In silico computation of coagulation factor II: a potential water treatment agent against gram negative bacteria

Nadya NATASYA, Tyniana TERA, Sava SAVERO, Nadine SWASTIKA, Arli Aditya PARIKESIT*

*Indonesia International Institute for Life Sciences, Bioinformatics Department, Pulomas Barat Kav. 88, East Jakarta, 13210, Indonesia; nadya.natasya@student.i3l.ac.id; tyniana.tera@student.i3l.ac.id; sava.savero@student.i3l.ac.id; nadine.melody@student.i3l.ac.id; arli.parikesit@i3l.ac.id (*corresponding author)*

Abstract

Water as one of the main sources of life, holds an important aspect to public health and safety. Up until now there have been many concerns about water pollution especially in developing countries. Heavy polluted water that is not treated well could cause many concerning diseases that can lead to deaths. Contaminants that came from chemical, physical, and biological compounds are commonly found in the water sources. Gram negative bacteria (GNB) has been shown to develop multiple drugs and antibiotics resistant, making the infection more fatal. This has become a major concern on public health especially in water treatment as it becomes more challenging. Our study investigates human coagulation factor II that is responsible for blood clotting as a possible method for water treatment against GNB. By investigating the coagulation protein interaction with several bacterial lipopolysaccharides proteins and calculating the binding affinity of the interaction, the results show factor II has a lower binding affinity compared to previously studied factor VII. This shows possibilities of factor II to hydrolyse several gram-negative bacteria to act as a potential treatment against these GNBs.

Keywords: coagulation factor; docking; FII; FVII; GNB; water

Introduction

Water is one of the main sources of life; every activity is related to water. Most people in big cities rely on treated water pipes, however unfortunately 663 million people around the world still rely on rivers or other water sources to do most of their activities such as cooking, bathing, drinking, and many more (WHO/UNICEF Joint Water Supply, 2015). The situation right now, especially in developing countries like Indonesia, is getting worse. The water has gone darker with less living creatures. People do not care about the water's health and still throw trash into the rivers. Most of the wastewater discharged untreated directly into the water. The rivers get more toxic and have a high level of contamination, especially from chemicals, bacteria, feces, metal, etc. These contaminations are oftentimes the root cause for waterborne diseases such as diarrhea, scabies, cholera, dysentery, hepatitis A, typhoid, polio, etc., which are responsible for 1.2 million deaths every

year (Hannah and Max, 2019; Hatami, 2013), becoming a major cause of mortality and morbidity (Cissé, 2019).

Specifically, gram negative bacteria (GNB) have become an increasingly concerning public health problem. Its resistance to multiple drugs and increasing resistance to available antibiotics make infection far more fatal especially in countries with poor medical infrastructure. To add to that, its ability to resist various drugs, also makes chemical-based water treatment a more challenging process.

A hundred to two hundred thousand of deaths were caused by *Vibrio cholerae* bacteria alone. This cholera infection occurs through the consumption of contaminated water or food that can lead to death within 12 hours of first symptoms (Conner *et al.*, 2016). Salmonella disease is caused by *Salmonella enterica* that first only infect animal farms but now can infect humans. While the *Escherichia coli* and *Shigella* sp are closely related. *E. coli* can cause severe diarrhea, vomiting, and stomach cramp while *Shigella* sp. can cause shigellosis infection through contaminated water or food and feces.

Gram-negative bacteria have an outer membrane composed of complex lipopolysaccharide, which is unique to the bacterial outer membrane for cell viability. The structure of lipopolysaccharide is important for the structural integrity of the gram-negative bacteria's outer membrane, as the gram-negative bacteria's survivability against chemical attack is dependent on it. The outer membrane also contains antigens to defend itself against biological attack. The disruption/destruction of the outer membrane would make the gram-negative bacteria prone to attacks, resulting in the death of the organism.

Recently, a paper from Chen *et al.* (2019) discussed the interaction of coagulation factors VII, IX, and X against the lipopolysaccharide. The paper discussed that they found that the light chain of coagulation factor VII, which is naturally used as a coagulation factor on an event bleeding occurs, can hydrolyse the lipopolysaccharide of extensively drug-resistant *Escherichia coli*. The paper also tested the light chain coagulation factor VII on other extensively drug-resistant gram-negative bacteria with positive results in the ability to hydrolyse the lipopolysaccharide of the tested bacteria.

The goal of this study is to explore and compare the ability of other coagulation factors to hydrolyse lipopolysaccharides of various gram-negative bacteria with molecular simulation tools. This is done to see the potential of other coagulation factors as well as its ability to hydrolyse multiple bacterial species. This process is done by molecular docking the light chain coagulation factors VII and II with the lipopolysaccharide to determine the interaction between the molecules. With light chains serving as a standard as its efficacy is previously known.

Materials and Methods

The flowchart of our methodology could be found in the Figure 1.

Protein structure

The bacterial lipopolysaccharide, the coagulation factors VII, and II were obtained from the RCSB PDB database. The PDB id was shown in Table 1. The GNBs used were *Salmonella enterica*, *Vibrio cholerae*, *Escherichia coli*, and *Shigella flexneri*. The protein structure was search based on the keywords of bacteria name and protein structure.

Preliminary structure retrieval and processing

The bacterial lipopolysaccharide protein structure and the coagulation factor VII were obtained from the RCSB PDB database. The PDB id was shown in Table 1. The raw protein structure was cleaned of its unnecessary solvents and parts using PyMOL. The clean protein structure was then assessed for its angular viability by obtaining its Ramachandran favoured value from its Ramachandran plot, via the SWISS-MODEL server. Structures were deemed viable if it scored a Ramachandran favour >85%.

Table 1. Protein name and the PDB id

PDB ID	Name
1F7M	Human blood coagulation FVII, NMR, minimized average structure
4LXB	Human blood coagulation FII (light chain), X-Ray diffraction
4N4R	Salmonella enterica structure basis of lipopolysaccharide biogenesis
6MJP	Vibrio cholerae Lipopolysaccharide export system protein LptC
6MI7	Escherichia coli Lipopolysaccharide export system permease protein LptF
6S8N	Shigella flexneri Lipopolysaccharide ABC transporter, ATP-binding protein LptB

Protein interaction analysis

The protein-protein interactions analysis was done using the ClusPro (Kozakov *et al.*, 2017) web server. The downloaded pdb file for both proteins, as seen on Table 1, were imported to the web server and the model selection was done based on its cluster size. The energy was recorded, and the visualization of the interaction was done using the PyMOL application.

Binding affinity analysis

In order to comparatively assess the docking results between the combinations of bacterial species origin and coagulation factor, its binding affinity was computed. The PRODIGY server (Xue *et al.*, 2016) by Bonvin lab was used to get the binding affinity or ΔG in kcal mol^{-1}. The protein-protein binding was selected and the molecular models that were output of ClusPro were uploaded. In the two interactors molecule and ligand chain name was specified respectively.
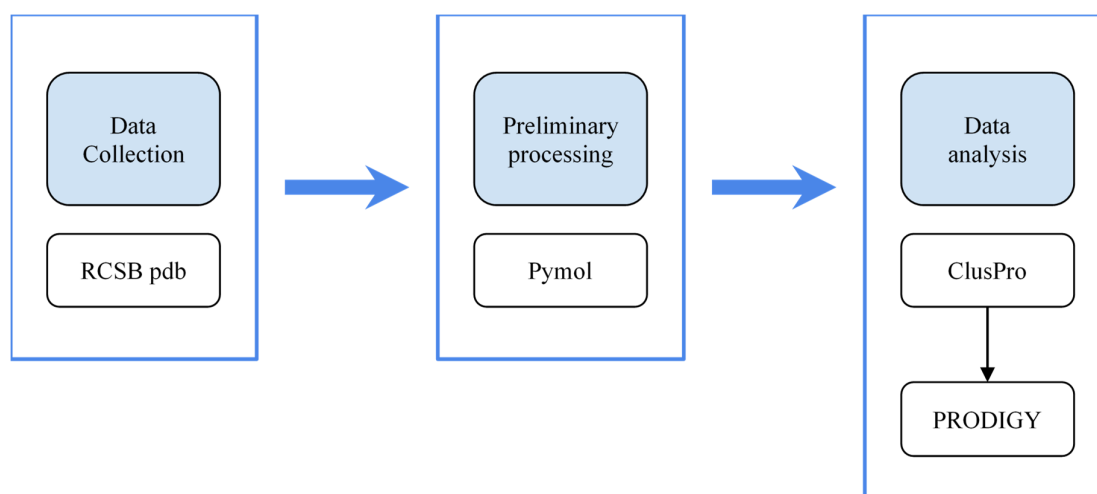
Figure 1. Workflow of the Molecular Simulation. The pipeline was taken and inspired from Arifin and Parikesit (2020); Lungidningtyas and Parikesit (2020)

The data were collected from RCSB pdb using the bacteria name and protein structure as the keyword for the GNB structure and the human coagulation factor VII or II for the coagulation factors. Preliminary processing was done beforehand to remove unwanted protein from the lipopolysaccharide structure of the bacteria. The analysis was done using the ClusPro webserver to compare the docking between the protein and PRODIGY webserver was used to retrieve the binding affinity.

Results and Discussion

Preliminary processing

Assessment of its structure quality using a Ramachandran plot on the SWISS MODEL server, evaluates the phi and psi backbone dihedral angles in a protein structure. The results indicate that almost all molecular structures had a Ramachandran favoured of >85%. Meaning that in each structure at least 85% of its regions have no steric clashes and are deemed allowed regions. However, the light chain coagulation factor VII had a Ramachandran favoured below 85%. Nevertheless, it was still used as it served as a comparative with a known efficacy.

Table 2. Ramachandran favoured values of macromolecules respectively

PDB ID	Ramachandran favoured
1F7M (light chain)	70.45%
4LXB (light chain)	97.56%
4N4R (lptD&E)	90.87%
6MJP (lptC)	93.60%
6MI7 (lptF)	89.04%
6S8N (lptB)	88.37%

Protein interaction analysis

ClusPro is a web server program for protein structure docking. The ligand was rotated 70,000 times, each rotation with x,y,z relative to the receptor on a grid. A total of 1000 rotations or translation combinations with the best score was chosen (lowest score). Then a greedy clustering algorithm was performed for the 1000 ligand position with a 9 Å C-alpha RMSD radius. Ligand positions with the most neighbours in 9 Å were found and became the cluster center, while the neighbours became the members of the clusters (Kozakov *et al.*, 2017). The first output is then removed from the set and the program will look for another cluster and so on.

The representative structure for the cluster is the ligand with the most neighbours in the cluster center. The output (Table 3 and 4) is the list of the clusters of docked structures in the order of cluster size. For each cluster, the table shows the cluster size, the PIPER energy of the cluster center (i.e., the structure that has the highest number of neighbours structures in the cluster), and the energy of the lowest energy structure in the cluster.

The docking step uses PIPER (Kozakov *et al.*, 2006), a Fast Fourier Transform (FFT) correlation approach that is used for protein-protein docking (Kozakov *et al.*, 2006). FFT can evaluate the energy of a billion docked conformation numbers in the form of a correlation function (Kozakov *et al.*, 2006). In the ClusPro web service, the protein model was selected based on its cluster size. The energy does not directly affect the binding affinity, but low energy regions tend to generate a larger number of docked structures clusters. The larger the cluster size the higher the probability of the interaction. The energy landscape indirectly determines the most likely conformation of the complex.

The results of the ClusPro docking of the bacterial LPS parts and coagulation factors VII and II can be comparatively seen on Tables 3 and 4 respectively. Cluster sizes are almost all found to be bigger on the coagulation factor VII docking. However, there is one exception of *Vibrio cholera*, having a larger cluster size when docking with coagulation factor II. Indicating that coagulation factor docking and indirect ability to hydrolyse may vary depending on the bacterial species. A similar pattern in cluster sizes within the same coagulation factor can be seen, with the cluster size decreasing across *Vibrio cholera, Escherichia coli, Salmonella enterica*, and *Shigella flexneri*. This may indicate the varying affinities to different parts of the LPS structure. In terms of weighted score, overall, there is seen a lower weighted score in dockings with coagulation factor II.

Table 3. Model scores for the docking coagulation factor VII to bacteria species. with the energy of the cluster center (i.e., the structure that has the highest number of neighbours structures in the cluster), and the energy of the lowest energy structure in the cluster

Bacterial molecule origin	Cluster size	Representative	Weighted Score
Vibrio cholerae	194	Center	-747.9
		Lowest energy	-774.3
Shigella flexneri	109	Center	-667.4
		Lowest energy	-749.2
Escherichia coli	125	Center	-811.2
		Lowest energy	-945.9
Salmonella enterica	111	Center	-877.6
		Lowest energy	-987.9

Table 4. Model scores for the docking coagulation factor II to bacteria species. with the energy of the cluster center (i.e., the structure that has the highest number of neighbours structures in the cluster), and the energy of the lowest energy structure in the cluster

Bacterial molecule origin	Cluster size	Representative	Weighted Score
Vibrio cholerae	258	Center	-870.6
		Lowest energy	-941.3
Shigella flexneri	64	Center	-970.7
		Lowest energy	-970.7
Escherichia coli	94	Center	-835.5
		Lowest energy	-1044.7
Salmonella enterica	81	Center	-1069.9
		Lowest energy	-1352.3

The models with the highest cluster sizes obtained from ClusPro were then predicted of its binding affinities using the PRODIGY server (Table 5). There is a difference of binding affinities between coagulation factors and across bacterial species and/or LPS components. When looking between the coagulation factors, the docking with *Vibrio cholerae* and coagulation FVII had a lower binding affinity compared to FII. Whereas all other species and components had a lower binding affinity when docked with FII. When looking down through each coagulation factor there can be seen variable binding affinities between bacterial species origin and LPS component. However further comparative studies are necessary to indicate the tendencies more definitively between the variables of bacteria origin, LPS component and coagulation factor. Overall, coagulation factor II has superior binding affinities to a greater range of bacterial species and LPS components, indicated by a lower binding affinity when docked with *Shigella flexneri*, *Escherichia coli*, and *Salmonella enterica*.

Table 5. Binding affinities of the ClusPro docking output (FVII: coagulation factor VII; FII: coagulation factor II)

Bacterial molecule origin	Binding Affinity ΔG (kcal mol^{-1})	
	FVII	FII
Vibrio cholerae (lptC)	-12.7	-9.6
Shigella flexneri (lptB)	-10.1	-10.5
Escherichia coli (lptF)	-8.1	-12.0
Salmonella enterica (lptD&E)	-10.0	-11.1

The visualization of the docking models found via ClusPro can be seen in Figures 2-5. Each of the docking sights were unique to its coagulation factor and bacterial species pairing.
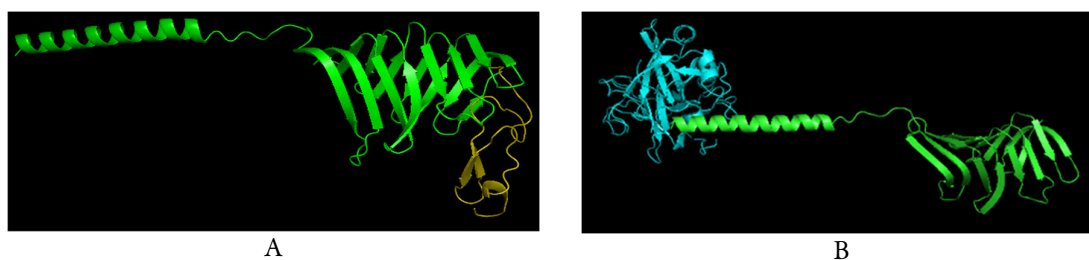

Figure 2. *Vibrio cholerae* lipopolysaccharide protein docking using PyMOL; (A) *Vibrio cholerae* lipopolysaccharide protein with coagulation factor VII (yellow); (B) *Vibrio cholerae* lipopolysaccharide protein with coagulation factor II (cyan)
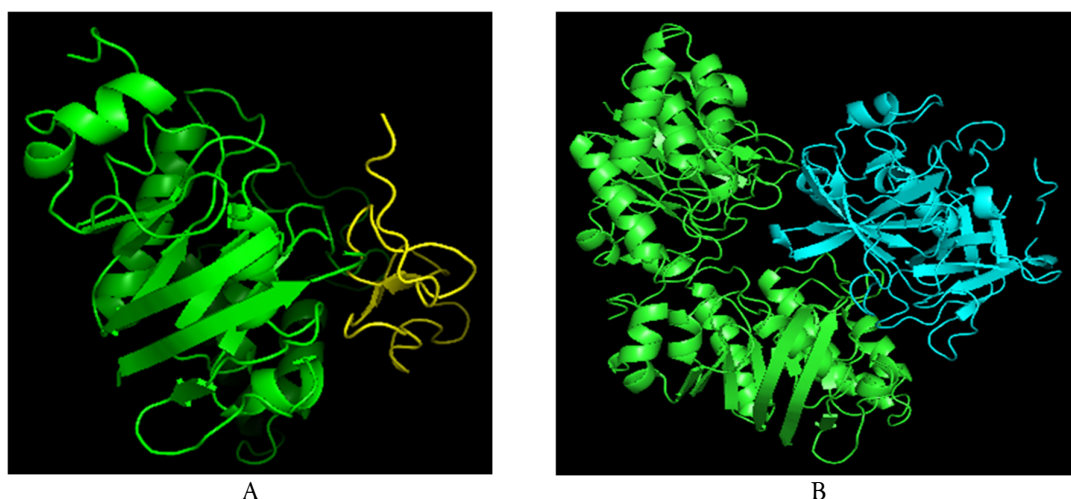

Figure 3. *Shigella flexneri* lipopolysaccharide protein docking using PyMOL; (A) *Shigella flexneri* lipopolysaccharide protein with coagulation factor VII (yellow); (B) *Shigella flexneri* lipopolysaccharide protein with coagulation factor II (cyan)
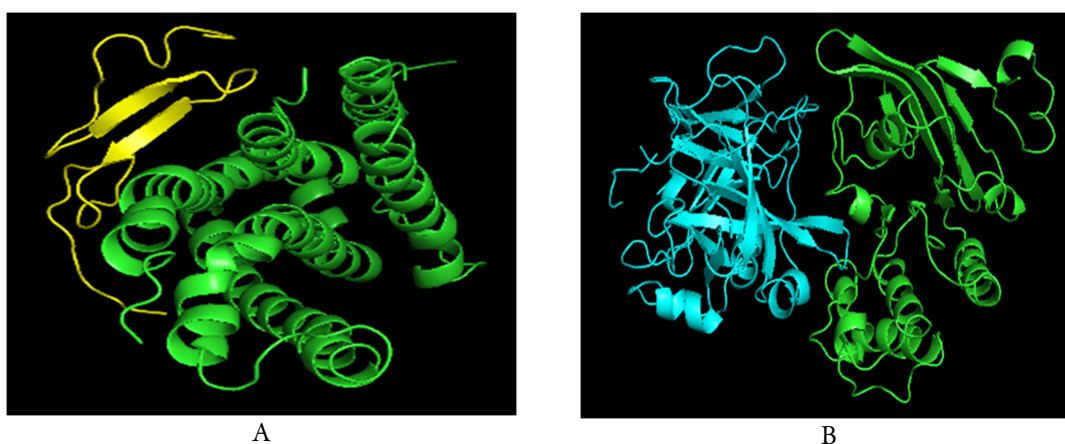

Figure 4. *Escherichia coli* lipopolysaccharide protein docking using PyMOL; (A) *Escherichia coli* lipopolysaccharide protein with coagulation factor VII (yellow); (B) *Escherichia coli* lipopolysaccharide protein with coagulation factor II (cyan)
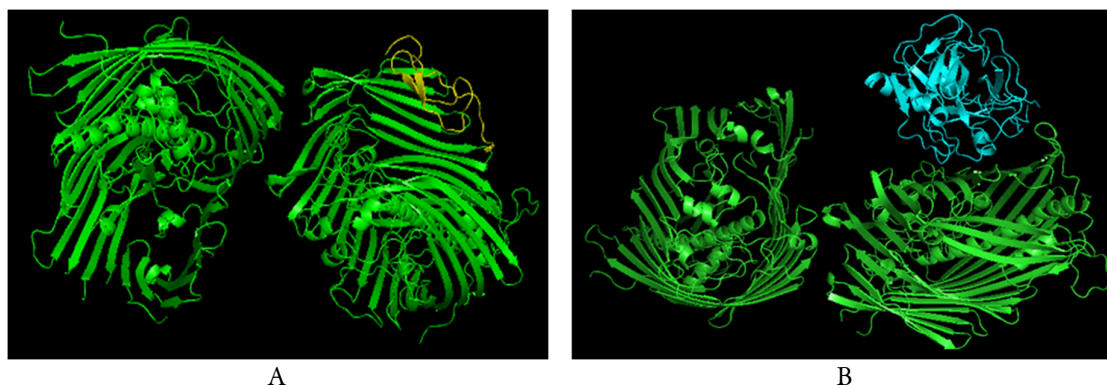
Figure 5. *Salmonella enterica* lipopolysaccharide protein docking using PyMOL; (A) *Salmonella enterica* lipopolysaccharide protein with coagulation factor VII (yellow); (B) *Salmonella enterica* lipopolysaccharide protein with coagulation factor VII (cyan)

The coagulation factor II (FII or F2) or prothrombin can be activated using prothrombinase to form thrombin that functions in thrombosis and hemostasis by transforming fibrinogen to fibrin during blood clot formation (Chinnaraj *et al.*, 2018). Other than maintaining vascular integrity during development and postnatal life, the peptides derived from the C-terminus of this protein have shown antimicrobial activity against *Escherichia coli* and *Pseudomonas aeruginosa* (F2 coagulation factor II, thrombin [*Homo sapiens* (human)], 2004). In humans, prothrombin is encoded by the F2 gene and synthesized in the liver by hepatocytes. Many other studies have been focusing on this gene, making the information clearer. The structure of the prothrombin has been previously determined (Chinnaraj *et al.*, 2018), opening possibilities to perform protein interaction analysis against different kinds of bacteria.

From the results, *Vibrio cholerae* have bigger cluster size in the FII than FVII but have a lower binding affinity. On the other hand, *Salmonella enterica*, *Shigella flexneri*, and *Escherichia coli* all have a smaller cluster size but bigger binding affinity. This shows that FII can be can act as an antimicrobial agent for several GNB like *Salmonella enterica*, *Shigella flexneri*, and *Escherichia coli*. However further research and experiment still need to be done to proof the capability of the FII.

With the abundance of water sources and the millions of people still dependent on it, any solution to combating the harmful microorganisms in these waters must be able to be produced for a global scale. The method of molecular cloning using recombinant plasmid could be used to mass produce this protein (Swiech *et al.*, 2017). These methods could produce many genes using a combination of bacteria plasmid and desired gene to produce numerous bacteria containing the combination plasmid. Thus, aiding in the production of the coagulation factors.

When introducing a foreign substance to bodies of water or having it be involved in a filtration process, the toxicity of the molecule must be considered. Coagulation factors VII and II are both naturally occurring molecules synthesized by the human body, thus assumptions of the lack of toxicity can be made (Palta *et al.*, 2014; Chinnaraj *et al.*, 2018; Chen *et al.*, 2019). However, further analysis regarding the toxicity of its recombinant molecule must be done to ensure safety.

Conclusions

Millions of people are still dependent on natural water sources for their livelihood, however as the quantity of waste and lack of consciousness increase, these imperative water source conditions continue to rapidly decline. With harmful microorganisms flourishing and creating further disease-ridden communities

with already poor healthcare infrastructures. This study attempts to explore the ability of coagulation factors to hydrolyse a range of gram-negative bacteria. It was found that coagulation factor II may have a greater potential in showing antimicrobial activities in comparison to the already known FVII, based on its lower binding affinities. Means to mass produce this protein to be used on a global scale can possibly be done via plasmid recombination. Furthermore, although its toxicity could not be tested, because coagulation factors are naturally occurring in the human body, its lack of toxicity can be assumed. However further analysis of the toxicity of its recombinant version must be conducted.

Authors' Contributions

The contributions of authors to the manuscript are specified as follows: Conceptualization: NN, TT, AAP; Data curation: NN, TT, SS; Formal analysis: NN, TT; Funding acquisition: AAP; Investigation: NN, TT, SS, NS, AAP; Methodology: NN, TT, AAP; Project administration: AAP; Resources: AAP; Software: NN, TT, SS, NS; Supervision AAP; Validation: NN; Visualization: NN, TT; Writing-original draft: NN, TT, SS, NS, AAP; Writing-review and editing: NN, AAP. All authors read and approved the final manuscript.

Acknowledgements

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

Arifin MZ, Agustriawan D, Parikesit AAP (2020). Molecular simulation oF MDM2 and E6AP proteins as P53 regulator in cervical cancer. Biointerface Research in Applied Chemistry 10(4):5875-5879. *https://doi.org/10.33263/BRIAC104.875879*

Chen J, Li X, Li L, Zhang T, Zhang Q, Wu F, … Song X (2019). Coagulation factors VII, IX and X are effective antibacterial proteins against drug-resistant gram-negative bacteria. Cell Research 29(9):711-724. *https://doi.org/10.1038/s41422-019-0202-3*

Chinnaraj M, Planer W, Pozzi N (2018). Structure of coagulation factor II: molecular mechanism of thrombin generation and development of next-generation anticoagulants. Frontiers in Medicine *https://doi.org/10.3389/fmed.2018.00281*

Cissé G (2019). Food-borne and water-borne diseases under climate change in low-and middle-income countries: Further efforts needed for reducing environmental health exposure risks. Acta Tropica 194:181-188. *https://doi.org/10.1016/j.actatropica.2019.03.012*

Conner JG, Teschler JK, Jones CJ, Yildiz FH (2016) Staying alive: *Vibrio cholerae's* cycle of environmental survival, transmission, and dissemination. Microbiology Spectrum 4(2):4-2. *https://doi.org/10.1128/microbiolspec.VMBF-0015-2015*

EMBO (2016). Protein-protein interaction prediction using docking. Retrieved 2021 June 18 from *http://aidanbudd.github.io/ppisnd/trainingMaterial/allegraVia/PPI/tutorial_on_protein_docking.html*

Hatami H (2013). Importance of water and water-borne diseases: on the occasion of the world water day. International Journal of Preventive Medicine 4(3):243-245.

Kozakov D, Brenke R, Comeau SR, Vajda S (2006). PIPER: an FFT-based protein docking program with pairwise potentials. Proteins: Structure, Function, and Bioinformatics 65(2):392-406.

Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S (2017). The ClusPro web server for protein–protein docking. Nature Protocols 12(2):255.

Lungidningtyas A, Parikesit AA (2020). *In silico* analysis of ethanol binding activity in neuronal nicotinic acetylcholine receptors. Malaysian Journal of Applied Sciences 5(1):54-61. *https://doi.org/10.37231/myjas.2020.5.1.243*

National Library of Medicine (US) (2004). F2 coagulation factor II, thrombin [*Homo sapiens* (human)]. National Center for Biotechnology Information. Retrieved 2021 June 20 from *https://www.ncbi.nlm.nih.gov/gene/2147*

Our World in Data (2019). Clean Water. Retrieved 2021 June 20 from *https://ourworldindata.org/water-access*

Palta S, Saroa R, Palta A (2014). Overview of the coagulation system. Indian Journal of Anaesthesia 58(5):515. *https://doi.org/10.4103/0019-5049.144643*

Schrödinger LLC (2010). The PyMOL molecular graphics system. Version 2.3.5.

Swiech K, Picanço-Castro V, Covas DT (2017). Production of recombinant coagulation factors: Are humans the best host cells? Bioengineered 8(5):462-470.

Vangone A, Bonvin AM (2015). Contacts-based prediction of binding affinity in protein–protein complexes. Elife 4:e07454.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, ... Schwede T (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research 46:W296-W303.

WHO/UNICEF Joint Water Supply (2015). Sanitation monitoring programme. World Health Organization. Progress on sanitation and drinking water: 2015 update and MDG assessment.

Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A (2016). PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. Bioinformatics 32(23):3676-3678.