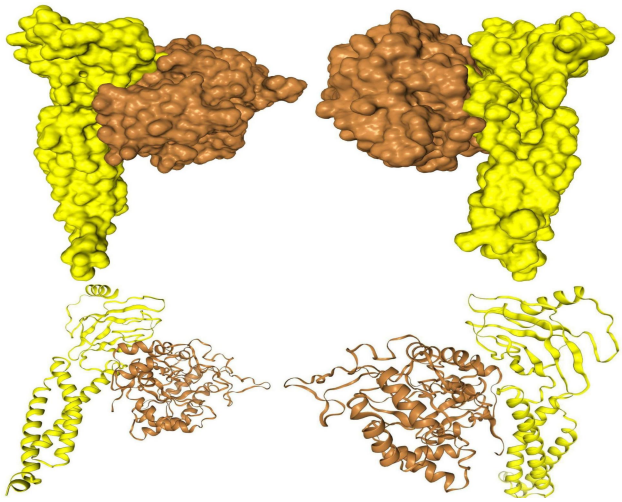
Results

1. M protein - AKR1C2 protein structures and interaction

The structure of SARS-CoV-2 M protein was predicted by the Robetta server using

the RoseTTAFold method [11]. The obtained structure had a confidence score of 0.74, which shows the obtained model is reliable. The experimental structure of human AKR1C2 was downloaded from PDB [9]. Both the structures are shown in figure 1.

The interaction between SARS-CoV-2 M protein (predicted structure) and AKR1C2 protein (experimental structure) was elucidated using the HDOCK server [12]. In total ten top ranking models were obtained as output which are shown in table 1. Model 1 was selected which has lowest docking energy score and highest ligand root mean square deviation shown in figure 1.



Figrue 1:M protein (yellow colour) - AKR1C2 protein (brown colour) interaction obtained using HDOCK server [12].

Table 1. Top 10 ranking models of M - AKR1C2 proteins interaction obtained using the HDOCK server [12].

Rank	Docking energy score (kcal/mol)	Ligand RMSD (Å)	Rank	Docking energy score (kcal/mol)	Ligand RMSD (Å)
1	-311.27	58.52	6	-265.85	66.69
2	-281.16	63.06	7	-264.75	82.93
3	-278.12	70.54	8	-262.67	84.77
4	-274.76	87.66	9	-261.88	39.25
5	-274.42	61.70	10	-261.51	91.48

2. Interface residues involved in M-AKR1C2 protein interaction

The interface residues involved in M-AKR1C2 protein interaction are identified by the HDOCK server [12] are shown in table 2 and figure 1.

3. Evolutionary analysis of interface residues of M protein

A total of 16,39,480 M protein sequences were downloaded from the database. After the removal of similar sequences, a total of 1907 unique sequences were obtained. The obtained sequences were then aligned using MAFFT version 7 [14]. Conservation analysis of SARS-Cov-2 M protein was carried out using ConSurf server[15] and shown in figure 2. Highly conserved (no amino acid variation in the sequence analyzed), conserved and variable residues in the interface region of M protein are shown in table 2.

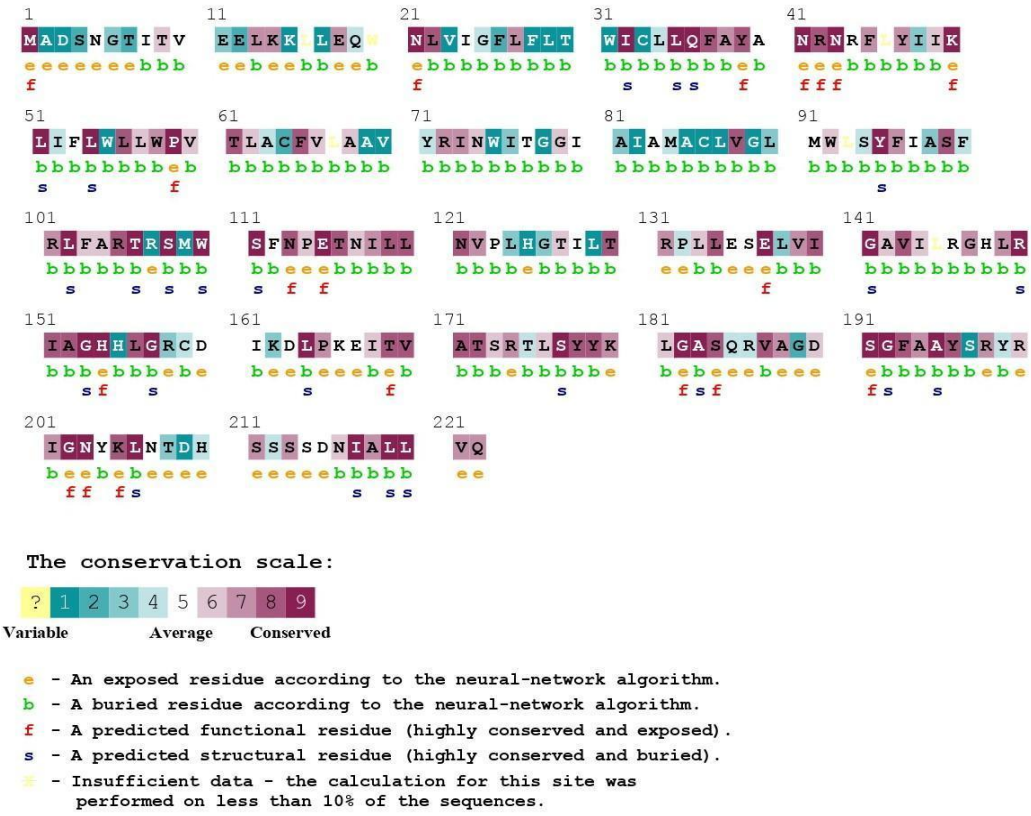
Table 1.Interface residues of M-AKR1C2 protein interaction.
Conserved residues are bolded. "*" indicate highly conserved residue.

Interface residues in M protein	Interface residues in AKR1C2 protein
Trp92, **Tyr95***, **Phe96**, Ile97, Phe100, **Phe103**, **Ala104**, Arg107,**Ser108***,Met109, **Trp110*,** Phe112, **Asn113**, **Thr127**, **Arg131**, **Leu134**, His148, **Arg150***, **Gly153**, **His154**, His155, **Leu156**, **Gly157***, Arg158	LYS4, TYR5, GLN6, CYS7, VAL8, LYS9, VAL18, ARG200, LYS201, LEU202, ASP204, PHE205, LYS207, SER208, ASP210, ARG258, LEU261, GLN262, ARG263, GLY264, PHE284, GLU285, GLN287, SER290, MET293, LYS294, TYR323

Discussion and Conclusion

SARS-Cov-2 has caused one of the most devastating pandemics faced by mankind. In the infectious life cycle of SARS-Cov-2, M protein plays an important role. In this ongoing pandemic mortality rate in males is far more than that of females due to SARS-Cov-2 infection. This is due to the fact that testosterone plays a crucial role in disease severity. SARS-CoV-2 infection reduces serum testosterone levels in the infected individuals. Recently it has been shown that the M protein of SARS-CoV-2 plays a critical role in availability of active testosterone via interaction with the enzyme AKR1C2[8]. In the present study, by analyzing the structural interaction we identified the residues involved in the M-ARK1C2 interaction.

A total of twenty four residues in the M protein and twenty three residues in the AKR1C2 protein are involved in the



Figrue 2:The evolutionary conservation of the SARS-Cov-2 M protein obtained from ConSurf server [15].

interaction. Of the twenty four residues identified in the M protein, sixteen residues are conserved, three residues are intermediately conserved and five residues are variable. Residues Trp92, Tyr95, Phe96, Phe103, Ala104, Ser108, Trp110, Asn113, Thr127, Arg131, Leu134, Arg150, Gly153, His154, Leu156 and Gly157 are important in the interaction with AKR1C2 protein, as these residues are conserved among different variants of SARS-CoV-2. Of these residues, Tyr95, Ser108, Trp110, Arg150 and Gly157 might play pivotal roles in the interaction as these residues are highly conserved. The conservation pattern in the interface region of M protein shows that the M-AKR1C2 interaction is universally conserved among different variants of SARS-CoV-2. Due to this the M-AKR1C2 interaction region of M protein could be a potential drug target region of M protein.

This study for the first time identified the residues involved in the M-ARK1C2 interaction. The ARK1C2 interaction region in M protein is conserved across various variants of SARS-CoV-2. Targeting the M-ARK1C2 interaction region with small drug-like molecules would inhibit the M-AKR1C2 interaction.

Competing interests

The authors declare all financial and non-financial competing interests.

Reference

[1]. Guy J S, Breslin J J, Breuhaus B, et al. Characterization of a coronavirus isolated from a diarrheic foal[J]. Journal of clinical microbiology, 2000, 38(12): 4523-4526.

[2]. Chan - Yeung M, Xu R H. SARS: epidemiology[J]. Respirology, 2003, 8: S9-S14.

[3]. Mohd H A, Al-Tawfiq J A, Memish Z A. Middle East respiratory syndrome coronavirus (MERS-CoV) origin and animal reservoir[J]. Virology journal, 2016, 13(1): 1-7.

[4]. Andersen K G, Rambaut A, Lipkin W I, et al. The proximal origin of SARS-CoV-2[J]. Nature medicine, 2020, 26(4): 450-452.

[5]. WHO coronavirus disease (COVID-19) dashboard. Available: https://covid19.who.int/. Date accessed: January, 2022.

[6]. V ' kovski P, Kratzel A, Steiner S, et al. Coronavirus biology and replication: implications for SARS-CoV-2[J]. Nature Reviews Microbiology, 2021, 19(3): 155-170.

[7]. Bienvenu L A, Noonan J, Wang X, et al. Higher mortality of COVID-19 in males: sex differences in immune response and cardiovascular comorbidities[J]. Cardio-vascular Research, 2020, 116(14): 2197-2206.

[8]. Darapaneni V, Jaldani A. Membrane protein of SARS-CoV-2 plays a pivotal role in the availability of active testosterone through its interaction with AKR1C2 enzyme leading to the upregulation of TMPRSS2 protease expression[J]. Microbiology Independent Research Journal (MIR Journal), 2021, 8(1): 38-40.

[9]. Berman H M, Westbrook J, Feng Z, et al. The protein data bank[J]. Nucleic acids research, 2000, 28(1): 235-242.

[10]. Zhang B, Hu X J, Wang X Q, et al. Human 3α -hydroxysteroid dehydrogenase type 3: structural clues of 5 α -DHT reverse binding and enzyme down-regulation decreasing MCF7 cell growth[J]. Biochemical Journal, 2016, 473(8): 1037-1046.

[11]. Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a three-track neural network[J]. Science, 2021, 373(6557): 871-876.

[12]. Yan Y, Tao H, He J, et al. The HDOCK server for integrated protein – protein docking[J]. Nature protocols, 2020, 15(5): 1829-1852.

[13]. Huang Y, Niu B, Gao Y, et al. CD-HIT Suite: a web server for clustering and comparing biological sequences[J]. Bioinformatics, 2010, 26(5): 680-682.

[14]. Katoh K, Rozewicki J, Yamada K D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization[J]. Briefings in bioinformatics, 2019, 20(4): 1160-1166.

[15]. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules[J]. Nucleic acids research, 2016, 44(W1): W344-W350.