ANNALS OF COMMUNICATIONS IN MATHEMATICS
Volume 6, Number 3 (2023), 165-176
ISSN: 2582-0818

A FAREY WAVELET-BASED MATHEMATICAL MODEL FOR BIOLOGICAL SERIES

AHMED S. ALANAZE[1], GHADA R. ALQADHI[1], WEJDAN S. ALATAWI[1], BASHAIR M. ALENAZI[1], ABDULLAZIZ ALANAZI[1], ANOUAR BEN MABROUK[*,1,2,3]

ABSTRACT. This work lies in the whole biomathematics framework which has as general goal the solving of biological problems with mathematical tools. The main objective of the present paper is to predict the transmembrane helices of proteins using wavelet denoising techniques. As a case of matter, we particularly highlight the interest in solving the problem of localizing these helices. Indeed, these helices play a vital role in the human body, notably in photosynthesis, respiration, neuronal signaling, immune response, absorption of nutrition, and have an important link with drugs as receptors coupled to many proteins. However, due to technical constraints, the crystallization of these helices remains very complex, which limits the exploration of their structure [27]. To overcome these difficulties, different prediction tools have been developed, initially based on hydrophobicity.

In this paper, we serve the Farey wavelet as a last alternative mother wavelet constructed in [4] to develop a mathematical model suitable for protein series description. We precisely apply a new type of wavelets constructed recently in [4] to localize the and/or predict the position of the transmembrane proteins alpha-helices in a coronavirus strain. The results are compared to existing works for performance, accuracy, and efficiency.

1. INTRODUCTION

The major aim of this article is to show the efficiency of some types of wavelets to model transmembrane (TM) proteins. Such a goal has been the object of many studies due to the importance of TM proteins in the functioning of organisms. Briefly, a TM protein is known to span the entire biological membrane. Such type of proteins also precipitates and aggregates in water. In our case, we are interested in the so-called alpha-helical proteins, which constitute the major category of TM proteins. These alpha-helices are the main parts of communication between cells. This makes them important stations in the body and also makes them major victims of attacks from the exterior. [6, 7, 9, 10, 11, 13, 16, 19, 20, 26, 29, 30]).

To localize the alpha-helices segments, a first step is needed subject to a conversion of the protein series into numerical time series to be able to apply mathematical tools. The main objective of the present study is to use Wavelets to predict the position of TM helices

2000 *Mathematics Subject Classification.* 42C40.
Key words and phrases. Wavelets; Farey wavelet; Biological series; Modeling.
Received: June 25, 2023. Accepted: September 25, 2023. Published: October 31, 2023.
*Corresponding author.

(TMHs) in protein sequences using hydrophobicity scales, by adapting wavelet-denoised hydropathy signals. (See also [8, 9, 13, 14, 18, 19, 25, 20, 22, 24]).

In our approach, the protein sequence is first converted into a hydropathy signal using Kyte and Doolittle (The hydropathy index of an amino acid is a number representing the hydrophobic or hydrophilic properties of its side-chain. The numerical sequence is then subjected to mathematical filtering, followed by biochemical filtering.

In mathematical filtering, the wavelet transform is applied using a mother wavelet. Then, in the filtering step, the wavelet coefficients corresponding to high frequencies are turned to zero, and finally, the signal is reconstructed and around zero by subtraction of its mean value, producing a zero mean filtered hydropathy signal whose maximum is used in the subsequent treatment.

In the biochemical filtering step, a tentative list of TMHs is created first, by the selection of all amino acids having a positive filtered hydropathy, and for each transmembrane segment, the number of amino acids is computed.

Wavelets are mathematical tools developed since the last two decades of the last century and proved to be powerful in many domains from theoretical fields to applied ([1, 2, 3, 5, 11, 12, 15, 17, 21, 23, 28]).

Wavelets analyze signals by computing their wavelet transforms via convolution products of the signals with copies of the mother/father wavelet. This indicates the importance of designating a function as an analyzing wavelet (father or mother). Wavelets have in addition the ability to localize analyzed signals in both time and frequency, which allows their adaptability to nonstationarity, non-seasonality, and irregularity.

The main common and essential point between all these frameworks is that any wavelet analysis starts with a source function called the mother wavelet, which gives rise next to the wavelet basis, and the wavelet transform. In our present work, we serve the Farey wavelet constructed already in [4] to conduct a wavelet model for TM proteins.

The rest of the paper will be structured as follows. Section 2 is devoted to a brief review of the Farey wavelet analysis. Section 3 is devoted to our main results on the Farey wavelet analysis and modeling of a special type of proteins issued from a coronavirus strain. Section 4 is the conclusion.

2. FAREY WAVELET REVISITED

In the pure mathematical framework, a wavelet may be defined as a function $\psi \in L^2(\mathbb{R})$ for which there holds the following essential properties.

- An admissibility assumption stating that,

$$\mathcal{A}_\psi = \int_{\mathbb{R}^+} |\hat{\psi}(\omega)|^2 \frac{d\omega}{|\omega|} < \infty. \tag{2.1}$$

- A zero mean assumption,

$$\widehat{\psi}(0) = \int_{-\infty}^{+\infty} \psi(u)du = 0. \tag{2.2}$$

- Localization (in time and frequency),

$$\|\psi\|_2^2 = \int_{-\infty}^{+\infty} |\psi(u)|^2 du = 1. \tag{2.3}$$

- Many vanishing moments or oscillations,

$$p = 0, ..., m-1, \quad \int_{\mathbb{R}} \psi(t)t^p dt = 0. \tag{2.4}$$

Analyzing a function $f \in L^2(\mathbb{R})$ passes through its The continuous wavelet transform at a position $u \in \mathbb{R}$ and a scale $s > 0$ defined by

$$d_{u,s}(f) = \int_{-\infty}^{\infty} f(t)\psi_{u,s}(t)dt, \quad \forall\, u, s, \tag{2.5}$$

where

$$\psi_{s,u}(x) = \frac{1}{\sqrt{s}}\psi(\frac{x-u}{s}). \tag{2.6}$$

The original analyzed function f may be re-obtained from its CWT as

$$f(x) = \frac{1}{\mathcal{A}_\psi}\int\int_{\mathbb{R}} d_{u,s}(f)\psi\left(\frac{x-u}{s}\right)\frac{dsdu}{s^2}. \tag{2.7}$$

The discrete wavelet transform (DWT) is obtained simply by setting $s = 2^{-j}$ and $u = k2^{-j}$, $j, k \in \mathbb{Z}$, we get here the dyadic version. The DWT of f is

$$d_{j,k} = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt, \tag{2.8}$$

where $\psi_{j,k}(.) = 2^{-j/2}\psi(2^j. - k)$. The DWT $d_{j,k}$ is sometimes called the wavelet coefficient or the detail coefficient of f. As previously, the function f may be reconstructed via its wavelet series decomposition as

$$f = \sum_{j,k} d_{j,k}\psi_{j,k} \tag{2.9}$$

It holds, indeed, that the set $(\psi_{j,k})_{j,k\in\mathbb{Z}}$ constitutes an orthonormal basis of $L^2(\mathbb{R})$ and called wavelet basis.

One of the strongest tasks is that the mother wavelet ψ may give rise to a second father wavelet (called also scaling function) φ satisfying the so-called 2-scale relation

$$\varphi = \sum_k h_k\varphi_{1,k} \tag{2.10}$$

where the coefficients h_k are

$$h_k = \int_{\mathbb{R}} \varphi(x)\varphi_{1,k}(x)dx.$$

In this case, there holds for ψ that

$$\psi = \sum_k g_k\varphi_{1,k}, \qquad \text{with } |; g_k = (-1)^k h_{1-k}.$$

Using these facts, the decomposition of the function f into a series of wavelets may be detailed more by involving the function φ. Let

$$f = \underbrace{\sum_k a_{J_0,k}\varphi_{J_0,k}}_{A_{J_0}} + \sum_{j=J_0}^{+\infty}\underbrace{\sum_k d_{j,k}\psi_{j,k}}_{D_j}, \tag{2.11}$$

where $J_0 \in \mathbb{N}$ is an arbitrary parameter fixed at advance and called the minimal decomposition level. The first sum in the decomposition A_{J_0} is called the approximation of f at the level J_0 and it describes its global shape or global behavior of it. The coefficients

$$a_{J_0,k} = \int_{\mathbb{R}} f(t)\varphi_{J_0,k}(t)dt$$

are called approximation coefficients. The sum D_j is called the detail of f at the level j and it describes its hidden or fluctuated behavior. Backgrounds may be found in [1, 2, 3, 5, 11, 12, 15, 17, 23, 28].

The (father,mother) farey wavelets are introduced in [4] starting from the Faray map (slightly modified) as

$$\varphi(\xi) = \begin{cases} K_0 \dfrac{1+\xi}{1-\xi} & , \quad -1 \leq \xi \leq 0, \\ \\ K_0 \dfrac{1-\xi}{1+\xi} & , \quad 0 \leq \xi \leq 1, \end{cases} \tag{2.12}$$

where $K_0 = \dfrac{1}{(4\log 2 - 2)}$, and

$$\psi(\xi) = K_1 \begin{cases} \dfrac{1}{3}\dfrac{1+2\xi}{1-2\xi} & ; \quad -\dfrac{1}{2} \leq \xi \leq 0, \\ \\ \dfrac{1}{3}\dfrac{1-2\xi}{1+2\xi} - \dfrac{\xi}{1-\xi} & ; \quad 0 \leq \xi \leq \dfrac{1}{2}, \\ \\ \dfrac{\xi-1}{\xi} - \dfrac{1}{3}\dfrac{1-2\xi}{3-2\xi} & ; \quad \dfrac{1}{2} \leq \xi \leq 1, \\ \\ -\dfrac{1}{3}\dfrac{3-2\xi}{1-2\xi} & ; \quad 1 \leq \xi \leq \dfrac{3}{2}. \end{cases}$$

Here, $K_1 = \dfrac{1}{\sqrt{3 - 4\log 2}}$ is the normalization constant.

These functions satisfy indeed many useful properties always needed in wavelet analysis, which are resumed in the following proposition.

Proposition 2.1. *The following assertions hold.*

(1) $\widehat{\varphi}(0) = 1$

(2) $\varphi = \sum_{k\in\mathbb{Z}} h_k \varphi_{1,k}$, *where* $h_1 = h_{-1} = \dfrac{1}{3\sqrt{2}}$, $h_0 = \dfrac{1}{\sqrt{2}}$ *and 0 else.*

(3) $\psi = \dfrac{K_1}{K_0} \sum_{k\in\mathbb{Z}} g_k \varphi_{1,k}$, *with* $g_k = (-1)^{k-1} h_{1-k}$.

(4) The function $\widetilde{\psi}(x) = \psi(x) - c\chi_{]-1/2,3/2[}(x)$ *with* $c = \dfrac{2\log 2 - 1}{6}$, *is admissible with one vanishing moment.*

(5) The function $\Gamma_\varphi(\omega) = \sum_{k\in\mathbb{Z}} |\widehat{\varphi}(\omega + 2k\pi)|^2$ *is bounded on* $\mathbb{R}$ *(called the overlap function associated to the function* φ*).*

(6) Let $\Phi \in L^2(\mathbb{R})$ *be defined by its Fourier transform* $\widehat{\Phi} = \dfrac{\widehat{\varphi}}{\sqrt{\Gamma_\varphi}}$. *Then, the system* $(\Phi_k = \Phi(. - k))_{k\in\mathbb{Z}}$ *is orthonormal in* $L^2(\mathbb{R})$.

3. Main results: A Coronavirus processing

To conduct our Farey wavelet analysis of the biological series in hand, we recall firstly that in wavelet analysis, it is natural to seek from (2.11) a finite series decomposition at

some level $J > J_0$ and consider the so-called J-level decomposition

$$f_J(t) = A_{J_0}(t) + D_{J_0}(t) + D_{J_0+1}(t) + \cdots + D_J(t). \tag{3.1}$$

We propose to localize the α-helices in a TM protein. using the length of the TM helix counting 19 residues at least, which in turn are always hydrophobic. Kyte and Doolittle [22] developed a hydrophobicity algorithm based on the amino acid-free energy scale transfer between water and a specific organic solvent. The following table (Table 1) resumes the results of such an algorithm and yields for each amino acid a corresponding numerical value, which allows the conversion of protein series into numerical ones.

TABLE 1. Kyte-Doolittle Hydrophobicity scale for protein series conversion.

Amino Acid	Symbol	Kyte-Doolittle Scale	Category
Isoleucine	Ile(I)	+4.5	Hydrophobic
Valine	Val(V)	+4.2	Hydrophobic
Leucine	Leu(L)	+3.8	Hydrophobic
Phenylalanine	Phe(F)	+2.8	Hydrophobic
Cysteine	CySH(C)	+2.5	Hydrophobic
Methionine	Met(M)	+1.9	Hydrophobic
Alanine	Ala(A)	+1.8	Hydrophobic
Glycine	Gly(G)	-0.4	Neutral
Threonine	Thr(T)	-0.7	Neutral
Serine	Ser(S)	-0.8	Neutral
Tryptophan	Try(W)	-0.9	Neutral
Tyrosine	Tyr(Y)	-1.3	Neutral
Proline	Pro(P)	-1.6	Neutral
Histidine	His(H)	-3.2	Hydrophilic
Glutamine	Gln(Q)	-3.5	Hydrophilic
Asparagine	Asn(N)	-3.5	Hydrophilic
Glutamic Acid	Glu(E)	-3.5	Hydrophilic
Aspartic Acid	Asp(D)	-3.5	Hydrophilic
Lysine	Lys(K)	-3.9	Hydrophilic
Arginine	Arg(R)	-4.0	Hydrophilic

The next step consists of the mathematical processing (filtering) based on the DWT of the numerical/statistical series obtained. On the Farey wavelet approximation of the numerical hydrophobicity scale series, we proceed by the following operations.

- We count the number of amino acids with positive hydrophobicity index for any TM segment S, let N_s be the corresponding number.
- If $N_s < 11$, the segment S is omitted.
- If $38 \leq N_s \leq 72$, we split the segment S into 2 equal parts, and 3 central acids are removed from the transmembrane portion (the first, the middle, and the last residues).
- If $Ns > 72$, the segment S is split into 3 sub-segments with at least 24 residues, and 3 plant acids are omitted.
- We next compute the mean hydropathy H of the whole protein.
- If $H < 0.1$, we act a second selection of TM segments.
- If $H < -0.2$, segments with a maximum hydropathy $m \geq 0.7$ are kept as correct.

- If $-0.2 \leq H \leq 0.1$, segments with a maximum hydropathy $m \geq 0.5$ are kept as correct.

The following figure (Figure 1) illustrates the flowchart of the algorithm above.

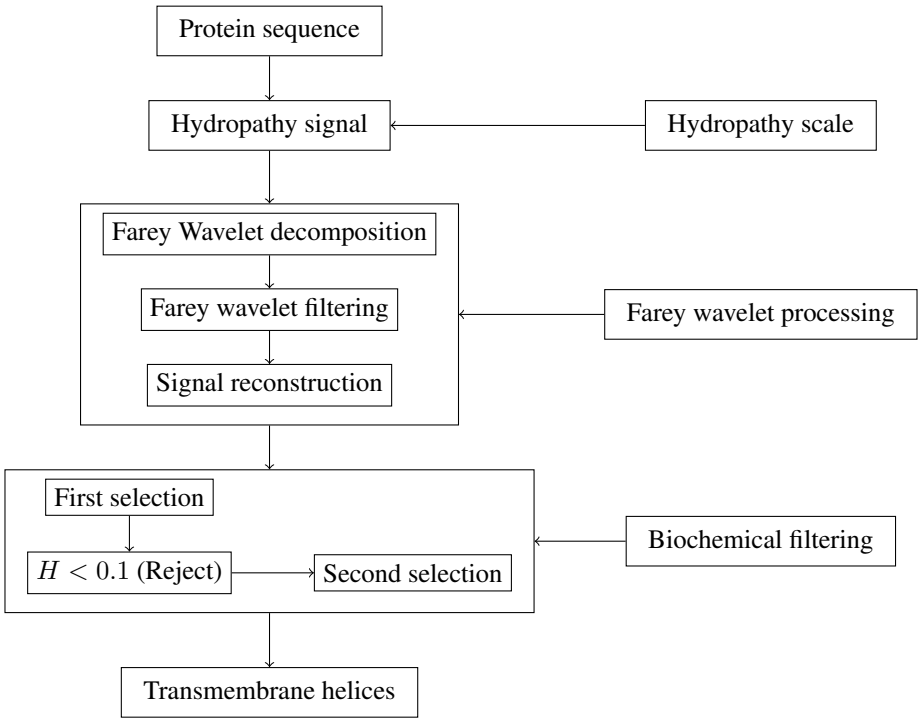

FIGURE 1. General flowchart of the Farey wavelet processing.

In the present paper, we consider as in [20], the strain of coronavirus recorded since 2002-2003 and due to a SARS case in Hanoi, Vietnam, See [26]. This strain is provided in Figure 2. We precisely purpose to conduct a Farey wavelet analysis of such a strain.

MPIPLLPLTLTSGSDLDRCTTPDDVQAPNYTQHTSSMRGVYYPDEIPRSDTLYLTQDLPLPPYSNVTGPHTINHTPGNPVIPPKD
GIYPAATEKSNVVRGWVPGSTMNNKSQSVIIINNSTNVVIRACNPELCDNPPPAVSKPMGTQTHTMIPDNAPNCTPEYISDAPS
LDVSEKSGNPKHLREPVPKNKDGPLYVYKGYQPIDVVRDLPSGPNTLKPIPKLPLGINITNPRAILTAPSPAQDIWGTSAAAYPVG
YLKPTTPMLKYDENGTITDAVDCSQNPLAELKCSVKSPEIDKGIYQTSNPRVVPSGDVVRPPNITNLCPPGEVPNATKPPSVYAW
ERKKISNCVADYSVLYNSTPPSTPKCYGVSATKLNDLCPSNVYADSPVVKGDDVRQIAPGQTGVIADYNYKLPDDPMGCVLAW
NTRNIDATSTGNYNYKYRYLRHGKLRPPERDISNVPPSPDGKPCTPPALNCYWPLNDYGPYTTTGIGYQPYRVVVLSPELLNAPA
TVCGPKLSTDLIKDQCVNPNPNGLTGTGVLTPSSKRPQPPQQPGRDVSDPTDSVRDPKTSEILDISPCSPGGVSVITPGTNASSEV
AVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVPQTQAGCLIGAEHVDTSYECDIPIGAGICASYHTVSLLRSTSQKSIVAYTM
SLGADSSIAYSNNTIAIPTNPSISITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSPCTQLNRALSGIAAEQDRNTREVPA
QVKQMYKTPTLKYPGGPNPSQILPPDPLKPTKRSPIEDLLPNKVTLADAGPMKQYECLGDINARDLICAQKPNGLTVLPPLLTDD
MIAAYTAALVSGTATAGWTPGAGAALQIPPAMQMAYRPNGIGVTQNVLYENQKQIANQPNKAISQIQESLTTTSTALGKLQD
VVNQNAQALNTLVKQLSSNPGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ
SKRVDPCGKGYHLMSPPQAAPHGVVPLHVTYVPSQERNPTTAPAICHEGKAYPPREGVPVPNGTSWPITHQRNPPSPQIITTD
NTPVSGNCDVVIGIINNTVYDPLQPELDSPKEELDKYPKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK
YEQYIKWPWYVWLGPIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKPDEDDSEPVLKGVKLHYT

FIGURE 2. The Coronavirus proteins' series strain.

In the first step, we provide the wavelet decomposition of the numerical time series associated with the strain of proteins due to the Kyte-Dolittle process. This is illustrated by Figure 3.

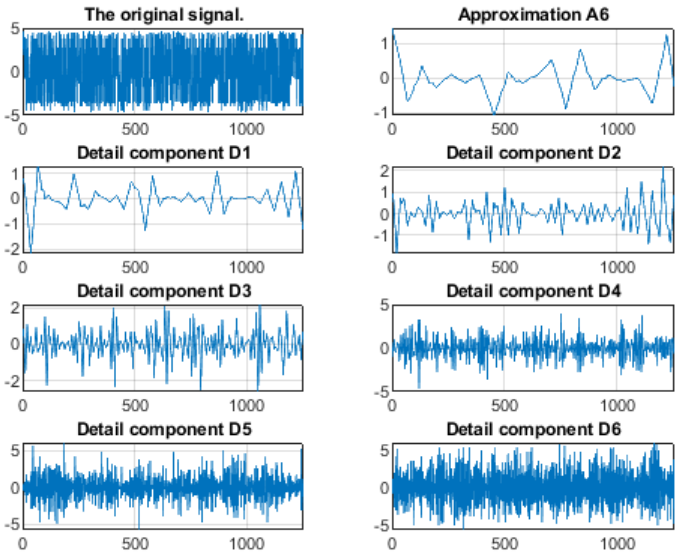

FIGURE 3. The Farey wavelet decomposition of the coronavirus strain at the level $J = 6$.

The next step is to localize the transmembrane segments in the strain above by using the Farey wavelet. To do this we plotted in Figure 4 the Kyte-Doolittle hydropathy signal for the strain. We indicate that the optima with scores greater than 1.8 are due to the possible transmembrane parts. We get here 8 helices (local maxima).

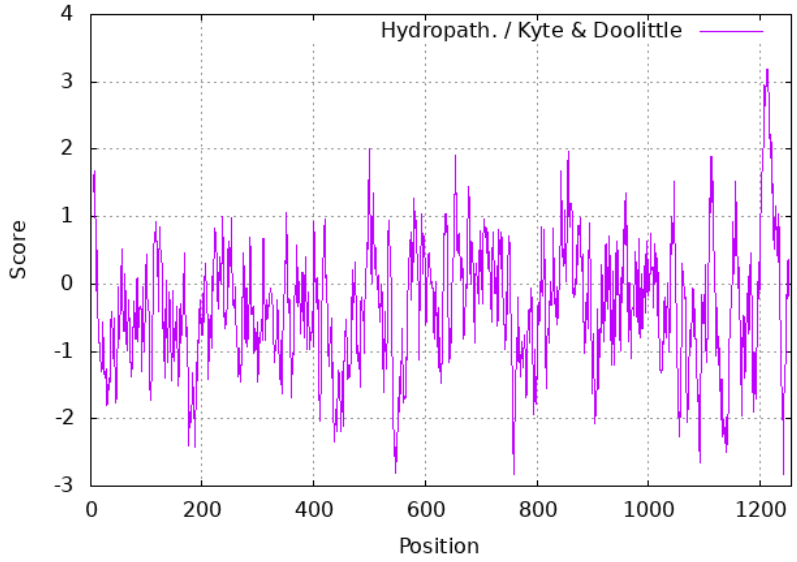

FIGURE 4. Kyte-Doolittle hydropathy signal for the coronavirus series.

Our aim is to show that effectively the Farey wavelet analysis or filtering permits the detection of the same segments due to the transmembrane helices parts. The result is shown in Figure 5.

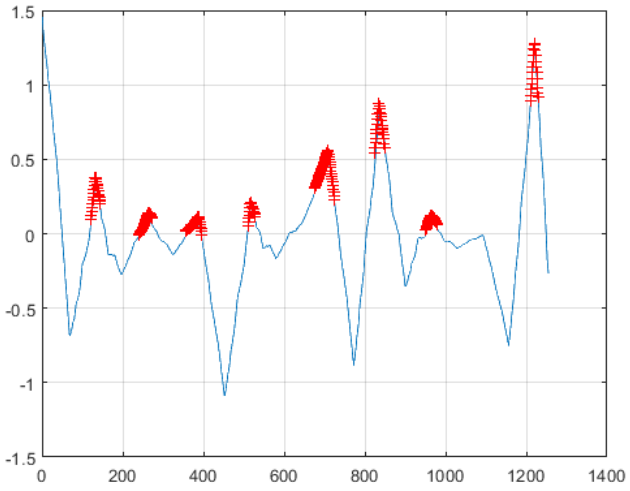

FIGURE 5. The TMHs Segments due to the Farey wavelet filtering of the coronavirus strain.

To illustrate the efficiency of our work, a comparison with the only reference [20] that investigated the similar strain, is provided in Table 2 in terms of the TMHs predicted.

TABLE 2. The TMHs Segments for HSch filtering of the coronavirus signal.

TMHs	Model [20]	Current model
1	120–134	121-145
2	233–253	239-273
3	359–373	356-396
4	505–523	510-529
5	678–699	675-725
6	824–842	824-849
7	1056–1069	950-979
8	1199–1212	1210-1230

To check more statistically, the performance, accuracy, and efficiency of the proposed method, a comparison to the results in [20] is developed via three statistical tests.

- The so-called percentage index
$$Q_p = \frac{N_1}{\sqrt{N_2 N_3}} \times 100\%,$$
where N_1 returns the number of the TMHs obtained via the Farey wavelet reconstruction, N_2 is the number of the TMHs obtained in [20] as a basis of comparison (like the observed case in experimental studies), and N_3 is the total number of predicted TMHs in the original hydropathy signal.
- An absolute deviation is evaluated as the difference between the first residue in [20] and the current first predicted residue. More precisely, we call Mean Absolute Error (MAE) the quantity
$$MAE_1 = \sum_{i=1}^{K} |x_i^1 - x_i^2|,$$
and
$$MAE_2 = \sum_{i=1}^{K} |y_i^1 - y_i^2|,$$
where $[x_i^1, y_i^1]$ are the TMHs segments due to [20], and recalled here in Table 2, column 2. $[x_i^2, y_i^2]$ are the current TMHs segments obtained via the Farey wavelet reconstruction, and subject of column 2 in Table 2. K is the total number of the TMHs segments. The MAE has to be as small as possible.
- The last measures is the so-called Jack-Knife test ([10, 20, 24]). Let N_j be the number of predicted segments due to the Farey wavelet component A_j, $1 \leq j \leq 6$. In the acting Jack-knife test, a single observation is recursively deleted from the sample, and an estimation is computed until we get N_j estimates for N_j, $1 \leq j \leq 6$. Once, the 6 estimates $\widehat{N}_1$, $\widehat{N}_2$, ..., $\widehat{N}_6$, are obtained, we compute a standard error
$$SEJK = \sqrt{\frac{5}{6} \sum_{i=1}^{6} \left(\widehat{N}_i - \overline{\widehat{N}} \right)^2},$$
where $\overline{\widehat{N}}$ is the arithmetic mean of the vector $(\widehat{N}_i)_{1 \leq i \leq J}$.

The readers may refer to [6, 7, 9, 19, 31] for more details on these measures. In the present work, the results obtained are resumed in the following Table 3.

TABLE 3. Statistical measures of performance.

Measures	Values
Q_p	100%
MAE_1	69
MAE_2	131
$SEJK$	97.16%

Notice from Table 3 that the index $Q_p = 100\%$ is compatible effectively with the fact that 8 helices exist in the strain. Besides, the Jack-Knife test is also performant. However, we notice a somehow acceptable MAE_1, which reflects that the first extremities of the TMHs segments are somehow close to each other, and a quite large MAE_2 value, which indicates an important translation of the second extremities of the TMHs segments. This last limitation may be reduced if the real observed segments are provided by the biologists.

4. CONCLUSION

In the present article, a strain of proteins issued from coronavirus is investigated by using a special type of wavelet known as the Farey wavelet. The wavelet processing permitted extraction and/or localize the TMHs segments in the protein strain. The results are compared to existing works and showed the ability of the present wavelet in the extraction of the correct TMHs segments. The performance, accuracy, and efficiency of the method are evaluated by means of some statistical measures such as the Jack-Knife test, the mean absolute error, and the percentage index.

As future work(s) and perspective(s) for the present work, they constitute parts of the continuity of the present work in the fields of biometrics and bioinformatics, it will be interesting to apply other types of wavelets to get more accurate predictions on a set of proteins. Besides, our work as well as a major part of existing works may be improved in a future direction to overcome an essential drawback or limitation related to the aspects of redundancy, self-similarity, fractality, multifractality and may be stochasticity or randomness in the hydropathy signal. We thus think about other mathematical theories of analysis of random signals issued from protein sequences, such as those studying aspects of redundancy and self-similarity in the hydropathy signal. We think that fractal theory, stochastic calculus, and also non-uniform wavelets can offer an alternative to the same extension in this context.

REFERENCES

[1] A. Antoniadis and G. Oppenheim, Wavelets and statistics. Lecture notes in statistics, 103, Berlin/New York, Springer-Verlag, 1995.

[2] S. Arfaoui, I. Rezgui, and A. Ben Mabrouk, Wavelet Analysis on the Sphere Spheroidal Wavelets. Walter de Gruyter (March 20, 2017).

[3] S. Arfaoui, A. Ben Mabrouk, and C. Cattani, Wavelet analysis Basic concepts and applications, CRC Taylor-Francis, Chapman Hall, Boca Raton, 1st Ed., April 21, 2021.

[4] S. Arfaoui, and A. Ben Mabrouk, The Farey map exploited in the construction of a Farey mother wavelet. Annals of Communications in Mathematics, Vol. 6 (2) (2023), 118-132..

[5] A. Z. Averbuch, P. Neittaanmaki, and V. A. Zheludev, Spline and Spline Wavelet Methods with Applications to Signal and Image Processing Volume I (2014). Springer. ISBN 978-94-017-8925-7.

[6] A. Ben Mabrouk, and M. M. Ibrahim Mahmoud, Multifractal study of some biological series, the case of proteins. IWSSIP, 2013.

[7] A. Ben Mabrouk, B. Rabbouch, and F. Saadaoui, A wavelet based methodology for predicting transmembrane segments. Poster Session,the International Conference of Engineering Sciences for Biology, and Medecine, 1-3 May 2015, Monastir, Tunisia.

[8] M. Bernhofer, B. Rost, TMbed: transmembrane proteins predicted through language model embeddings. BMC Bioinformatics 23, 326 (2022).
[9] Y. Bin and Y. Zhang, A Simple Method for Predicting Transmembrane Proteins Based on Wavelet Transform. Int. J. Biol. Sci., 9(1) (2013), 22-33.
[10] J. A. Calvo, K. C. Millett, E. J. Rawdon, and A. Stasiak. Physical and Numerical Models in Knot Theory Including Applications to the Life Sciences. Series on Knots and Everything - Vol. 36, World Scientific, Singapore 2005.
[11] C. Cattani and J. J. Rushchitski, Wavelet and Wave Analysis as applied to Materials with Micro or Nanostructure, World Scientific 2007.
[12] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied mathematics, Philadelphia, PA, USA (1992).
[13] S.H. Feng, W.X. Zhang, J. Yang, Y. Yang, H.B. Shen, Topology Prediction Improvement of α-helical Transmembrane Proteins Through Helix-tail Modeling and Multiscale Deep Learning Fusion. J Mol Biol. 432(4) (2020), 1279-1296.
[14] P. Fischer, G. Baudoux, and J. Wouthers, Wavepred: A wavelet-based algorithm for the prediction of transmembrane proteins. Comm. Math. Sci. 1(1) (2003), 44-56.
[15] W. Hardle, G. Kerkyacharian, D. Picard, A. Tsybakov, Wavelets, approximation and statistical applications. Seminar Berlin-Paris (1997).
[16] T. Hegedus, M. Geisler, G.L. Lukacs, B. Farkas, Ins and outs of AlphaFold2 transmembrane protein structure predictions. Cell Mol Life Sci. 79(1) (2022), ID 73, 12 pages.
[17] M. Holschneider, Wavelets An Analysis Tool, Mathematical Monographs. Clarendon Press. Oxford. 1995.
[18] Z. Liu, Y. Gong, Y. Bao, Y. Guo, H. Wang and G.N. Lin, TMPSS: A Deep Learning-Based Predictor for Secondary Structure and Topology Structure Prediction of Alpha-Helical Transmembrane Proteins. Front. Bioeng. Biotechnol. 8 (2021), ID 629937, 12 pages.
[19] M. M. Ibrahim Mahmoud, A. Ben Mabrouk and M H. Abdallah Hashim, Wavelet multifractal models for transmembrane proteins series. Interna. J. Wavelets Multires and Information Processing, Vol. 14, No. 6 (2016) 1650044 (36 pages).
[20] M. Jallouli, M. Zemni, A. Ben Mabrouk, and M.A. Mahjoub, Toward new multi-wavelets: associated filters and algorithms. Part I: theoretical framework and investigation of biomedical signals, ECG, and coronavirus cases. Soft Comput 25, 14059–14079 (2021).
[21] M. Jallouli, S. Arfaoui, A. Ben Mabrouk and C. Cattani, Clifford wavelet entropy for fetal ECG extraction. Entropy, 2021 (23), 844. https://doi.org/10.3390/e23070844
[22] J. Kyte and R. F. Doolittle, A simple method for displaying the hydrophathic character of a protein, J. Mol. Biol., 157 (1982), 105-132.
[23] S. Mallat, A Wavelet Tour of Signal Processing The Sparse Way, 3rd Edition, 2009.
[24] M. A. Rezaei, P. Abdolmaleki, Z. Karami, E. B. Asadabadi, M. A. Sherafat, H. Abrishami-Moghaddam, M. Fadaie, M. Forouzanfar, Prediction of membrane protein types by means of wavelet analysis and cascaded neural networks. Journal of Theoretical Biology 254 (2008), 817–820.
[25] S. Téletchéa, J. Esque, A. Urbain, C. Etchebest, A.G. de Brevern, Evaluation of Transmembrane Protein Structural Models Using HPMScore. BioMedInformatics. 3(2) (2023), 306-326.
[26] S. Van Der Werf, Nouvelle souche de coronavirus associe au SRAS et ses applications. EP1 694 829 B1, Fascicule de brevet européen, 2010.
[27] J.K. Varga, G.E. Tusnady, TMCrys: predict propensity of success for transmembrane protein crystallization, Bioinformatics, 34(18) (2018), 3126–3130.
[28] D. F. Walnut, An Introduction to Wavelet analysis, Applied and Numerical Harmonic Analysis, Birkhauser, Boston, Basel, Berlin, 2002.
[29] S. Wang, S. Fei, Z. Wang, Y. Li, J. Xu, F. Zhao, X. Gao, PredMP: a web server for de novo prediction and visualization of membrane proteins, Bioinformatics, 35(4) (2019), 691–693.
[30] X. Yin, J. Yang, F. Xiao, et al, MemBrain: An Easy-to-Use Online Webserver for Transmembrane Protein Structure Prediction. Nano-Micro Lett. 10(2) (2018), 8 pages.
[31] M. Zemni, M. Jallouli, A. Ben Mabrouk, and M. A. Mahjoub, Explicit Haar–Schauder multiwavelet filters, and algorithms. Part II: Relative entropy-based estimation for optimal modeling of biomedical signals. International Journal of Wavelets, Multiresolution, and Information Processing, Vol. 17, No. 05, 1950038 (2019).

[1]Department of Mathematics, Faculty of Sciences, University of Tabuk, King Faisal Road, 47512 Tabuk, Saudi Arabia.
[2]Laboratory of Algebra, Number Theory and Nonlinear Analysis LR18ES15, Department of Mathematics, Faculty of Sciences, University of Monastir, Boulevard of the Environment, 5000 Monastir, Tunisia.
[3]Department of Mathematics, Higher Institute of Applied Mathematics and Computer Science, University of Kairouan, Street of Assad Ibn Al Fourat, Kairouan 3100, Tunisia.

Email address: 391000151@stu.ut.edu.sa; 421009239@stu.ut.edu.sa;
Email address: 411008509@stu.ut.edu.sa; 391000093@stu.ut.edu.sa;
Email address: aalanazi@ut.edu.sa; anouar.benmabrouk@fsm.rnu.tn