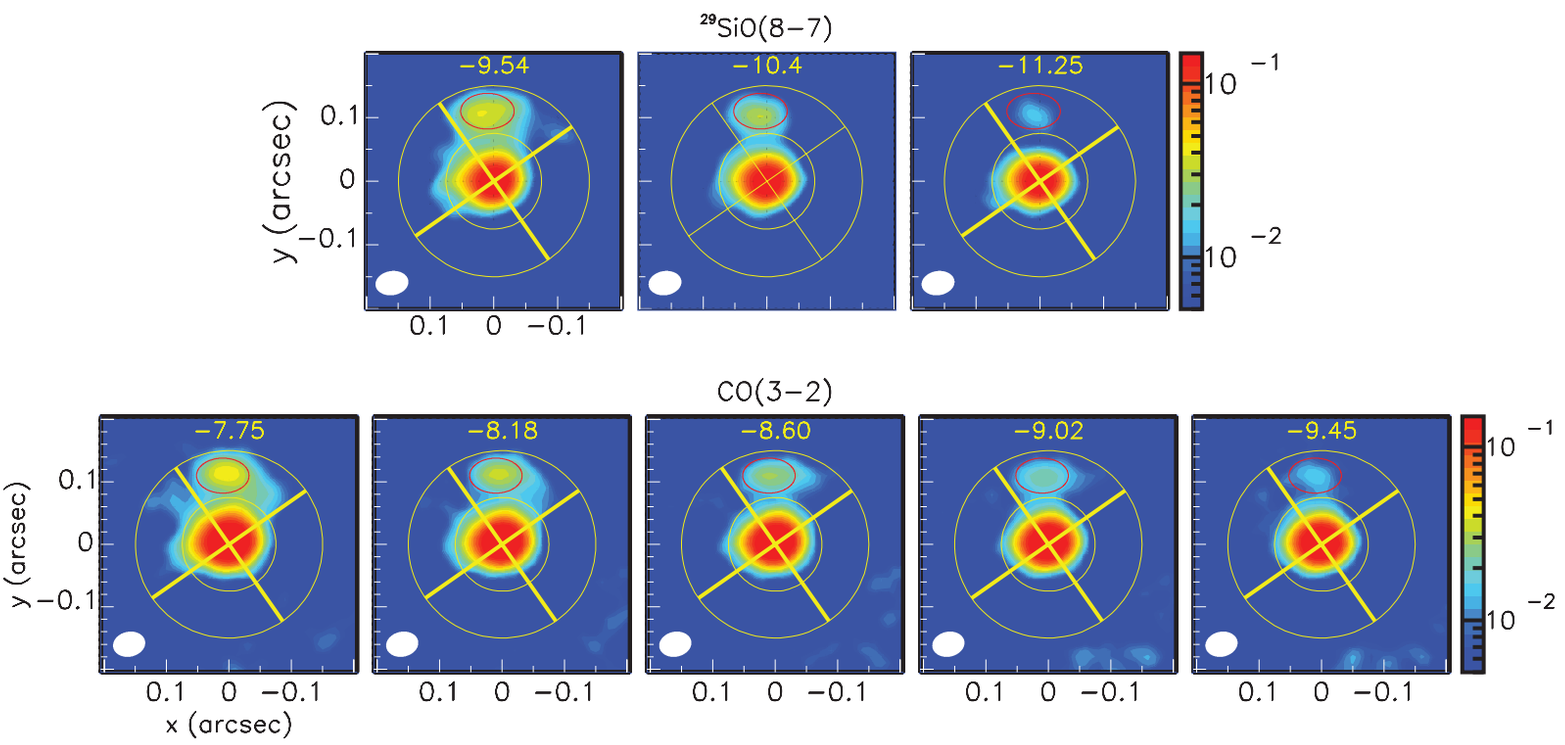
P-ISSN 2525-2461 • E-ISSN 2615-9937

Vietnam Journal of Science, Technology and Engineering

DECEMBER 2022 • VOLUME 64 NUMBER 4



Recent mass ejection from AGB star W Hya

VIETNAM JOURNAL OF SCIENCE, TECHNOLOGY AND ENGINEERING * DECEMBER 2022 * VOLUME 64 NUMBER 4
vietnamscience.vn or vietnamscience.vjst.vn

Vietnam Journal of Science, Technology and Engineering (VJSTE) **is a quarterly double-blind peer-reviewed academic journal. VJSTE publishes research articles and review articles in all areas of science, technology and engineering with the focus on the physical sciences, life sciences and environmental sciences including the following topics:**

- **Physical Sciences**
- **Agriculture**
- **Medicine**
- **Pharmacology**
- **Biology**
- **Biotechnology**
- **Biomedical Applications**
- **Ecology**
- **Climatology**
- **Environmental Science.**

CLASSIFICATION SCHEME

1. **Mathematics and Computer Science**
   - **1.1. Mathematics**
   - **1.2. Computer Science**
   - **1.3. Computational Science**
2. **Physical Sciences**
   - **2.1. Physics**
   - **2.2. Chemistry**
   - **2.3. Engineering**
3. **Life Sciences**
   - **3.1. Agriculture**
   - **3.2. Medicine**
   - **3.3. Pharmacology**
   - **3.4. Biology**
   - **3.5. Biotechnology**
   - **3.6. Biomedical Applications**
4. **Earth Sciences**
   - **4.1. Geography**
   - **4.2. Geology**
   - **4.3. Geophysics**
   - **4.4. Oceanography**
5. **Environmental Sciences**
   - **5.1. Ecology**
   - **5.2. Climatology**
   - **5.3. Environmental Science**

CONTENTS

DECEMBER 2022 • VOLUME 64 NUMBER 4

37

Baxter

Yumi

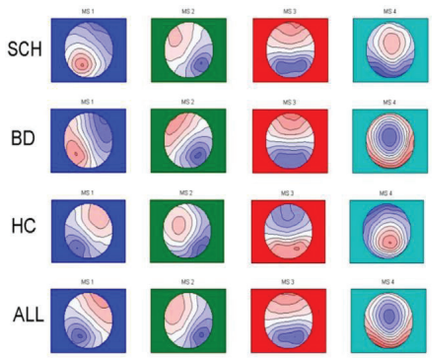
LBR iiwa

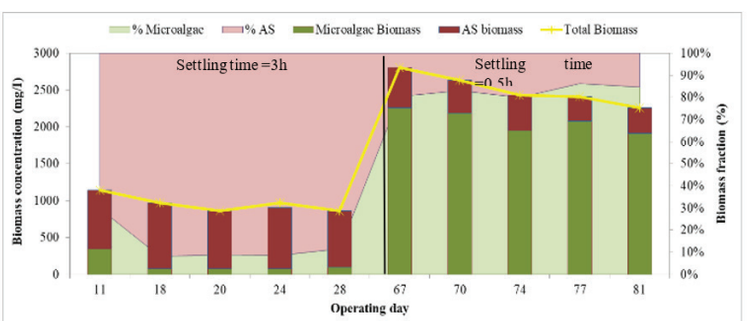
CR-35iA

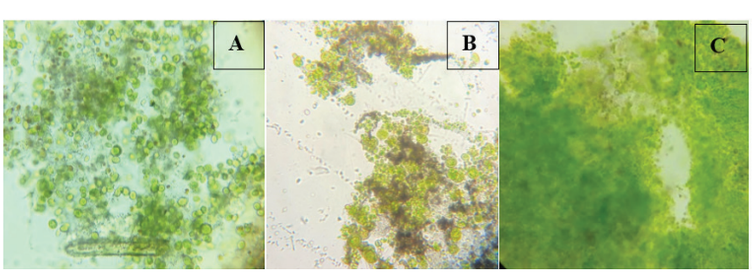
Universal Robot

CONTENTS

DECEMBER 2022 • VOLUME 64 NUMBER 4







DOI: 10.31276/VJSTE.64(4).03-09

Implementation of Boneh - Lynn - Shacham short digital signature scheme using Weil bilinear pairing based on supersingular elliptic curves

Nhu-Quynh Luc*, Quang-Trung Do, Manh-Hung Le

Academy of Cryptography Techniques, 141 Chien Thang Street, Tan Trieu Ward, Thanh Tri District, Hanoi, Vietnam

Received 4 May 2022; revised 28 June; accepted 14 July 2022

Abstract:

One option for a digital signature solution for devices with low memory and low bandwidth transmission over channels uses a short digital signature scheme based on Weil bilinear pairing aimed at short processing times, fast computation, and convenient deployment on applications. The computational technique of non-degenerate bilinear pairings uses supersingular elliptic curves over a finite field F_{p^l} (where p is a sufficiently large prime number) and has the advantage of being able to avoid Weil-descent, Menezes-Okamoto-Vanstone (MOV) attacks, and attacks by the Number Field Sieve algorithm. Compared to Elliptic Curve Digital Signature Algorithm (ECDSA) digital signature schemes, generating a digital signature for a Boneh-Lynn-Shacham (BLS) scheme using Weil bilinear pairing on a supersingular elliptic curve is simple. In this study, the authors replace non-degenerate bilinear pairing calculations on a supersingular elliptic curve with a Weil pairing with $P \epsilon E(F_p)$, $Q \epsilon E(F_p 1)$ and a higher security multiplier $\alpha=12$ in the BLS short digital signature scheme. The execution time of the BLS short digital signature program showed improvement compared to the commercial ECDSA digital signature scheme.

Keywords: **digital signature, ECDSA (Elliptic Curve Digital Signature Algorithm), elliptic curve cryptography, tate pairing, Weil pairing.**

Classification number: **1.2**

1. Introduction

Information exchange between devices and applications requires security and authentication with high reliability per the demanding strict standards of this digital era. New requirements for digital signature solutions such as short digital signatures, fast processing speeds, message authentication without transmissions, and digital signature on short message and low bandwidth channel transmissions are essential for today's applications [1-5]. To date, short digital signature solutions and signature authentication using the calculation of an elliptic curve, such as ECDSA, Elliptic Curve-based Schnorr Digital Signature Algorithm (ECSDSA), or Edwards-Curve Digital Signature Algorithm (EdDSA) have been applied widely in commercial products [1, 2, 6-9]. Among these, the digital signature solution with a short digital signature using the calculation of Weil and Tate bilinear pairing of the authors Boneh, Lynn, Schacham (2001) (denoted by the BLS short digital signature scheme) proves to meet the requirements [2, 10].

The BLS scheme uses a special supersingular curve with p=3, which raises the security level of the BLS scheme to be equivalent to the Digital Signature Algorithm (DSA) using a 1024-bit prime number [11-13]. The BLS short digital signature scheme is secure against attack with selected messages (according to a random oracle model), given that "Computational Diffie-Hellman based on an elliptic curve over finite field F_{p^l} (where p is a sufficiently large prime number) being difficult to solve" [1, 2]. The advantage of the BLS scheme when generating a digital signature is its simplicity as both the digital signature and signature verification processes use a non-degenerate bilinear pairing (Weil and Tate bilinear pairings) on the elliptic curve [2, 6, 10, 14-18]. Since this non-degenerate bilinear pairing calculus technique uses a supersingular elliptic curve over finite field F_p, such that both generic discrete log algorithm in $E(F_p)$ and the Number Field Sieve in F_{p^l}* are intractable, it is resistant to some Weil descent and MOV attacks [11, 12], as well as attacks by the Number Field Sieve algorithm [19-21]. Several publications have shown that elliptic curve cryptography (ECC) built on non-degenerate bilinear pairing could be a secure cryptosystem for today's applications with one particular development being the supersingular isogeny Diffie-Hellman (SIDH) [7, 22, 23].

This solution aims towards short processing time, fast computation, and convenient deployment on applications, making it fit for devices with low memory and transmission over low bandwidth channels. The authors have used computational techniques of Weil non-degenerate bilinear pairing (with a higher security multiplier α=12) in building a BLS short digital signature scheme based on a supersingular elliptic curve with functions for key generation, digital signature, and signature verification.

*Corresponding author: Email: quynhln@actvn.edu.vn

2. Related works on the BLS short digital signatures scheme

2.1. Mathematical basis of Weil and Tate pairing based on Supersingular Elliptic curves

Torsion points play an important role in the calculations of Weil and Tate bilinear pairings on elliptic curves and usually torsion points are points of finite order [1, 7].

Definition 1: Given an elliptic curve E over a field K and a positive integer n. Then, the set of n-torsion points is defined as the set $E[n] = \{P \in E(\overline{K}) | nP = \infty\}$ [1].

Since the characteristic of K is not divisible by n, the equation $x^n=1$ does not have multiple solutions, but has n solutions in $\overline{K}$ and μ_n is a cyclic group of order n. An element $\zeta \in \mu_n$ satisfies $\zeta^k=1$ if and only if n is divisible by K, then ζ is called a primitive root of degree n [1].

Definition 2: Let there be an elliptic curve E over K and n be an integer not divisible by the characteristic of K such that $E[n] \subseteq E[K]$. Then, the Weil pairing is the mapping e_n:E[n]×E[n]→μ_n [2].

Given $T \in E$[n], there exists a function f such that $div(f)=n[T]-n[\infty]$. Then choose $T' \in E[n^2]$ with $nT'=T$, there exists g such that $div(g)=\sum_{R \in E[n]}([T'+R]-[R])$. For $S \in E[n], P \in E[\overline{K}]$, then $g(P+S)^n=f[n(P+S)]=f(nP)=g(P)^n$. Thus $\frac{g(P+S)}{g(P)} \in \mu_n$ and $\frac{g(P+S)}{g(P)}$ do not depend on P. Hence, the Weil pairing is $e_n(S,T) = \frac{g(P+S)}{g(P)}$.

Definition 3 [2]: Let p be a prime power, and E/F_p an elliptic curve with m points in $E(F_p)$. Let P in E/F_p be a point of primer order q where $q^2 \nmid m$. We say that the subgroup ⟨P⟩ has a security multiplier α, for some integer α>0, if the order of p in F_q^* is α. In other words:

$q | p^\alpha - 1$ and $q \nmid p^k - 1$ for all $k = 1, 2, \ldots, \alpha - 1$.

The security multiplier of $E(F_p)$ is the security multiplier of the largest prime order subgroup in $E(F_p)$.

Theorem 1 [2, 7, 17, 24]: Let E be an elliptic curve defined over a field F_p. Let n be an integer so that $n|(q-1)$. The elements of $E(F_p)$ of n are denoted by $E(F_p)[n]$ in dividing order, and let $\mu_n=\{x \in F_p | x^n=1\}$. Assume $E(F_p)$ contains an element of order n. Then, there exists a non-degenerate bilinear mapping:

$$\begin{cases} \langle .,. \rangle_n : E(F_p)[n] \times E(F_p)/nE(F_p) \to F_p^\times/(F_p^\times)^n \\ \tau_n : E(F_p)[n] \times E(F_p)/nE(F_p) \to \mu_n \end{cases}$$

The first pairing is called Tate-Lichtenbaum pairing. The second one, τ_n, is called the modified Tate-Lichtenbaum pairing [2, 7, 17, 24]. Each element in $E(F_p)/nE(F_p)$ has the form $Q+nE(F_p)$, so it is usually written as $\langle P,Q \rangle_n$ and $\tau_n(P,Q)$ instead of $\langle P,Q+nE(F_p) \rangle_n$ and $\tau_n(P,Q+nE(F_p))$. Since $F_p^\times$ is a cyclic group of order n, the $\frac{p-1}{n}$ powers of $\langle P,Q \rangle_n$ and $\tau_n(P,Q)$ give an isomorphism $F_p^\times/(F_p^\times)^n \to \mu_n$. Hence

$$\tau_n(P,Q) = \langle P,Q \rangle_n^{\frac{p-1}{n}} \quad (1)$$

2.2. Compute the Tate pairing according to Miller's algorithm [3, 7, 17, 24]:

Given an elliptic curve E over F_p; P, Q are points with prime order n and $P,Q \in E(F_p)$. Draw the line n_1 through P and Q, which intersects E at another point called R_1. Draw the vertical line n_2, which is the line connecting R_1 and the point ∞. The line n_2 intersects E at the third point, which is R_2 ($R_2=P+Q$). The lines n_1 and n_2 are functions on E and have a main divisor [2]:

$$\begin{cases} div(n_1) = [P] + [Q] + [R_1] - 3[\infty] \\ div(n_2) = [R_1] + [R_2] - 2[\infty] \end{cases}$$

Divisor [Q']-[S] will be equivalent to D_Q=[Q]-[∞], so S is chosen at random. Calculate gD_P at D_Q, where at each step in the algorithm T_1 is the point obtained by computing mP where m is an integer represented in binary of the binary expansion of n. Calculate f_1 to be the value at [Q']-[S] of the function f satisfying m([P]-[∞])=[T_l]-[∞]+$div(f)$. At the end of the algorithm the value reaches $T_l=\infty, f_l=gD_P$. It follows that f_1 is the value at [Q']-[S] of the function gD_P satisfying m([P]-[∞])=$div(gD_P)$ as required by the definition of the Tate pairing. For $P \in E(F_p), Q \in E(F_{p^l})$ the Tate pairing is calculated according to the formula $\langle P,Q \rangle_n$ and the modified Tate-Lichtenbaum pairing is calculated by formula (1) with powers (p^l-1)/n [1, 3, 7].

Algorithm 1: Miller's algorithm for computation with Tate bilinear pairings [2, 7]

Input: Let the elliptic curve E over the field F_p. Two points P and Q on E are points of order n.

Output: The value f_1 satisfies the definition of a Tate pairing (Theorem 2).

1. Randomly select $S \in E(F_{p^l})$ and calculate $Q'=Q+S \in E(F_{p^l})$.
2. Let $l=[log_2(n)]-1, T_l=P, f_l=1$
3. While $l \geq 1$ do
   - Write equations for the lines n_1 and n_2 with the multiplication of T_1.
   Calculate $T_l=2T_l, f_l=f_l^2((n_1(Q')n_2(S))/(n_2(Q')n_1(S))$
   - If the l^{th} bit of n is 1, then
   write equations for the lines n_1 and n_2 with the addition of points of T_1 and P.
   Calculate $T_l=T_l+P, f_l=f_l^2((n_1(Q')n_2(S))/(n_2(Q')n_1(S))$
   - Decrease l.
4. Return f_1

The input is an elliptic curve E chosen as a supersingular curve E over the field F_p, $p>3$ (the curve E over the field F_p is said to be supersingular if the curve E satisfies $E[P]=[\infty]$); The subgroup $E(F_p)[n]$ has an influence on the computation in Miller's algorithm, so the number of iterations is $[log_2(n)]$ [2, 7]. For Tate pairing, it is necessary to pay attention to the field characteristic of 2,3 and make sure the order of the group $E(F_p)$ is appropriate, so choose the prime number n as the largest prime divisor of the group order $E(F_p)$. In Miller's algorithm, integer n is calculated by Schoof's algorithm and using the point multiplication algorithm kP [1, 4, 16, 25-27].

According to Algorithm 1, calculating the Tate pairing $\langle P,Q \rangle_n$, (with $P \in E(F_p)$, $Q \in E(F_{p^l})$) on security applications, the line coefficients n_i belongs to the subfield of F_p, the finite field is used to calculate the value of f_1 with a large length field. At that time, the attacker who wants to attack the Miller algorithm must solve the problem "The point P to be found belongs to $E(F_p)$ when knowing the public point Q belongs to $E(F_{p^l})$, then finding the point P is more complicated" [2, 23]. Formula $e_n(P,Q) = \frac{\langle P,Q \rangle_n}{\langle Q,P \rangle_n}$ is often used to calculate Weil

pairing [3, 7]. In addition, the Weil pairing is also calculated according to the formula $e_n(P,Q) = \frac{f_P(R)f_Q(P)}{f_P(Q+R)f_Q(\infty)}$ but it is not favourable [1, 3, 7]. So, the Weil pairing is considered as another way of calculating the Tate pairing when the conditions for the Weil pairing occur.

When $P \in E(F_p)$, $Q \in E(F_{p^l})$, both Tate and Weil pairing calculations are time consuming. Therefore, the calculation time for the required Weil pairing takes twice as much as the calculation of the Tate pairing. In this study, the authors have replaced the non-degenerate bilinear pairing calculations on the supersingular elliptic curve with the Weil pairing in the BLS short digital signature scheme. Then, the performance of the BLS short digital signature scheme is evaluated by comparison with the classic ECDSA scheme commonly used today.

3. Building a BLS short digital signature scheme based on the non-degenerate bilinear pairing of supersingular elliptic curves

3.1. The BLS key generation scheme

With the BLS short digital signature scheme, the curve E used is $y^2=x^3+Ax+B \bmod p$. The input for key generation consists of a set of parameters *(A, B, p, q, l, P)* denoted *BTS-BLS* (Table 1) [2]. This parameter set is used by the author for all key generation, digital signatures, and signature verification processes of the BLS short digital signing scheme.

Table 1. Parameter sets used in the BLS short digital signature scheme.

Parameters	Functions
A, B	The coefficients of the supersingular elliptic curve equation
p	Modulo
q	Greatest prime divisor of $\#(E/F_{p^l})$
l	Key length belongs to F_{p^l}
Point $P \in E/F_{3^l}$	Base point with order *q*

In Algorithm 2, the generated key pair consists of the public key *PK* and the private key *SK* in which the public key is the parameter set *PK=(l, q, P, R)* and the private key $SK=x$, with x is a random number belonging to Z_p^* (with a large enough prime p). When generating the key for the BLS short digital signatures scheme, the BLS scheme only uses the *kP* point multiplication algorithm and choses a random number belonging to Z_p^*. This shows that the key generation process for the BLS short digital signatures scheme is efficient and simple.

Algorithm 2: Generate keys for the short digital signature scheme BLS [2, 6]

- Input: Let l, the curve (E/F_{p^l}) and q is the greatest prime divisor of $\#(E/F_{p^l})$, the point P has order q
- Processing steps: Chosen random number $x \in Z_p^*$ and alculate $R \leftarrow xP$
- Output: The public key $PK=(l, q, P, R)$ and the private key $SK=x$

3.2. The BLS short digital signature scheme

According to Algorithm 3, the signing process of the BLS short digital signatures scheme also uses the input parameters of the supersingular elliptic curve E on the field F_{p^l}; the parameters of the curve used for digital signature are the number of the corresponding *BTS-BLS* tuple in the key generation scheme for the BLS scheme.

Algorithm 3: The BLS short digital signature [2, 6, 7]

- Input: message $M \in \{0,1\}^*$, private key $SK=x$
- Parameter set: *BTS-BLS*
- Processing steps:

+ Using $MaptoGroup_{h'}$ algorithm [2], map message M to point $P_M=(x_M,y_M) \in \langle P \rangle$ belonging to E/F_{p^l}

+ Calculate $S_M=xP_M$

- Output: signature $\sigma=x_{S_M} \in F_{p^l}$ of the point $S_M=(x_{S_M},y_{S_M})$

In this algorithm, embedding the message M to be signed into a point $P_M=(x_M,y_M) \in E/F_{p^l}$ and using the kP multiplier algorithm to create a signature for the message M is necessary. The message M, before embedding into a point $P_M \in E/F_{p^l}$ will be hashed using a hash function [5]. The mapping of this hash value to a component x_M coordinate of point P_M is accomplished using the $MapToGroup_{h'}$ algorithm [2, 6, 7]. Thus, the process of creating a digital signature of the BLS short digital signature scheme is more complicated than that of the key generation algorithm of the ECDSA scheme [16, 28, 29]. In the BLS short digital signature scheme, the signature generation process requires the use of a cryptographic hash function and the technique of embedding the message into a point of the curve. This keeps the value of the digital signature generated by the BLS short digital signature scheme small.

3.3. The BLS signature verification scheme

In Algorithm 4, signature verification of the BLS scheme is done using the same set of input parameters of the curve as above Table 1. To verify the digital signature, first one must check whether the obtained signature belongs to the curve. Secondly, two values of Weil pairings will be computed, as the first one is being calculated from the base point and the digital signature, and the second one from the public key and the message M. If these two values are equal or the inverse of the first value is equal to the second value, then the signature is valid.

Algorithm 4: The BLS signature verification [2, 6, 7]

- Parameter set: *BTS-BLS*
- Input: The public key $PK=(l, q, P, R)$, the message $M \in \{0,1\}^*$, and the signature σ
- Output: The signature σ is valid or invalid
- Processing steps:

Step 1: Check the condition that the signature σ is the coordinates x_{S_M} of the point $S_M=(x_{S_M},y_{S_M}) \in E/F_{p^l}$. If such a point does not exist, the signature is invalid.

Step 2: Calculate $u \leftarrow e[P,\phi(S)]; v \leftarrow e[R,\phi(h(M))]$, where e is a non-degenerate bilinear mapping (Weil pairing) on the curve $E/F_{p^{6l}}$ and $\phi: E \rightarrow E$ is a Frobenius endomorphism.

Step 3 (check condition u, v): If $u=v$ or $u^{-1}=v$, then the signature is valid, otherwise the signature is invalid.

The correctness of the BLS short digital signature verification algorithm (algorithm 4) is confirmed in step 3 of the algorithm, whether the signature is valid or not. Specifically, with (σ, y) and $(\sigma, -y)$ being two points on E/F_{p^l}, where σ is the x coordinate, one of the two

points can be point S_M or can be used to generate digital signatures in the BLS short digital signatures scheme. From $(\sigma,y)=-(\sigma,-y)$ on the curve, then $e(P,\phi(-S))=e(P,\phi(-S))^{-1}$. Therefore, the $u=v$ condition is to check that $(P, R, h(M), S)$ is a Diffie-Hellman tuple, while the $u^{-1}=v$ condition is to check that $(P, R, h(M), -S)$ is a Diffie-Hellman set [6, 7].

3.4. Theoretical model to prove the security of the BLS short digital signature scheme

In Ref. [2], a secure proof theory for the BLS short digital signatures scheme was propose. The theoretical model that proves the security of BLS is based on the difficulty level of the Hidden Field Equation (HFE), co-CDH (Computational co-Diffie-Hellman), co-DDH (Decision co-Diffie-Hellman), and GDH (Gap Diffie-Hellman groups) problems. It is shown that when an isomorphism $\psi:G_2\rightarrow G1$ exists, the short digital signatures scheme BLS is vulnerable to the discrete log problem by MOV attacks [11, 12], and attacks by the Number Field Sieve algorithm [19-21] on the extended field F_{p^l}.

For Co-GDH signatures from elliptic curves [2], the security level of the BLS short digital signatures scheme is equivalent to the difficulty of the co-CDH (Computational co-Diffie-Hellman) problem on (G_1,G_2). In other words, it is the computational requirements of a discrete log in G_1 or the computation of a discrete log in $F^*_{p^l}$. According to [2], when the BLS scheme uses a special supersingular curve with p=3, the security level of the BLS scheme is equivalent to DSA using a 1024-bit prime (MOV attack [11-13]. This is a weakness of the BLS short digital signatures scheme when the number p is small. To use the BLS schema in this case, we would have to use a curve $E(F_{3^l})$ where 3^{6l} is much larger than 1024 bits.

In the case of a BLS schema using a non-supersingular curves over fields of high characteristic with the security multiplier α=6, [2] shows that with l=159-bit (Signature size $[log_2q]$ of the BLS scheme) is equivalent to "DLog Security $[log_2p]$ of 158 bits" and "MOV Security $[6log_2q]$ of 954 bits". Signatures using this curve are 168 bits while the best algorithm for co-CDH on $E(F_p)$ requires either (Formula (1) in [2]) a generic discrete log algorithm taking time approximately 2^{83}, or (Formula (2) in [2]) a discrete log in a 1008-bit finite field of large characteristic.

Finally, consider the BLS schema in the case of higher security multipliers (Definition 3). D. Pointcheval, J. Stern (2000) [30] proposed certain Abelian varieties. However, to obtain security comparable to DSA using a 2048-bit prime with α=6, we get signatures size l=342 bits. Then, with α=12, the signature is shorter but the security level is guaranteed (equivalent to 2048-bit discrete-log security) [31]. The result is an n-bit signature where the pairing reduces the discrete log problem to a finite field of size approximately $2^{7.5n}$.

4. Results and discussion

4.1. Architectural design of BLS short digital signature scheme

Figure 1 details the implementation steps of the key generation algorithm of the BLS schema, the diagram shows that the key generation modulo is simply designed using only a random function and multiplication points (kP) on the elliptic curve. The key generation modulo then will generate the private key and the public key, which

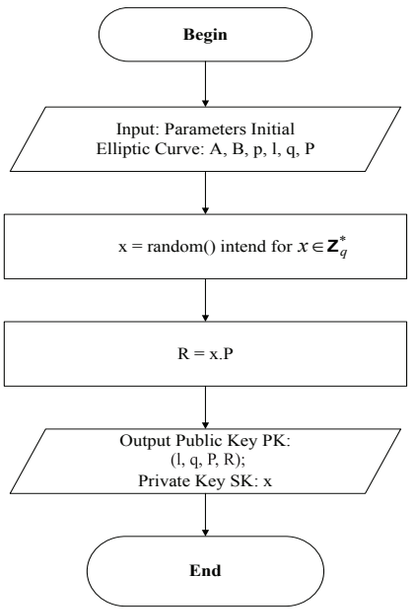


Fig. 1. Scheme of the BLS key generation.

are saved as "bls_private.key" and "bls_public.key," respectively. After executing the key pair generation, the program modulo will issue a notice about the key pair generation time.

Figure 2 shows details the steps of implementing the DSA of BLS schema with a digital signature called "bls_signature.sig". First, when performing a digital signature according to the BLS scheme, the message to be signed, M, will be passed through a secure hashing algorithm that outputs a summary (hash value) [5]. This summary is combined with the private key (the key generated by the BLS key generator modulo), which is then fed into the digital signature program modulo, which results in the digital signature bls_signature. The digital signature program can sign data files of any content with text

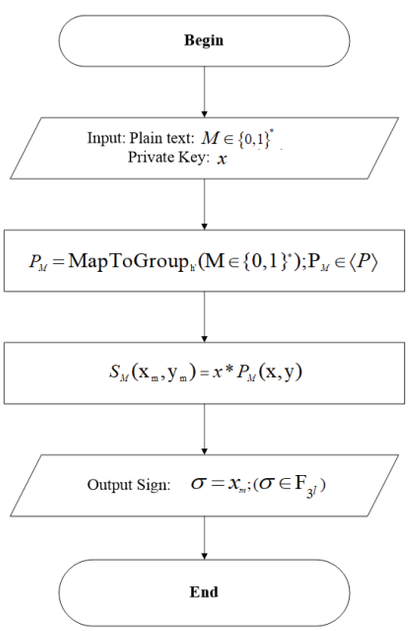


Fig. 2. BLS digital signature scheme.

file formats, image files, audio files, video files, etc. When performing digital signature, the program will create a digital signature file (bsl_signature.sig) and output the execution time of the digital signature process.

Figure 3 details the implementation of the signature verification algorithm steps of the BLS short digital signature scheme. The program verifies the content of the signed data file and calculates the signature verification time of the BLS short digital signature scheme. The received message is passed through the hashing algorithm that obtains the hash value. The process of checking the digital signature of the BLS scheme is done by calculating and checking the input parameters of the hash digest, digital signature, and public key. If the conditions are satisfied, then the signature is valid.

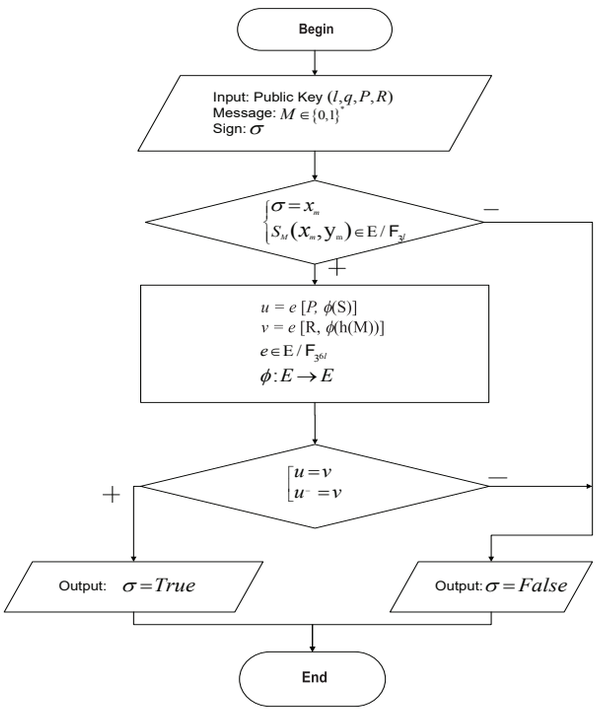


Fig. 3. The BLS signature verification.

4.2. Results of the short digital signature program BLS

In this study, the authors have built a program with 3 main modules: key generation, digital signature and signature verification according to BLS scheme. First, the key generation modulo generates a public key and a private key, then the digital signature modulo performs digital signature with the newly generated private key in the key generation modulo. Finally, the signature verification modulo will perform the signature verification with the public key. In addition, in order to facilitate the performance evaluation of the BLS short digital signature scheme, the authors also built a program following the ECDSA digital signature scheme including the key generation module, digital signature module, and signature verification module [16, 28, 29]. Comparisons of key generation, digital signature, and signature verification program against the BLS and ECDSA scheme were executed on the computer using Intel(R) Core i5-4200U, CPU @ 1.60GHz, up to 2.30 GHz; RAM: 4.00 GB.

Based on the security analysis and evaluation for such a BLS scheme, in this study the authors have selected the parameters for the supersingular elliptic curve over finite field F_p such that both a generic discrete log algorithm in $E(F_p)$ and the Number Field Sieve in $F_{p^l}^*$ are intractable, with p=7DDCA613A2E3DDB17 49D0195BB9F14CF44626303, the security multiplier α=12, and signature size l=159. The coefficients of the supersingular elliptic curve are A=-3, B=21C3F3AC7864D1F99273D0F828D3657D8CFD4E ($y^2=x^3+Ax+B$). This parameter set was evaluated by the National Institute of Standards and Technology (NIST, US Department of Commerce), which minimised the risk of being attacked [2, 6, 7, 28].

Table 2 details the execution time of the BLS key generation, digital signature, and signature verification computations. To check the correctness of the program, the authors tested the program with 2 scenarios, specifically:

Table 2. Results of digital signature and signature verification according to the BLS scheme.

Input data	Digital signature time (ms)	Signature verification time (ms)
535 KB	31	98
1.56 MB	119	161
9.47 MB	577	646
9.79 MB	618	671
25.5 MB	1643	1638

Scenario 1: The authors modified the contents of the input data files of the BLS short digital signature program, kept the key and signature, then checked the authenticity of the data. Fig 4 details the process of modifying the input data, where the results showed that the digital signature is invalid and the processing time was given (Fig. 5).

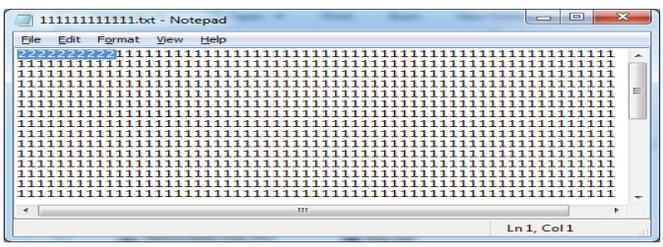


Fig. 4. Modification of the contents of the signed data file.

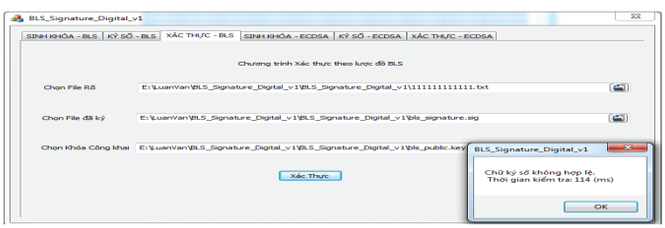


Fig. 5. Signature verification after the message was modified.

Scenario 2: The program generated an original signature (Fig. 6). Then, the author modified the signature (Fig. 7) but did not change the message and the public key. The data verification process for the modified signature resulted in an invalid signature (Fig. 8). Moreover, to evaluate the BLS short digital signature program performance, the

authors tested the digital signature and signature verification program according to the BLS short digital signature scheme with several data files of different lengths (Tables 2, 3).

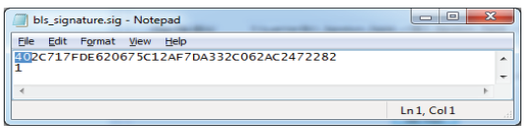


Fig. 6. Original unmodified signature.

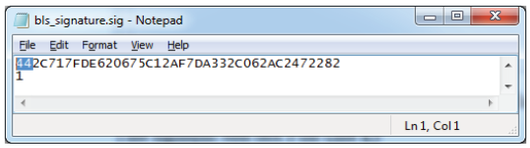


Fig. 7. Signature after modification.

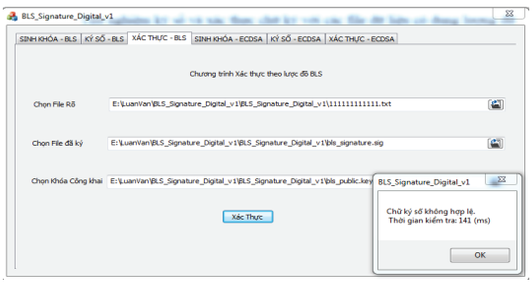


Fig. 8. Signature verification after the signature was modified.

Table 3. Runtime comparison of BLS scheme and ECDSA scheme.

Input data (mb)	Digital signature time (ms)			Signature verification time (ms)		
	BLS	*ECDSA*	*Diff. in %*	*BLS*	*ECDSA*	*Diff. in %*
1.02	108	350	69.14%	166	347	52.16%
1.56	131	523	74.95%	201	523	61.57%
2.00	186	720	74.17%	241	713	66.20%
3.68	298	1227	75.71%	335	1230	72.76%
4.07	313	1353	76.87%	350	1337	73.82%
5.03	376	1637	77.03%	418	1664	74.88%
6.01	450	1928	76.66%	473	1955	75.81%

4.3. Analysis and evaluation of the results achieved by the short digital signature program BLS

In previous publications, the authors evaluated the execution speed and occupied resources of the Tate pairing computation and kP point multiplication algorithm on a Spartan6 XC6SLX150T FPGA hardware platform [25, 32].

In this study, the authors tested the execution time of the program under two scenarios. The first was to evaluate the execution speed between the two program functions, i.e., digital signature and signature verification. Second, the authors evaluated the execution speed between the BLS short digital signature program and the ECDSA digital signature program. For each function of the program, the authors ran the test three times and took the average execution time.

Execution speed of digital signature and signature verification BLS: Table 2 details the execution time results of the digital signature modulo and BLS signature validation. Fig. 9 shows the corresponding graph comparing the running time between digital signature and signature verification. Experimental results of the BLS scheme show that the signing time is faster than the validation time. Theoretically, the digital signature of the BLS scheme uses one-point multiplication, while the validation uses two values of the Weil pairing for calculation. In the Weil non-degenerate bilinear pairing values calculation, a point multiplication is used for each value of u and v. Therefore, calculating u, v requires two-point multiplications, which makes the signature verification time longer than the digital signature time.

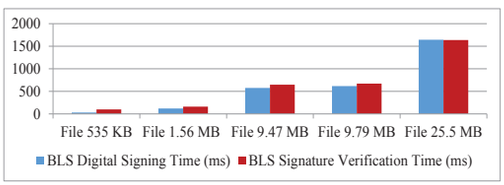


Fig. 9. Digital signature time and signature verification time of the BLS scheme.

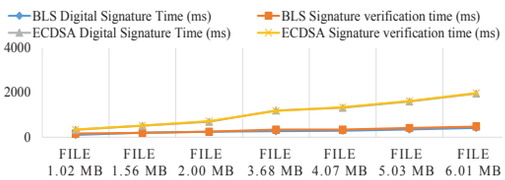


Fig. 10. Runtime comparison of BLS short digital signature and ECDSA schemes.

Execution speed of BLS short digital signature program and ECDSA digital signature program: Both the BLS and ECDSA digital signature schemes are designed with a 160-bit key-length key for the same data input. Table 3 and the diagram in Fig. 10 present the run-time details of the digital signature function for both the BLS and ECDSA short digital signature scheme.

Table 3 shows that the running speed of the BLS scheme's digital signature/signature verification algorithm (with a key length of 160 bits) is better than that of the ECDSA scheme. Specifically, BLS's digital signature generation performs at least 69% faster than that of ECDSA, while the signature verification process of BLS is at least 52% faster than ECDSA.

With the same key length (160 bits), the same digital signature, and signature verification data, the BLS short digital signature scheme had a faster execution time than the ECDSA scheme. Moreover, with the larger size of the input data file, the execution time of the BLS short digital signature scheme linearly increased with the input data file size as shown in Fig. 10. This can be explained by two main reasons:

For digital signature function: The number of operations used for the digital signature function of the BLS schema includes a mapping of a point on the curve and a point multiplication kP. Meanwhile, the number of operations used for the digital signature function of the ECDSA scheme includes one kP point multiplication, one inverse operator modulo, and two scalar point multiplications. The DSA of the BLS scheme obviously requires less operations than ECDSA digital signature.

For the signature verification function: The number of operations using signature verification for the BLS scheme includes the Weil non-degenerate bilinear pairing value calculation that uses two points multiplications to calculate the two values u and v. Meanwhile, the number of operations used in the signature verification function of the ECDSA digital signing scheme includes one modulo inverse operator, two points multiplications, and two scalar multiplications. The larger number of operations makes the ECDSA scheme operate slower than the BLS scheme.

5. Conclusions

In this paper, the authors used the calculation technique of Weil non-degenerate bilinear pairing (with $P \in E(F_p)$, $Q \in E(F_{p^l})$ and a higher security multiplier α=12) in building a BLS short digital signature scheme based on supersingular elliptic curves with key generation, digital signature, and digital verification functions. The set of supersingular elliptic curve parameters (with a sufficiently large prime p and a higher security multiplier α=12) initialised for the selected BLS scheme ensures that the signature size is short and the security of the BLS scheme remains theoretically safe. The execution time of the BLS short digital signature program was much improved compared to the ECDSA digital signature scheme, which makes BLS short digital signature scheme a candidate for applications that require short processing time, fast computation, and for devices with low memory and low bandwidth transmission.

CRediT author statement

Nhu-Quynh Luc, Quang-Trung Do, Manh-Hung Le: Data curation, Software, Visualisation, Investigation, Writing original draft preparation, Conceptualisation, Methodology, Supervision.

ACKNOWLEDGEMENTS

The authors are grateful to the Academy of Cryptography Techniques for supporting this work.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] H. Cohen, G. Frey, R. Avanzi, et al. (2005), *Handbook of Elliptic and Hyperelliptic Curve Cryptography*, Chapman and Hall/CRC, DOI: 10.1201/9781420034981.

[2] D. Boneh, B. Lynn, H. Shacham (2001), "Short signatures from the weil pairing", *Advances in Cryptology - CRYPTO 2002*, **2248**, pp.514-532.

[3] S. Wang (2017), *Efficient Computation of Miller's Algorithm in Pairing-Based Cryptography*, Electronic Theses and Dissertations, University of Windsor, 86pp.

[4] M. Masoumi, H. Mahdizadeh (2012), "Efficient hardware implementation of an elliptic curve cryptographic processor over GF(2^{163})", *Int. J. Comput. Electr. Autom. Control Inf. Eng. 2012 Int.*, **6(5)**, pp.725-732, DOI: 10.5281/zenodo.1329671.

[5] D. Moody, R. Peralta, R. Perlner, et al. (2015), "Report on pairing-based cryptography", *J. Res. Natl. Inst. Stand. Technol.*, **120**, DOI: 10.6028/jres.120.002.

[6] A. Markel, L. Nemirovskiy (2014), "Pairing-based short signatures", https:// markel.co/projects/ecc/2/article.pdf, accessed 14 January 2022.

[7] V.S. Miller (2004), "The Weil pairing, and its efficient calculation", *J. Cryptol.*, **17**, pp.235-261, DOI: 10.1007/s00145-004-0315-8.

[8] J. Shallit, P.C.V. Oorschot, S.A. Vanstone (1999), "Handbook of applied crytography", *Am. Math. Mon.*, **106(1)**, DOI: 10.2307/2589608.

[9] S.S. Dhanda, B. Singh, P. Jindal (2020), "Lightweight cryptography: A solution to secure IoT", *Wirel. Pers. Commun.*, **112(3)**, pp.1947-1980.

[10] P.S.L.M. Barreto, H.Y. Kim, B. Lynn, et al. (2002), "Efficient algorithms for pairing-based cryptosystems", *Advances in Cryptology - CRYPTO 2002*, **2442**, pp.354-369, DOI: 10.1007/3-540-45708-9_23.

[11] A.J. Menezes, T. Okamoto, S.A. Vanstone (1993), "Reducing elliptic curve logarithms to logarithms in a finite field", *IEEE Trans. Inf. Theory*, **39(5)**, pp.1639-1646, DOI: 10.1109/18.259647.

[12] J. Shikata, Y. Zheng, J. Suzuki (2000), "Realizing the Menezes-Okamoto-Vanstone (MOV)", *IECE Trans. Fundam.*, **E83-A(4)**, pp.756-763.

[13] R. Barbulescu, P. Gaudry, A. Joux, et al. (2013), "A quasi-polynomial algorithm for discrete logarithm in finite fields of small characteristic", https://arxiv.org/abs/1306.4244.

[14] O. Abid (2012), "New digital signature protocol based on elliptic curves", *Int. J. Cryptogr. Inf. Secur.*, **2(4)**, pp.13-19.

[15] S. Koppula, J. Muthukuru (2016), "Secure digital signature scheme based on elliptic curves for internet of things", *Int. J. Electr. Comput. Eng.*, **6(3)**, DOI: 10.11591/ijece.v6i3.9420.

[16] M.A. Mehrabi, C. Doche, A. Jolfaei (2020), "Elliptic curve cryptography point multiplication core for hardware security module", *IEEE Trans. Comput.*, **69(11)**, pp.1707-1718, DOI: 10.1109/TC.2020.3013266.

[17] M.H.T. Tran, T.T. Ton, T.D. Nguyen, et al. (2017), "Multilinear mappings based on weil pairing over elliptic curves", *2017 4th NAFOSTED Conference on Information and Computer Science*, DOI: 10.1109/NAFOSTED.2017.8108053.

[18] D.P. Le, C.H. Tan (2013), "Improved Miller's algorithm for computing pairings on edwards curves", *IEEE Trans. Comput.*, **63(10)**, pp.2626-2632, DOI: 10.1109/TC.2013.125.

[19] O. Schirokauer, D. Weber, T. Denny (1996), "Discrete logarithms: The effectiveness of the index calculus method", *International Algorithmic Number Theory Symposium*, **1122**, DOI: 10.1007/3-540-61581-4_66.

[20] R. Padmavathy, C. Bhagvati (2010), "Solving the discrete logarithm problem for ephemeral keys in chang and chang password key exchange protocol", *J. Inf. Process. Syst.*, **6(3)**, pp.335-346, DOI: 10.3745/JIPS.2010.6.3.335.

[21] D. Hankerson, A.J. Menezes, S. Vanstone (2004), *Guide to Elliptic Curve Cryptography*, Springer, 312pp.

[22] D.B. Roy, D. Mukhopadhyay (2019), "High-speed implementation of ECC scalar multiplication in GF(p) for generic montgomery curves", *IEEE Trans. Very Large Scale Integr. Syst.*, **27(7)**, pp.1587-1600, DOI: 10.1109/TVLSI.2019.2905899.

[23] C. Costello, P. Longa, M. Naehrig (2016), "Efficient algorithms for supersingular isogeny Diffie-Hellman", *Annual International Cryptology Conference*, **9814**, DOI: 10.1007/978-3-662-53018-4_21.

[24] M. Scott (2005), "Computing the tate pairing", *Cryptographers' Track at The RSA Conference*, **3376**, DOI: 10.1007/978-3-540-30574-3_20.

[25] L.N. Quynh, D.V. Son, M.A. Tuan (2017), "Enhancement of implementing cryptographic algorithm in FPGA built-in RFID tag using 128 bit AES and 233 bit kP multitive algorithm", *VNU J. Sci. Math. - Phys.*, **33(2)**, pp.82-87, DOI: 10.25073/2588-1124/vnumap.4206.

[26] I. Yavuz, S.B.Ö. Yalçin, Ç.K. Koç (2008), "FPGA implementation of an elliptic curve cryptosystem over GF(3^m)", *2008 International Conference on Reconfigurable Computing and FPGAs*, DOI: 10.1109/ReConFig.2008.66.

[27] J. López, R. Dahab (1999), "Fast multiplication on elliptic curves over GF(2^m) without precomputation", *International Workshop on Cryptographic Hardware and Embedded Systems*, DOI: 10.1007/3-540-48059-5_27.

[28] National Institute of Standards and Technology (2013), *Digital Signature Standard (DSS)*, DOI: 10.6028/NIST.FIPS.186-4.

[29] D. Johnson, A. Menezes, S. Vanstone (2001), "The elliptic curve digital signature algorithm (ECDSA)", *Int. J. Inf. Secur.*, **1(1)**, pp.36-63, DOI: 10.1007/s102070100002.

[30] D. Pointcheval, J. Stern (2000), "Security arguments for digital signatures and blind signatures", *J. Cryptol.*, **13(3)**, pp.361-396, DOI: 10.1007/s001450010003.

[31] P.S.L.M. Barreto, B. Lynn, M. Scott (2003), "Constructing elliptic curves with prescribed embedding degrees", *International Conference on Security in Communication Networks*, **2576**, DOI: 10.1007/3-540-36413-7_19.

[32] L.N. Quynh, D.V. Son, M.A. Tuan (2019), "Performance of 697-bit Tate pairing based on Elliptic curve implementation for Spartan6 XC6vlx760-2ff1760 FPGA", *4th International Conference on Advanced Materials and Nanotechnology*, pp.166-169.

 DOI: 10.31276/VJSTE.64(4).10-15

Applying Bayesian neural network to evaluate the influence of specialized mini projects on final performance of engineering students: A case study

Minh Truong Nguyen[1], Viet-Hung Dang[2], Truong-Thang Nguyen[2*]

[1]University of Sciences, Vietnam National University - Hanoi, 334 Nguyen Trai Street, Thanh Xuân Trung Ward, Thanh Xuan District, Hanoi, Vietnam

[2]Faculty of Building and Industrial Construction, Hanoi University of Civil Engineering, 55 Giai Phong Street, Dong Tam Ward, Hai Ba Trung District, Hanoi, Vietnam

Received 10 June 2022; revised 24 August 2022; accepted 8 September 2022

Abstract:

In this article, deep learning probabilistic models are applied to a case study on evaluating the influence of specialized mini projects (SMPs) on the performance of engineering students on their final year project (FYP) and cumulative grade point average (CGPA). This approach also creates a basis to predict the final performance of undergraduate students based on their SMP scores, which is a vital characteristic of engineering training. The study is conducted in two steps: (i) establishing a database by collecting 2890 SMP and FYP scores and the associated CGPA of a group of engineering students that graduated in 2022 in Hanoi; and (ii) engineering two deep learning probabilistic models based on Bayesian neural networks (BNNs) with the corresponding architectures of 8/16/16/1 and 9/16/16/1 for FYP and CGPA, respectively. The significance of this study is that the proposed probabilistic models are capable of (i) providing reasonable analysis results such as the feature importance score of individual SMPs as well as an estimated FYP and CGPA; and (ii) predicting relatively close estimations with mean relative errors from 6.8 to 12.1%. Based on the obtained results, academic activities to support student progress can be proposed for engineering universities.

Keywords: **data, engineering, machine learning, neuron network, project.**

Classification numbers: **1.3, 2.3**

1. Introduction

Nowadays, universities are capable of collecting data with reference to their students in electronic format. As a result, there is an urgent need to effectively transform large volumes of data into knowledge to improve the quality of managerial decisions and to predict academic performance of students at an early stage. As a part of artificial intelligence (AI) techniques recently adopted in a wide variety of human life applications [1, 2], various machine learning (ML) approaches have been increasingly applied to analyse educational data, such as student scores, to concentrate academic assistance on students as well as to improve the university training programs. ML is an especially appealing alternative in the field of engineer training and education as it is difficult or unfeasible to develop conventional algorithms to perform required tasks [3, 4].

S.S.A. Naser, et al. (2015) [4] developed an artifical neural network (ANN) model for predicting student performance at the Faculty of Engineering and Information Technology, Al-Azhar University, based on the registration records of 1407 students using a feed forward back propagation algorithm for training. The model was tested with an overall result of 84.6%. E.Y. Obsie, et al. (2018) [5] developed a neural network model for predicting student cumulative grade point averages for the 8th semester (CGPA8) and designed an application based on the predictive models. The real dataset employed in the study was gathered from 134 students at the Hawassa University School of Computer Science that graduated in 2015, 2016, and 2017. It is shown that the student progress performance, which is measured by CGPA8, can be predicted using scores from their first-, second-, and third-year courses. Z. Iqbal, et al. (2017) [6] utilized collaborative filtering (CF), matrix factorization (MF), and restricted Boltzmann machine (RBM) techniques to systematically analyse real-world data collected from 225 undergraduate students enrolled in the Electrical Engineering program at the Information Technology University (ITU) from which the academic performance of the ITU students was evaluated. It was shown that the RBM technique was better than the other techniques in predicting

*Corresponding author: Email: thangnt2@huce.edu.vn

student performance in the particular course. S.D.A. Bujang, et al. (2021) [7] introduced a comprehensive analysis of machine learning techniques to predict final student grades in first semester courses by improving the performance of predictive accuracy. The performance accuracy of six well-known machine learning techniques, namely, decision tree (J48), support vector machine (SVM), naïve bayes (NB), K-nearest neighbour (K-NN), logistic regression (LR), and random forest (RF) using 1282 student course grades were presented and followed by a multiclass prediction model to reduce the over-fitting and misclassification results caused by imbalanced multi-classification using the Synthetic Minority Oversampling Technique (SMOTE) with a two feature selection method. It was shown that the proposed model integrated with RF had significant improvement with the highest f-measure of 99.5% [7].

It is worth mentioning that most of the aforementioned ML approaches were conducted in a deterministic manner. Hence, there is a need to develop a probabilistic model that is capable of providing well-predicted results as well as estimating the confidence of the results through associated intervals. Such a result is more relevant for experimental data on student exam scores rather than a single point estimation because, even with the same student, scattered results can be obtained from different series of experiments.

In training programs at engineering universities, specialized mini projects (SMPs) play an important role as they progressively provide knowledge as well as accumulate conceiving, designing, implementing, and operating skills necessary for their FYP, which is an integrated topic to solve a practical problem of the particular field of engineer training. This article applies a machine learning approach to predict FYP and final CGPA results from those SMPs based on which the influence of SMPs on the FYP and CGPA can be evaluated in a data-driven manner. A case study is conducted by collecting 2890 datapoints in the form of score results from eight SMPs, one FYP, and the CGPA of a group of 289 engineering students that graduated in 2022 in Hanoi. Then, two deep learning probabilistic models based on BNNs are established for FYP and CGPA predictions. It is shown from the obtained results that the proposed approach is a practical tool providing quick and reasonable analysis results such as the feature importance score of an individual SMP and the estimated FYP and CGPA results. Furthermore, a relatively close estimation can be captured from the BNN model for CGPA, providing useful information for academic management.

2. Stochastic model using BNNs

As pointed out by various authors, a major obstacle to the data-driven method is the scarcity of relevant data, and this problem becomes accentuated when studying the obtained scores of various individual students [8-12]. Even with data in hand, there exist unavoidable deviations between them [13-14]. Thus, this study proposes to engineer a probabilistic machine learning model on the basis of BNNs rather than deterministic ones as done in the reviewed works. The advantage of such a probabilistic model is that it is capable of predicting quantities of interest such as the FYP score or CGPA, as well as estimating the amount of uncertainty that is associated with the prediction values. It is evident that the more data available, the more accurate the model, and vice versa. In summary, the key contribution of this article is to propose a probabilistic ML model to predict the FYP score and CGPA results from the given scores of SMPs so that the effect of an individual SMP as well as FYP on the final performance of students can be evaluated.

We begin by briefly reviewing the ANN to set up mathematical symbols and terminology. Given a dataset $D=[X,Y]$, the ith data sample is denoted by $X_i=[x_{i,1},\ldots,x_{i,n}]$ with n being the number of features. Herein, each feature is an input related to SMP and FYP results. It is desirable to develop a non-linear mapping from X to Y, i.e., $Y_i=f(X_i)$. A standard architecture of ANN consists of an input layer, an output layer, and one or many hidden layers with the total number of layers of the ANN being L. Each layer consists of various neurons that are fully connected with all neurons in neighboring layers. Mathematically, a neuron j at layer l could be described by a linear transformation plus a non-linear activation function, as follows:

$$x_j^l = h\left(\sum_{k=1}^{N_{l-1}} w_{j,k}^l\, x_k^{l-1} + b_j^{l-1}\right) \quad (1)$$

where x_j^l is output of neuron j at layer l; x_k^{l-1} is output of neuron k at the previous layer l-1; $w_{j,k}^l$ is a weight assigned to the connection between the former and the latter; N_{l-l} is the total number of neurons of layer l-1; b_j^{l-1} is a real value to be determined, also known as bias; and h is an activation function. Here, the sigmoid activation function is used to squeeze values into the range (0,1). Moreover, the function is continuous and differentiable everywhere, thus rendering the training process of the neural network via a gradient descent-based algorithm faster and more effective.

By setting W^l as the matrix of weights corresponding to layer l, with l=1,…,L, the ANN could be described by the following equation derived from the description of neural networks in [15]:

$$\hat{Y}_i = F(X_i|W) = f_L(\ldots(f_2(f_1(X_i|\mathrm{W}^1)|\mathrm{W}^2)\ldots|\mathrm{W}^L) \quad (2)$$

where $\hat{Y}_i$ is a prediction of Y_i, and f_l with l=1,…,L denotes transformation operations at layer l in the ANN. The network will be iteratively trained to determine the optimal values of W^l that minimize the discrepancy between $\hat{Y}_i$ and Y_i.

BNN is a probabilistic deep learning model that combines the high prediction performance of ANN with the ability to estimate uncertainty of the Bayes theory [16]. In the authors'

opinion, the model is especially suitable for working with not-so-abundant collected data owing to two reasons: (i) in practice, similar series of experiments with identical input parameters still provide different results due to unavoidable uncertainty; and (ii) fitting an ANN with many parameters to a limited database may cause the over-fitting problem, i.e., ANN is likely to yield low-accuracy results on new data despite being well trained. In other words, it is necessary to not only perform prediction of FYP and CGPA results but also to estimate how much confidence we have about the prediction results.

For this purpose, rather than assigning the deterministic values for weight W of the neural network, BNN uses a Gaussian probability distribution for W as below:

$$W = \mu + \sigma \times \epsilon \ \ with\ \epsilon \sim \mathcal{N}(0, I) \tag{3}$$

where μ and σ denote the mean and standard deviation matrices of W and ϵ is the noise drawn from a zero-mean unity-variance normal distribution. Then, μ, σ are parameters to determine through the learning process.

Note that the output of BNN is a probability distribution, thus a specialized loss function L is required to measure the model's performance. The adopted metric is the Kullback-Leibler divergence (KL) whose formula is:

$$L = KL(q(W|\mu,\sigma)||p(W|D)) = E_{q(W|\mu,\sigma)}\, log\, \frac{q(W|\mu,\sigma)}{p(W|D)} \tag{4}$$

Next, the optimal values of μ^* and σ^* are the solutions of the following minimization problem:

$$\mu^*, \sigma^* = \underset{\mu,\sigma}{\operatorname{argmin}}\, KL(q(W|\mu,\sigma)||p(W|D)) \tag{5}$$

Via the Bayes' rule, $p(W|D)$ can be calculated as below:

$$p(W|D) = \frac{p(D|W)p(W)}{p(D)} \tag{6}$$

Substituting Eq. (6) into Eq. (4), the loss function L is rewritten as follows:

$$L = E_{q(W|\mu,\sigma)} \log p(D) - E_{q(W|\mu,\sigma)} \log p(W) - E_{q(W|\mu,\sigma)} \log p(D|W) \tag{7}$$

This loss function can be approximated from observed discrete data as follows:

$$L = \frac{1}{N_s} \sum_{i=1}^{N_s} [\log q(W_i|\mu,\sigma) - \log p(W_i) - \log p(D|W_i)] \tag{8}$$

where N_s is the total number of samples.

Next, the gradients of the loss function with respect to μ and σ are derived by:

$$\Delta\mu = \frac{\partial L(W|\mu,\sigma)}{\partial W} + \frac{\partial L(W|\mu,\sigma)}{\partial \mu}$$
$$\Delta\sigma = \frac{\partial L(W|\mu,\sigma)}{\partial W} \times \epsilon + \frac{\partial L(W|\mu,\sigma)}{\partial \sigma} \tag{9}$$

Finally, μ and σ are updated using a small learning rate α as follows:

$$\sigma \leftarrow \sigma - \alpha \times \Delta\sigma;\ \ \mu \leftarrow \mu - \alpha \times \Delta\mu. \tag{10}$$

3. Case study - Database on score results of SMPs, FYP and CGPA

In the postgraduate training program of civil engineers, a student is required to pass all the subjects including eight specialized mini projects before being qualified to conduct his/her final year project. The SMPs consist of: Design of Architecture (DoA), Design of Foundation (DoF), Mechanism of Reinforced Concrete Structures (MoRCS); Design of Reinforced Concrete Buildings (DoRCB); Design of Structural Steel Buildings (DoSSB); Construction Technology 1 (CT1); Construction Technology 2 (CT2); and Construction Management (CM) that will be numbered from SMP.1 to SMP.8, respectively. Each individual SMP provides students with the corresponding knowledge and professional skill that will be integrated in his/her FYP to design and build a civil/industrial building in real situation. It is noteworthy that the SMPs' and FYP's exams are all conducted in the form of oral defence. As a result, all the aforementioned SMPs significantly influence the FYP, which together with all theoretical subjects and SMPs contributes to the CGPA (Fig. 1).

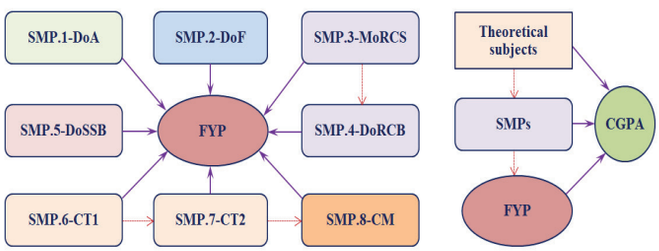


Fig. 1. Projects in training program of civil engineers in HUCE.

In this research, a dataset of 2890 scores of eight SMPs, one FYP, and the CGPA is collected from 289 civil engineers who graduated from Hanoi University of Civil Engineering (HUCE) in 2022. Figs. 2-5 display the histograms of all the score results of their SMPs, FYP, and CGPA on the 4-point scale, showing clear visualization of the range of values as well as their distributions.

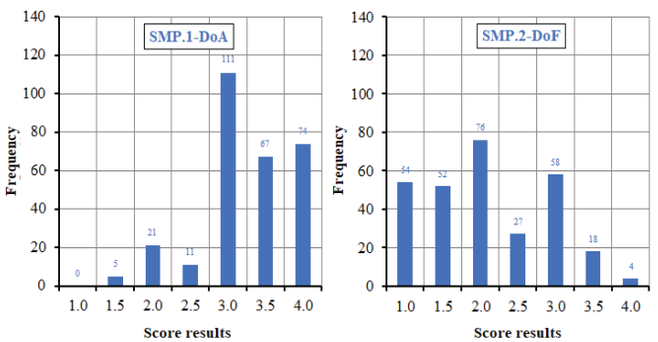


Fig. 2. Score histograms of SMP.1-DoA and SMP.2-DoF.

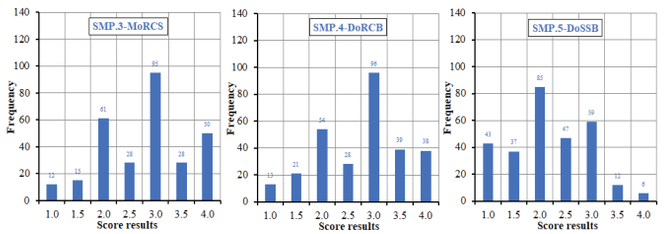


Fig. 3. Score histograms of SMP.3-MoRCS, SMP.4-DoRCB and SMP.5-DoSSB.

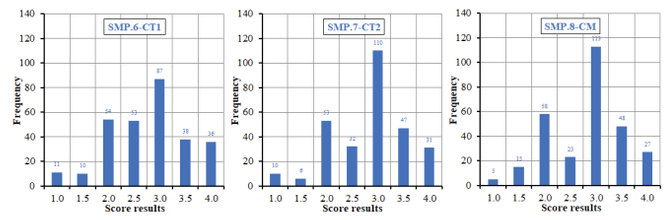


Fig. 4. Score histograms of SMP.6-CT1, SMP.7-CT2 and SMP.8-CM.

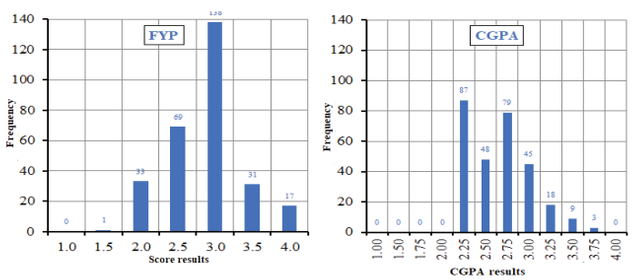


Fig. 5. Histograms of FYP and CGPA results.

It can be seen in Fig. 2 that the score results of SMP.1-DoA of the investigated 289 students were quite high, among which 74 students obtained 4.0, whereas the biggest group was 3.0 with 111 students. There were 67 students that earned a 3.5 and the number of students receiving scores of 2.5, 2.0, and 1.5 were 11, 21, and 5, respectively. It is noted that although this is the first SMP of engineering students in the university program, there is not much calculation in this task of architectural design. The remaining seven SMPs are all critical to students as they are curtailed to train them on the design as well as construction of buildings. It can be observed in the distributions of these SMP scores that the group of 3.0 is almost dominant, except the cases of SMP.2-DoF and SMP.5-DoSSB, of which the dominant group was 2.0 (Figs. 2 and 3). It is shown in Figs. 3 and 4 that the distribution of the main SMPs in the program were similar to each other. Meanwhile, it can be observed in Fig. 5 that the distribution of FYP results was quite standard, whereas there is a regression trend of the number of students with the higher scores in the CGPA results.

4. Analysis results on FYP and CGPA using BNN model

In this study, the adopted architecture of the BNN for FYP scores is 8/16/16/1. The model consists of an input layer with 8 neurons, 2 hidden layers with 16 neurons, and an output layer with one neuron corresponding to the FYP result, as graphically illustrated in Fig. 6A. The neurons in the input layer correspond to the score results of SMP.1 to SMP.8. Since the data size is moderate, it is reasonable to avoid using too deep architectures of many hidden layers as well as wide layers with a significant number of neurons, which may lead to a pronounced increment of parameters to determine. Besides, the number of neurons for the hidden layer is set to 16 since it should be a power of 2 to be convenient for the memory of the computer. For the proposed BNN model, each neuron has two parameters to be determined, characterizing the probability distribution of its weight. At the beginning of learning, they are initialized as a normal distribution with zero mean and unity standard deviation. For prediction of CGPA, the corresponding BNN architecture is 9/16/16/1 in which the FYP score in the dataset is combined with the SMPs as the 9th element of the input layer (Fig. 6B).

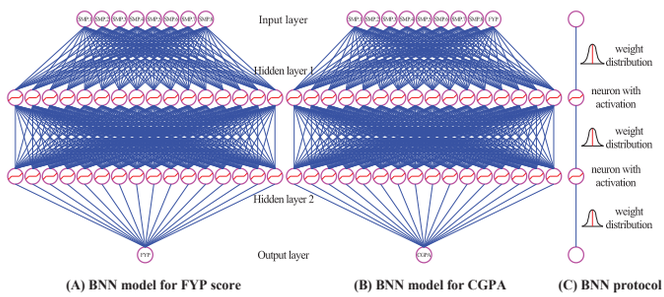


Fig. 6. Graphical representation of BNN whose weights are characterized by probability distributions.

Updating the model's weights as described in Eq.(10) is based on the Adam Optimization Algorithm belonging to the first-order gradient descent optimization family, which gradually adapts the model's weights by a small amount after each iteration to reduce the loss function. The number of updates is controlled through a hyper-parameter of 0.001. This value can also be referred to as the learning rate that was determined via a preliminary test to ensure the learning process is convergent within a reasonable learning time. It is noted that a small learning rate will unnecessarily increase learning time, whereas a large value could lead to premature results. On the other hand, pre-processing standardization is adopted to obviate the scale difference issue between different features with different physical meanings. By utilizing the aid of the deep learning library Pytorch for building the overall framework, the deep probabilistic library Pyro [17] for establishing the BNN-based data-driven model, Pandas for data management, scikit-learn library [18] for data standardization, and Matplotlib for data visualization, the implementation of the proposed data-driven framework can be realized in this study.

In this study, the investigated database is split into three non-overlapping datasets, namely, the training, validation, and testing sub-sets with a ratio of 60:20:20, corresponding to data from 173, 58, and 58 students.

After being built, the probabilistic models were trained with the training database mentioned in the previous section. Figs. 7A and 7B depict the learning curves of the BNN models of FYP and CGPA results, respectively, showing how loss functions evolve versus the number of training iterations (epochs) on both training and validation datasets. In the following paragraphs, the corresponding results of the BNN model for CGPA will be given in parentheses. The KL loss function of FYP (CGPA) quickly dropped for epochs from 0 to 500, before gradually decreasing to values around 0.035 and 0.02 (0.03 and 0.025) on the training and validation datasets, respectively. After that, a steady trend is observed, i.e., no clear improvement is obtained, until the number of epochs reached 2000. For epochs 1300 to 1400, there is less fluctuation in loss function than the other intervals. Hence, an iteration value of 1300 is selected as the final configuration of the BNN model.

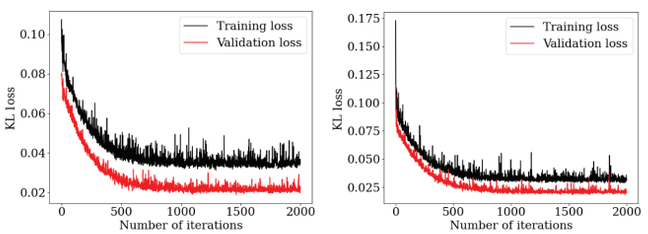


Fig. 7. Evolution of the KL loss function against number of iterations on training and validation datasets for (A) FYP and (B) CGPA.

The effect of each SMP on FYP and CGPA results can be investigated via a feature importance study. Importance scores will be assigned for all SMPs, a high score means that the corresponding SMP has a significant impact on the FYP/CGPA and a low score means there is less impact. The SMPs are ranked based on their importance score. Such results provide understanding about the correlation between the projects and helps students optimize their learning strategy to achieve desired final scores. The permutation feature importance method is used along with the proposed data-driven model. Herein, a feature refers to an SMP score. The simple, yet effective, core idea of this method is to permute the values of features and evaluate the change in prediction errors. A permuted feature means that original values of this feature are shuffled among data samples while other features remain unchanged. If a permuted feature incurs large errors, this feature is important and contributes significantly to the prediction results. Hence, all features will be permuted one by one, and the respective error will be calculated for each case. Next, these errors are sorted in descending order, and importance scores are derived. Since the BNN is a probabilistic model, the feature importance results obtained with BNN are random variables. Thus, one repeats the feature importance calculations 100 times and then derives their statistical characteristics such as mean, min, and max values. The influence of SMPs on a student's final performance on their FYP and CGPA in terms of importance score are shown in Fig. 8A and Fig. 8B, respectively.

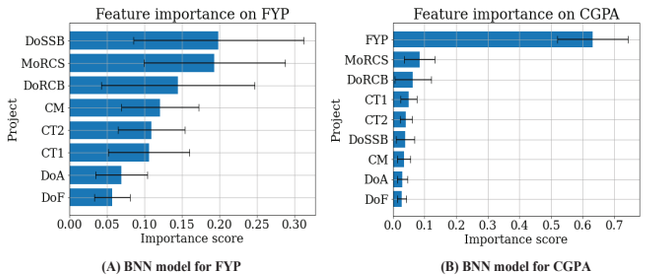


Fig. 8. Feature importance graph of SMPs' influence on (A) FYP and (B) CGPA.

It can be observed from Fig. 8A that, compared to the remaining SMPs, the mini projects of DoSSB and MoRCS have more pronounced influence on FYP results. It is noted that in the training program the SMP and FYP have 1 and 10 credits, respectively (each credit equals to 15 hours of lecturing). It can be seen in Fig. 8B that the influence of the FYP on CGPA, which is counted from a total of 168 credits, is most significant while MoRCS holds the second position. The large gap between the feature importance scores of FYP and MoRCS is reasonable due to their credit relative ratio of 10:1. Fig. 8 also proves that reinforced concrete and steel structures are the leading applications with significant distributions currently in the field of construction.

Next, the final performance of the trained model is evaluated on the test dataset and the prediction results of FYP and CGPA are demonstrated in Fig. 9.

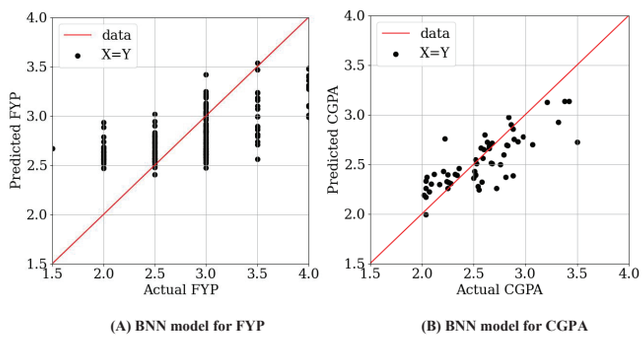


Fig. 9. Prediction results of (A) FYP scores and (B) CGPA.

It is shown in Fig. 9 that for each data point (shown in the dot symbol), its X-coordinate denotes true FYP scores from the database while the Y-coordinate is a value predicted by the model. Ideally, a perfect model will provide the same results as those from the database, as highlighted by the red 45-degree line. While the BNN model for FYP is not as close as expected in Fig. 9A, the

predicted points of CGPA lie relatively close to the ideal line in Fig. 9B. Specifically, the mean relative errors of predicted FYP and CGPA results are 12.1% and 6.8%, respectively. These results qualitatively confirm the viability of the proposed BNN model for CGPA.

5. Conclusions

This study applied a data-driven method for assessing the FYP score and CGPA based on the results of eight SMPs, which are considered as a characteristic of the training program for engineering students. Probabilistic machine learning models based on the BNN were introduced and the theoretical foundation of the model and key parameters of the proposed approach were described, followed by a case study using a database of 2890 score results of SMPs, FYP, and CGPA from a group of civil engineers that graduated in 2022 in Hanoi. One of the main results of the proposed model is that it can be utilized to evaluate the influence of an individual SMP on a student's final performance in terms of FYP and CGPA. It was shown from the results of the case study that the subjects commonly applied in practice also contribute a more significant influence. In addition, the BNN model for CGPA is capable of providing relatively close predictions with a mean relative error of 6.8%. Furthermore, the application of the data-driven model is straightforward as it is built based on open source libraries and the user-friendly programming language, Python, without requiring any specialized software.

For the next step of the study, other models such as straight artificial neural network and drop-out neural network can also be incorporated for comparison purposes. Furthermore, one can complement the database with the score results of all students that graduated before, during, or after the COVID-19 pandemic to gain an overall picture to propose appropriate solutions to improve academic activities of engineering universities, where the projects play very important role for undergraduate students.

CRediT author statement

Minh Truong Nguyen, Viet-Hung Dang, Truong-Thang Nguyen: Data curation, Software, Visualisation, Investigation, Writing original draft preparation, Conceptualization, Methodology, Supervision.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] T. Mitchell (1997), *Machine Learning*, McGraw-Hill Education, 432pp.

[2] J. Hu, H. Niu, J. Carrasco, et al. (2020), "Voronoi-based Multi-robot autonomous exploration in unknown environments via deep reinforcement learning", *IEEE Transactions on Vehicular Technology*, **69(12)**, DOI: 10.1109/TVT.2020.3034800.

[3] Y. Bengio, Y. LeCun, G. Hinton (2015), "Deep learning", *Nature*, **521(7553)**, pp.436-444, DOI: 10.1038/nature14539.

[4] S.S. Abu-Naser, I.S. Zaqout, M. Abu-Ghosh, et al. (2015), "Predicting student performance using artificial neural network: In the Faculty of Engineering and Information Technology", *International Journal of Hybrid Information Technology*, **8(2)**, pp.221-228, DOI: 10.14257/ijhit.2015.8.2.20.

[5] E.Y. Obsie, S.A. Adem (2018), "Prediction of student academic performance using neural network, linear regression and support vector regression: A case study", *International Journal of Computer Applications*, **180(40)**, pp.39-47, DOI: 10.5120/ijca2018917057.

[6] Z. Iqbal, J. Qadir, A.N. Mian, et al. (2017), "Machine learning based student grade prediction: A case study", https://arxiv.org/abs/1708.08744.

[7] S.D.A Bujang, A. Selamat, R. Ibrahim, et al. (2021), "Multiclass prediction model for student grade prediction using machine learning", *IEEE Access*, **9**, pp.95608-95621, DOI: 10.1109/ACCESS.2021.3093563.

[8] S.T. Jishan, R.I. Rashu, N. Haque, et al. (2015), "Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique", *Decis. Analyst.*, **2(1)**, DOI: 10.1186/s40165-014-0010-2.

[9] A. Polyzou, G. Karypis (2016), "Grade prediction with models specific to students and courses", *Int. J. Data Sci. Analyst.*, **2(3-4)**, pp.159-171, DOI: 10.1007/s41060-016-0024-z.

[10] I. Khan, A.A. Sadiri, A.R. Ahmad, et al. (2019), "Tracking student performance in introductory programming by means of machine learning", *Proc. 4th MEC Int. Conf. Big Data Smart City (ICBDSC)*, pp.1-6, DOI: 10.1109/ICBDSC.2019.8645608.

[11] M.A.A. Barrak, M. Al-Razgan (2016), "Predicting students final GPA using decision trees: A case study", *Int. J. Inf. Educ. Technol.*, **6(7)**, pp.528-533, DOI: 10.7763/IJIET.2016.V6.745.

[12] E.C. Abana (2019), "A decision tree approach for predicting student grades in research project using WEKA", *Int. J. Adv. Comput. Sci. Appl.*, **10(7)**, pp.285-289.

[13] F. Ahmad, N.H. Ismail, A.A. Aziz (2015), "The prediction of students' academic performance using classification data mining techniques", *Appl. Math. Sci.*, **9**, pp.6415-6426, DOI: 10.12988/ams.2015.53289.

[14] T. Anderson, R. Anderson (2017), "Applications of machine learning to student grade prediction in quantitative business courses", *Glob. J. Bus. Pedagog.*, **1(3)**, pp.13-22.

[15] N. Andrew (2019), *Stanford University CS229: Lecture Notes*, 216pp.

[16] C.M. Bishop (1997), "Bayesian neural networks", *Journal of The Brazilian Computer Society*, **4**, pp.61-68.

[17] E. Bingham, J.P. Chen, M. Jankowiak, et al. (2019), "Pyro: Deep universal probabilistic programming", *The Journal of Machine Learning Research*, **20(1)**, pp.973-978.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, et al. (2011), "Scikit-learn: Machine learning in Python", *The Journal of Machine Learning Research*, **12(85)**, pp.2825-2830.

DOI: 10.31276/VJSTE.64(4).16-27

Recent mass ejection from AGB star W Hya

D.T. Hoai*, P.T. Nhung, P. Darriulat, P.N. Diep, N.B. Ngoc, T.T. Thai, P. Tuan-Anh

Department of Astrophysics, Vietnam National Space Center, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Nghia Do Ward, Cau Giay District, Hanoi, Vietnam

Received 3 August 2022; revised 9 October 2022; accepted 21 October 2022

Abstract:

We analyse archival ALMA (Atacama Large Millimeter/submillimeter Array) observations of two molecular line emissions, ^{12}CO(3-2) and ^{29}SiO(8-7), from oxygen-rich AGB (Asymptotic Giant Branch) star W Hya. Together with results of earlier VLT observations at visible and infrared wavelengths, our results suggest a two-component picture of the morpho-kinematics of the circumstellar envelope (CSE), one stable over time, at the scale of centuries, and the other variable, at the scale of years. The stable component consists of an approximately spherical shell of gas and dust expanding radially to a terminal velocity of ~5 km s^{-1} at a distance of ~30 au from the star. It is found to display comparable features as seen in the CSE of R Dor, a star similar to W Hya. The variable component projects on the plane of the sky over a region confined to the neighbourhood of the star and elongated toward the north. Its very high density and sudden acceleration suggest an interpretation in terms of mass ejection initiated a few years ago. We discuss its properties in relation with earlier observations of dust formation in the same region. Our results offer a picture of the wind of W Hya that differs significantly from the picture that could be suggested by earlier analyses, giving evidence for a mass ejection that had been previously overlooked and underscoring new relations between the dust and gas emissions. They have an impact on the evidence published earlier for the presence of CO masers. They favour an interpretation in terms of convective cell ejections playing the main role in the generation of the nascent wind, the stable component of the CSE being seen as the result of many successive such events occurring in different directions at short time intervals.

Keywords: **circumstellar matter, radio lines: stars, stars: AGB and post-AGB, stars: individual: W Hya.**

Classification number: **2.1**

1. Introduction

W Hya is an oxygen-rich AGB star at a distance of only 104^{+14}_{-11} pc from the Sun [1]; a shorter distance of 78±6 pc [2] is often used in the published literature. It is a semi-regular variable with a period of ~361 days [3], often quoted as a Mira [4] belonging to spectral class M7.5e-M9ep. Single dish observations suggest a mass-loss rate at the level of ~10^{-7} M$_\odot$yr^{-1} [5, 6] and a main sequence mass between 1 and 1.5 solar masses [7, 8].

The close neighbourhood of the star has been recently observed with the Very Large Telescope (VLT) at wavelengths covering from visible to near-infrared, using NACO [9], SPHERE/ZIMPOL [10-12], AMBER/VLTI [10, 11, 13] and MIDI/VLTI [14]. These observations have given evidence for a clumpy and dusty layer, displaying important variability at short time scale. Dust grains, mostly aluminium composites, are found to have sizes ranging between 0.1 and 0.5 µm. W.H.T. Vlemmings, et al. (2017) [15], W.H.T. Vlemmings, et al. (2019) [16] using the Atacama Large Millimeter-submillimeter Array (ALMA) to observe continuum and CO(v=1, J=3-2) emissions close to the star have given evidence for shocks induced by pulsations and convective cell ejections, producing an effective line broadening of the molecular emission observed on lines of sight close to the centre of the star [17].

Some of the most relevant observations are displayed in the Appendix (Table A1). These observations have made it possible to refine and improve our earlier understanding of the morpho-kinematics of the CSE of W Hya, which T. Khouri, et al. (2015) [18] had summarized in a model based on a large number of earlier observations over a broad range of wavelengths, including from Herschel and other single-dish telescopes [6, 7]. In particular, the analysis of K. Ohnaka, et al. (2016) [10], K. Ohnaka, et al. (2017) [11] suggests the presence of a shell of large transparent grains, such as of Al_2O_3, with a 550 nm optical depth of 0.6±0.2 and inner and outer radii of 1.3 R$_*$ (R$_*$=25 mas) and 10±2 R$_*$, respectively; it gives evidence for time variability, with predominance of small (0.1 µm) grains at minimum light and larger (0.5 µm) grains at pre-maximum light; it gives also evidence for anisotropy with a dust density enhancement of a factor ~4 in the north, over a cone of ~45° half-opening angle. These results have been confirmed by ALMA millimetre observations of the emission of AlO lines [19], found to be confined within ~3 R$_*$ from the centre of the star, in contrast with the emission of the ^{29}SiO(8-7) line, observed to extend to much larger radii. Together with A. Takigawa, et al. (2019) [20], these authors suggest that the dust shell near the star (2 to 3 R$_*$) is dominated by alumina-rich dust grains, and that the overall wind acceleration, up to some 5.5 km s^{-1}, occurs beyond this shell where silicate dust does not form efficiently, SiO molecules remaining mostly in the gas phase; such inefficient silicate formation is claimed to be consistent with the SiO/AlO ratio estimated from mid-infrared dust emission, where only ~6% of the Si atoms should condense as silicate dust if all Al atoms are bound in aluminium oxide.

Finally, W.H.T. Vlemmings, et al. (2021) [21] have recently observed CO masers at a Doppler velocity of 5.5 km s^{-1} using ALMA millimetre observations of the ^{12}CO(3-2) line emission.

*Corresponding author: Email: dthoai@vnsc.org.vn

The present article revisits earlier analyses of archival ALMA observations of the ^{29}SiO(8-7) and ^{12}CO(3-2) molecular emissions in order to draw a significantly improved picture of the morpho-kinematics of the CSE of the star.

2. Observations and data reduction

The present work uses archival observations of W Hya from project ADS/JAO.ALMA#2015.1.01446.S (PI: A. Takigawa), which were carried out for a total of ~2 hours on source between 30 November and 5 December 2015 (stellar pulsation phase φ=0.3) with ALMA in Cycle 3. The antennas, respectively 33 and 41 in number, were configured in such a way that the baseline lengths were distributed in two groups, one covering between ~15 m and ~200 m, the other between ~400 m and ~8 km; the former group included 10 antennas in a circle of 100 m radius providing insufficient *uv* coverage for addressing properly the short spacing problem. As a result, reliable imaging is limited to projected angular distances smaller than 0.15-0.20 arcsec from the star [17]. The emissions of the ^{29}SiO(8-7) line, with a frequency of 342.9808 GHz, and of the ^{12}CO(3-2) line, with a frequency of 345.7960 GHz, were observed in different spectral windows, with channel spacing of 0.854 km s^{-1} and 0.4236 km s^{-1}, respectively. The data have been calibrated with CASA[1] using standard scripts provided by ALMA without continuum subtraction. Imaging, using GILDAS[2]/IMAGER[3] with robust weighting parameter of 1, produces beams of 52×38 mas^2 (PA=99^o) and 50×42 mas^2 (PA=101^o) for the SiO and CO lines respectively. Noise levels (σ) are of ~2 and ~3 mJy beam^{-1}, respectively.

The observations of the CO line have been analysed earlier by W.H.T. Vlemmings, et al. (2017, 2019, 2021) [15, 16, 21] who used the data in conjunction with observations of a second epoch (project ADS/JAO.ALMA#2016.A.00029.S) on 25 Nov 2017 with an on-source observing time of ~35 min. The observations of the SiO line have been analysed earlier by A. Takigawa, et al. (2017) [19], T. Danilovich, et al. (2019) [22] and D.T. Hoai, et al. (2021) [17]. Details related to the observations, calibration and reduction of the data have been presented in these publications.

We use coordinates centred at the peak of continuum emission [16], x pointing east, y pointing north and z pointing away from Earth. The projected distance to the star is calculated as $R=\sqrt{(x^2+y^2)}$. Position angles, ω, are measured counter-clockwise from north. The Doppler velocity, V_z, spectra are referred to a systemic velocity of 40.4 km s^{-1}.

3. Morpho-kinematics of the CSE

3.1. A radially expanding shell

Doppler velocity spectra integrated over the circle R<0.2 arcsec are displayed in the left panels of Fig. 1. Contrasting with the small angular range being covered, these spectra probe the CSE along the line of sight over large distances where temperature and density conditions allow for emission of the observed lines. Both spectra show a clear peak at V_z~−5 km s^{-1}, SiO in absorption and CO in emission. This suggests the presence of a gas layer between star and observer, optically thin for CO and thick for SiO emissions, over which the wind has reached a terminal velocity of ~5 km s^{-1}: a similar situation is observed in several other AGB stars as illustrated in the right panels of Fig. 1 by the case of R Dor [23] where the terminal velocity is ~4 km s^{-1}. In both stars, the CO spectrum shows a small emission peak at terminal velocity on the red-shifted side suggesting that the expanding layer is in fact part of a spherical shell surrounding the star.

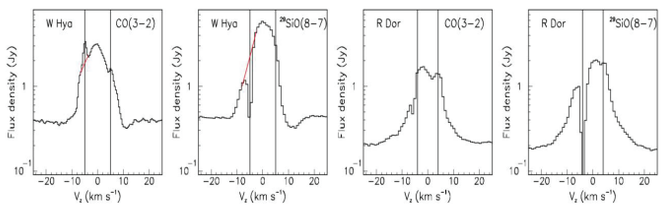


Fig. 1. Doppler velocity spectra integrated over the circle R<0.2 arcsec for the CO(3-2) (left) and ^{29}SiO(8-7) (centre-left) line emissions of W Hya. The red lines show the interpolation over the peak range of the emission observed on the sides of the peak. The right panels show the same line emissions for R Dor, integrated over a circle R<0.35 arcsec and divided by a factor 3.1, accounting for the different distances (104 pc for W Hya vs 59 pc for R Dor). Vertical lines are shown at V_z=±5 km s^{-1} for W Hya and ±4 km s^{-1} for R Dor.

However, in the case of W Hya, such interpretation raises a few questions: a) The Doppler velocity of ~5 km s^{-1} is the same as the velocity where W.H.T. Vlemmings, et al. (2021) [21] observe a CO maser; b) The wind terminal velocity evaluated by T. Khouri, et al. (2014a, 2014b) [6, 7] is 7.5 km s^{-1} rather than 5 km s^{-1}; c) While similar, the spectra of the two stars display small but significant differences; in particular the CO emission peak of W Hya is narrower than that of R Dor and the SiO absorption peak is shallower. We address these questions in the remaining of the present section.

Spectral maps of both lines in a square $|x,y|$<75 mas centred on the star are displayed in Fig. A1 of the Appendix. Significant continuum contribution is seen to extend to R~50 mas, in agreement with earlier analyses [15, 16]. In all panels, the distributions show a peak between ~−4 and ~−6 km s^{-1}, in emission for the CO line and in absorption for the SiO line. Under the peak, the CO spectra display a broad distribution, covering between V_z~−10 and +10 km s^{-1}; red-shifted absorption is seen over the stellar disc, particularly strong for the SiO line, suggesting the presence of in-falling gas. Also shown in this figure is a same set of spectral maps for the ^{29}SiO(8-7) emission of R Dor [23], scaled to account for the difference in distances. It displays remarkable similarity with the W Hya set as illustrated in Fig. 2 for the spectra covering the stellar discs. Yet, significant differences are visible: the absorption peak associated with the terminal velocity is broader for R Dor than for W Hya and the broad absorption spectrum underneath the peak is more red-shifted for W Hya than for R Dor.

[1]https://casa.nrao.edu/.
[2]https://www.iram.fr/IRAMFR/GILDAS/.
[3]https://imager.oasu.u-bordeaux.fr/.

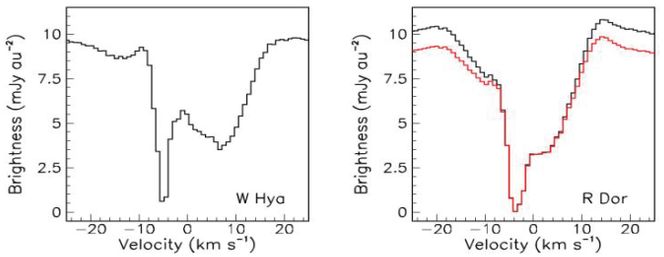


Fig. 2. Comparing the spectral distributions of the central brightness in W Hya (left) and R Dor (right). The difference in distances is taken into account by proper scaling and the difference in beam sizes (53×40 mas^2 for W Hya and 41×35 mas^2 for R Dor) is taken into account by averaging over 30×30 mas^2 for W Hya and over both 45×45 mas^2 (black) and 55×55 mas^2 (red) for R Dor, providing an estimate of the attached uncertainty.

3.2. Terminal velocity

The absence of evidence for a terminal velocity exceeding ~5 km s^{-1} may cast doubt on the evaluation of 7.5 km s^{-1} proposed by T. Khouri, et al. (2014a) [6]. These authors fit single dish infrared to sub-millimetre observations of CO line profiles (Herschel, APEX, SMT and SEST) using simple non-LTE radiative transfer modelling with an initial CO/H_2 abundance of 4×10^{-4}. The best fit gives a terminal velocity of 7.5±0.5 km s^{-1}, a turbulent velocity of 1.4±0.1 km s^{-1} and a radial dependence of the temperature of the form T[K]=2500×(23/r[mas])$^{0.65\pm0.05}$. It requires a photo-dissociation radius of the CO molecules 2.5 times smaller than expected from G.A. Mamon, et al. (1988) [24] modelling and a slow start of the wind acceleration for expansion velocities below 5.5-6.0 km s^{-1}, followed by a fast injection of momentum to reach the terminal velocity. These results contrast with those obtained by A. Takigawa, et al. (2017) [19], who suggest that acceleration takes place beyond an AlO-rich dust shell that covers between 50 and 70 mas from the centre of the star and reaches quickly a terminal velocity of 5.5 km s^{-1}. Also, W.H.T. Vlemmings, et al. (2021) [21] find it difficult to adopt the radial dependence of the expansion velocity proposed by T. Khouri, et al. (2014a) [6] when modelling the CO masers: their model requires a larger velocity at the inner envelope, ~4.2 km s^{-1}, a steeper acceleration and a larger turbulent velocity component.

In summary, large uncertainties are attached to the results obtained by T. Khouri, et al. (2014a) [6] using single dish line profiles: the Doppler velocity spectra of the ^{12}CO(3-2) and ^{29}SiO(8-7) ALMA observations presented here can reasonably be interpreted as evidence for a terminal velocity of 5-6 km s^{-1} rather than 7.5 km s^{-1}. In this context, we remark that the escape velocity from a solar mass star is ~20 km s^{-1} beyond the dust shell and reaches 5 km s^{-1} at only ~700 mas: The precise form of the radial dependence of the velocity in this interval depends on the details of the forces acting on the dust grains and of their collisions with the gas molecules, in a way which is far from being well understood. We also note that the observation of narrow peaks at terminal velocity implies that this velocity has been reached early enough to dominate much of the explored regions of the line of sight. We remark that an OH maser observed in 1986 at Nançay [25] is at this same velocity of 5-6 km s^{-1}.

Of course, we cannot exclude that further acceleration, beyond 5 km s^{-1}, takes place at distances from the star where the emission of the CO line has become too weak to be detected, because the density or/and the temperature have become too low. This however would be irrelevant to the arguments made in the present article, the expansion velocity of ~5 km s^{-1} being effectively terminal over the explored radial range.

3.3. Peak emission of the CO line

Between ~−6 and ~−4 km s^{-1}, CO emission and SiO emission (Fig. 1) show an emission and an absorption peak, respectively. In order to obtain an approximate evaluation of what they correspond to, we evaluate their contribution by subtracting the linear interpolation over the peak range of the emission observed on the sides of the peak (Fig. 1). The result is mapped in Fig. 3, giving evidence for very similar morphologies, the CO intensity being positive and the SiO intensity negative. The right panel of Fig. 3 gives indeed evidence for a clear correlation between the peak emissions of the two lines. This is a surprising result in relation with the maser interpretation proposed by W.H.T. Vlemmings, et al. (2021) [21] for the emission of the CO line: it is not clear why the morphology observed for the CO maser emission should match the region of SiO absorption.

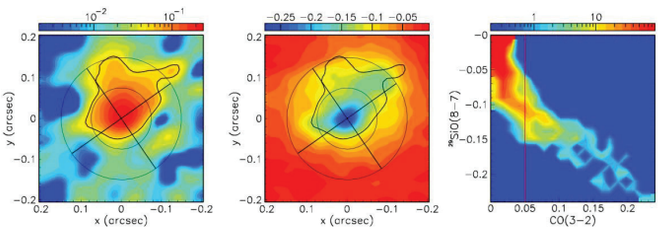


Fig. 3. Left: W Hya CO intensity map of the peak integrated between −4 and −6 km s^{-1} with linearly interpolated background subtracted, colour scale is Jy beam^{-1} km s^{-1}; Centre: same as left but for SiO, the contour of the CO line at 50 mJy beam^{-1} km s^{-1} is shown on both the left and central panels and defines region C; Right: correlation between the intensities mapped in the left and central panels. For each pixel, we plot the CO intensity in abscissa and the SiO intensity in ordinate, the colour scale is in number of pixels per bin. The red line shows the intensity associated with the CO contour displayed in the left and central panels.

The peak emission displayed in Fig. 3 shows clear anisotropy, with enhancement in a wedge covering approximately from 305° to 35° counter-clockwise in position angle. We refer to this region of enhanced emission as "region C"; concretely, for convenience, we define it as the region where the intensity of the emission of the CO peak exceeds 50 mJy beam^{-1} km s^{-1}. It qualitatively matches the anisotropy of the dust distribution observed by K. Ohnaka, et al. (2016, 2017) [10, 11] at two epochs bracketing the observation of the molecular lines (8 July 2015 and 23 March 2016 compared

with end of November 2015). Moreover, it also matches the observation of a blue-shifted blob of SiO emission by D.T. Hoai, et al. (2021) [17] at Doppler velocities around 10 km s^{-1}, for which these authors failed to find a sensible interpretation. As illustrated in Fig. 4, which displays channel maps of both the CO and SiO lines in the range of relevant Doppler velocities, a similar blob is observed in the CO emission, however at less negative Doppler velocities.

3.4. Overall view

In order to shed light on these observations, we consider separately 8 regions covering the environment of the star within 150 mas (6 R_*) projected distance from its centre: four quadrants covering $0<R<75$ mas and four others covering $75<R<150$ mas. The quadrants, shown in Fig. 3, are numbered in increasing order from north counter-clockwise, from 1 to 4 for the inner quadrants and from 5 to 8 for the outer quadrants. Quadrants 1 and 5 cover the wedge $|\omega-350^o|<45^o$. Fig. 5 displays, for each quadrant and for each of the CO and SiO lines, the observed Doppler velocity spectrum. Together with the inner quadrants, outer quadrant 5 covers region C where the CO peak emission is enhanced (Fig. 3) and where the blue-shifted blobs are observed (Fig. 4). The inner quadrants cover the range where dust is being formed and starts accelerating the gas [11, 19] and where shocks anticorrelated with the dust clouds are seen to heat the extended atmosphere of the star [15]. Tables 1 and 2 list a number of parameters that describe the spectra displayed in Fig. 5. They are described as the sum of three components: a flat continuum emission averaged over $15<|V_z|<26$ km s^{-1}, a broad component and a peak, positive for CO and negative for SiO. To separate the contribution of the peak from that of the broad component we define a peak interval as $4<|V_z|<6$ km s^{-1} over which we evaluate the contribution of the broad component by linear interpolation from the sides of the peak interval. The peak is then redefined as the difference between the observed distribution and this linearly interpolated distribution, the broad component being defined accordingly. Tables 1 and 2 list the continuum level and, for the broad component, its value at maximum, its FWHM and its central velocity; for the peak, Table 1 lists the peak emission at maximum and Table 2 the emission at maximal absorption. Uncertainties are typically ±1 km s^{-1} on Doppler velocities and ±10 mJy on flux densities.

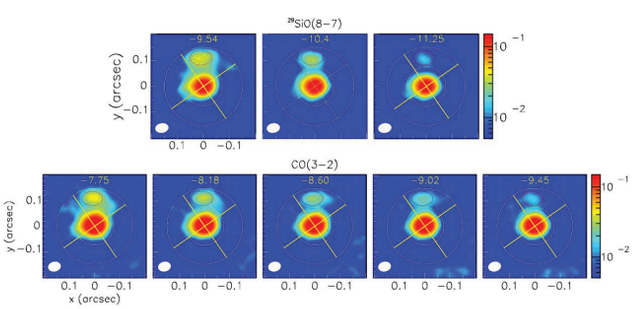


Fig. 4. Blue blob emission: Channel maps of the ^{29}SiO(8-7) emission around −10.4 km s^{-1} and of the CO emission around −8.6 km s^{-1}, the colour scales are in Jy beam^{-1}.

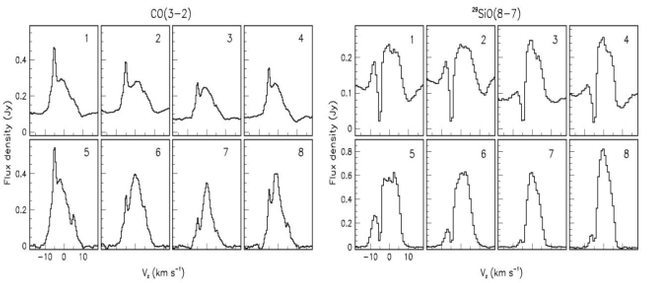


Fig. 5. Doppler velocity spectra (Jy) integrated in each of 8 quadrants for the ^{12}CO(3-2) emission (left) and ^{29}SiO(8-7) emission (right) of W Hya. The quadrants cover the projected radial ranges $0<R<75$ mas (upper row of each pair) and $75<R<150$ mas (lower row of each pair), respectively. In each row the quadrants are numbered in increasing order starting from north counter-clockwise, from 1 to 4 for the inner quadrants and from 5 to 8 for the outer quadrants. Quadrants 1 and 5 cover position angles $|\omega-350^o|<45^o$.

Table 1. Parameters describing the ^{12}CO(3-2) spectra of quadrants 1 to 8.

Quadrant	Continuum (Jy)	Peak height (Jy)	Broad emission		
			Maximum (Jy)	*FWHM (km s^{-1})*	*Central velocity (km s^{-1})*
1	0.11	0.25	0.18	10	–2
2	0.12	0.19	0.16	11	+1
3	0.07	0.11	0.18	10	–1
4	0.08	0.18	0.19	11	–2
5	0	0.28	0.37	10	–2
6	0	0.12	0.40	9	0
7	0	0.09	0.35	6	0
8	0	0.15	0.40	6	–2

Table 2. Parameters describing the ^{29}SiO(8-7) spectra of quadrants 1 to 8.

Quadrant	Continuum (Jy)	Flux at maximal absorption (mJy)	Broad emission		
			Maximum (Jy)	*FWHM (km s^{-1})*	*Central velocity (km s^{-1})*
1	0.12	25	0.24	14	–1
2	0.14	25	0.23	13	0
3	0.08	25	0.24	12	0
4	0.10	25	0.25	12	–1
5	0	50	0.58	13	0
6	0	50	0.62	11	1
7	0	50	0.62	8	0
8	0	50	0.82	9	0

The same exercise using sextants rather than quadrants is reported in the Appendix (Fig. A2, Tables A2 and A3) where a parameterized description of the spectra, in the form of a flat continuum and a Gaussian for each of the peak and broad component, is presented. We used it to cross-check the results obtained from the quadrant analysis. These results call for a number of comments.

Continuum is seen to contribute to the inner quadrants exclusively and the levels observed in the SiO and CO bands agree to a precision of ±7 mJy km s^{-1} $arcsec^{-2}$. They display anisotropy comparable with the dust distribution observed by K. Ohnaka, et al. (2017) [11], namely a depression in the south-western direction: the ratio between the average emission in sextants 4 and 5 and that in the other four sextants (see Appendix) is 58% and the ratio between the average emission in quadrants 3 and 4 and that in quadrants 1 and 2 is 67%. This seems to contradict the detailed analysis of W.H.T. Vlemmings, et al. (2017) [15], which gives evidence for a south-western hot spot dominating the asymmetry of the continuum emission.

The emission peaks of the CO spectra are at −5.1±0.2 km s^{-1} and the absorption peaks of the SiO spectra at −4.7±0.2 km s^{-1}. The difference is 0.4±0.2 km s^{-1}, barely significant and possibly associated with an extra acceleration in the region probed by the CO line emission compared with that probed by the SiO line emission. The rms deviation from the peak mean velocity is 1.0±0.4 km s^{-1} for CO and 0.8±0.2 km s^{-1} for SiO. The red-shifted peak of CO emission is precisely at +5 km s^{-1}, supporting the shell picture.

The amplitude of the CO emission peak is 3 times larger in quadrants 1 and 5 than in the opposite quadrants, 3 and 7. The quadrants in between have intermediate peak heights: at the level of 3/4 of quadrant 1 in the inner circle and at the level of 1/2 of quadrant 5 in the outer ring. Accounting for the factor 3 difference in area between the inner circle and the outer ring, the mean brightness of the peak is 4.5 times larger in the inner quadrants than in outer quadrants 6 to 8. These asymmetries quantify the anisotropy illustrated in the left panel of Fig. 3. They suggest an enhancement of density and/or an anticipated reach of the terminal velocity in region C. We discuss this further in the next section.

The minima of the absorption peaks of the SiO spectra are at the approximate level of ~25 mJy in the inner circle and ~50 mJy in the outer ring, independently from the position angle. They reveal the important optical thickness and provide information on temperature that is further discussed later.

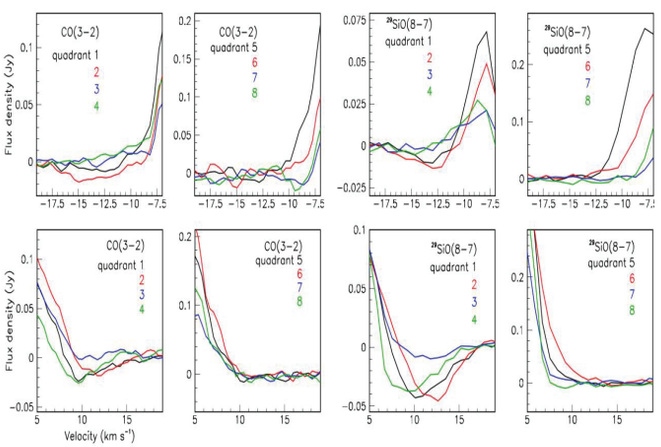


Fig. 6. Continuum-subtracted blue-shifted parts (upper panels) and red-shifted parts (lower panels) of the spectra (Jy) displayed in Fig. 5.

The emission above continuum and outside of the terminal velocity interval, between ~−6 and ~−4 km s^{-1}, is centred at ~−1 km s^{-1} for the CO line and at ~0 for the SiO line, with a precision (rms) of ±1 km s^{-1}. In quadrants 1 to 6, it has a FWHM of 10±1 km s^{-1} for the CO line and of 12.5±1.0 km s^{-1} for the SiO line. In quadrants 7 and 8, the FWHM is only 2/3 of these values. The flux density at maximum displays little dependence on position angle. In the inner circle, it is ~0.18 and ~0.24 Jy for the CO and SiO lines, respectively and in the outer ring, ~0.38 and ~0.66 Jy. In terms of brightness, on average, the outer ring brightness is ~83% of that of the inner circle. Altogether, this emission displays less anisotropy and less inhomogeneity than in the terminal velocity interval, the main deviations from uniformity being a broader spectrum for SiO emission than for CO emission and for the C region than opposite to it. Broader SiO than CO spectra are commonly observed in other oxygen-rich AGB stars, the difference being caused by shorter distances to the star, where the linewidth is larger, being probed by SiO emission than by CO emission.

The high V_z wings of the CO and SiO emissions are illustrated in Fig. 6. They show an important excess of flux density on the blue-shifted side of quadrant 5. The contribution of the blue blob in this region is shown in Fig. 7, which displays the difference between the Doppler velocity spectrum integrated over the blob region (75<R<150 mas, |ω|<30°) and that integrated over the rest of the 75<R<150 arcsec ring and divided by 5 to account for the broader ω interval. The blob is seen as a blue-shifted source of intense emission centred on the terminal velocity and covering some ~10 km s^{-1}. The CO emission shows a narrow peak at terminal velocity over a broader distribution, each reaching ~0.1 Jy at maximum. The SiO emission is more difficult to interpret. The persistence of an absorption peak at terminal velocity implies that the SiO line is much more opaque over the blob than on the rest of the ring. The structure seen at Doppler velocities exceeding ~−3 km s^{-1} results from the different shapes displayed by the spectra (see more Fig. 5).

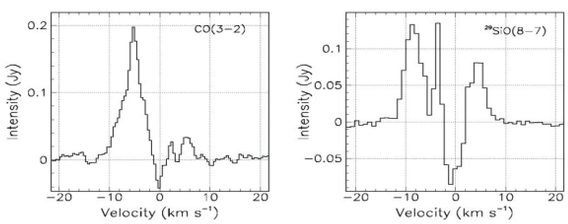


Fig. 7. Difference between the Doppler velocity spectrum integrated over the blob region (75<R<150 mas, |ω|<30°) and that integrated over the rest of the 75<R<150 mas ring and divided by 5 to account for the broader ω interval. Left panel is for the CO line and right panel for the SiO line.

4. Temperature dependence and radiative transfer modelling of the observed line emissions

4.1. Formalism

In the LTE approximation, for a given velocity V_z and a brightness I_0 [Jy $arcsec^{-2}$] on the entrance side of a thin slab of column density N [molecules cm^{-3} arcsec] and temperature T [K], the brightness I [Jy $arcsec^{-2}$] at the exit side, reads:

$$I = I_0\, e^{-\tau} + S(1-e^{-\tau}) = S + (I_0 - S)e^{-\tau} \quad (1)$$

where S is the source function defined as $S=\varepsilon/\tau$ with ε[Jy arcsec^{-2}] describing the unabsorbed emission of the slab and τ being its optical thickness. In the LTE approximation, S is the Planck function: $S=\varepsilon/\tau=(2h\ f^3)/(c^2)(e^{\Delta E/T}-1)-1=f\ [GHz]^3/(2.9\times10^7)\ (e^{\Delta E/T}-1)^{-1}$ Jy arcsec^{-2}. The frequency f and the energy ΔE of the transition take similar values for the CO and SiO lines: 345.80 and 342.98 GHz, and 16.7 and 16.6 K, respectively. When the column density increases, I varies from I_0 to ε/τ. In case of absorption, $I_0>\varepsilon/\tau$, I is always larger than ε/τ, a lower limit to the emission of an optically thick shell: however large absorption may be, the outer part of the layer cannot be prevented from shining. When the energy of the transition is much smaller than the temperature T, one can approximate $\varepsilon/\tau\sim f^3T/(2.9\ 10^7\Delta E)\sim T[K]/11.8$.

Figure 2 shows that the minimal brightness observed in a 30×30 mas^2 region is ~0.5 mJy au^{-2}=5.5 Jy arcsec^{-2}, which must therefore exceed T[K]/11.8, implying T<65 K (75 K using the exact relation). The same result is obtained with the minimal flux density of ~25 mJy observed in each of the 4.4×10^{-3} arcsec^2 inner quadrants. This result is independent from the value of ε, namely of the column density. At r=1 arcsec, this temperature is only one third of that obtained from the law used by T. Khouri, et al. (2014a) [6] and A. Takigawa, et al. (2017) [19] close to the star: T[K]=2500×(23/r[mas])$^{0.65}$, suggesting a steeper radial decrease of the temperature than commonly assumed in the published literature.

The emission parameter ε [Jy arcsec^{-2}] is obtained from Relation 2 below:

$$\varepsilon = hf/(4\pi\Delta f)(ndz)\ A_{ji}\ f_{pop} = hc/(4\pi\Delta V)(ndz)\ A_{ji}\ f_{pop}$$
$$= 5.5\ 10^3\ (\Delta V)^{-1} d\ N\ A_{ji}\ f_{pop}\ ,\ f_{pop}=k(2J+1)e^{-E\,u/T}/T \quad (2)$$

where d [pc] is the distance from the Sun to the star, J and E_u [K] the spin and energy of the upper level, A_{ji} [s^{-1}] the Einstein coefficient, k [K] the ratio between the temperature and the partition function, N [molecules cm^{-3} arcsec] the column density, Δf and ΔV the line widths in frequency and velocity respectively and T [K] the temperature. The parameters for the CO and SiO lines are given in Table 3. The values of ε_0 and τ_0 for ΔV=1 km s^{-1} listed in the last columns are defined from the expressions $\varepsilon=\varepsilon_0 Ne^{-Eu/T}/T$ and $\tau=\tau_0 Ne^{-Eu/T}/T(e^{\Delta E/T}-1)$. The dependence on temperature of ε/N and τ/N is illustrated in Fig. 8.

Table 3. Parameters related to the emissions of the CO and SiO lines. The values of frequency f, transition energy ΔE, upper level energy E_u, emission coefficient A_{ji} and partition function are taken from CDMS [26].

Line	*F* [GHz]	ΔE [K]	E_u [K]	2J+1	A_{ji} [s^{-1}]	*k* [K]	*d* [pc]	ε_0	τ_0
^{12}CO(3-2)	345.80	16.7	33.19	7	2.50×10^{-6}	2.77	104	27.7	19.4
^{29}SiO(8-7)	342.98	16.6	74.08	17	21.0×10^{-4}	1.03	104	21.0×10^3	16.0×10^3

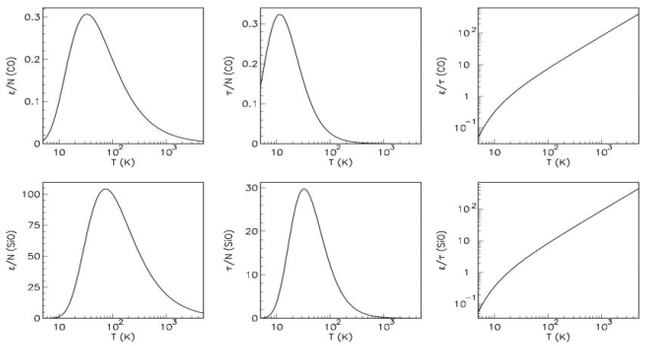


Fig. 8. Dependence on temperature of *ε/N* (left), *τ/N* (centre) and *ε/τ* (right) for the ^{12}CO(3-2) (upper panels) and ^{29}SiO(8-7) (lower panels), respectively. Units are Jy arcsec^{-2} for ε and mol. cm^{-3} arcsec for *N*.

4.2. Comparing W Hya and R Dor outside region C

Figure 9 compares the CO and SiO spectra in quadrants 3 and 7, opposite to region C, of W Hya to the equivalent for R Dor [23], divided by 3.1 and integrated over a circle R<0.13 arcsec and a ring 0.13<R<0.26 arcsec, to account for the different distances from the Earth and divided by 4 to compare with a quadrant. In the outer ring, which is free of region C, the SiO R Dor spectra are seen to be lower (0.25 compared to 0.62 Jy) than the W Hya spectra and broader (width at 1/5 maximum ~15 km s^{-1} compared to ~10 km s^{-1}). Qualitatively, however, the similarity between the two spectra suggests that a same dynamics is at stake in both stars. The broader linewidth of the R Dor spectra, as noted earlier by D.T. Hoai, et al. (2021) [17], suggests a stronger impact of shocks induced by pulsations and convective cell ejections in R Dor than in W Hya away from region C. The difference in intensity (~3.7 vs ~6.2 Jy km s^{-1}) suggests a higher density for W Hya than for R Dor. In the inner circle, the difference between W Hya and R Dor spectra is qualitatively the same as in the outer ring.

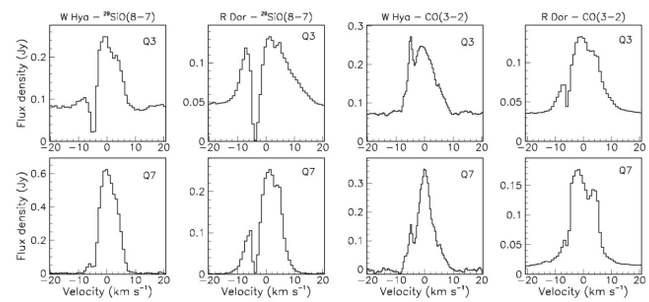


Fig. 9. Comparison between W Hya and R Dor, ^{29}SiO(8-7) and CO(3-2) lines, in quadrants 3 and 7, opposite to region C, as specified on top of the upper panels. The R Dor data have been scaled down for the different distance from the Earth. The upper row is for the inner circle (quadrant 3), the lower row for the outer ring (quadrant 7).

The CO emission spectra are smeared by the broader beam size in the R Dor case (180×140 mas^2 instead of 52×40 mas^2), making the comparison with W Hya difficult.

In the remaining of this section, we use a very simple LTE model to compare the CSE of R Dor with that of W Hya in quadrant 7, away from region C. We restrict the analysis to the SiO data for R Dor, the beam size being too large for a reliable comparison with the CO data. The R Dor data have been scaled to correct for the different distances from the Earth and to account for the solid angle coverage of quadrant 7. The model assumes that the wind velocity V is radial and increases linearly from r=50 mas (~5 au or ~2 R_*) to a terminal velocity V_{term} at $r=r_{term}$. The radial decrease of the temperature is taken of the form $T=T_0/(r/0.05)^p$ where T_0 is the temperature at r=50 mas. The CO density takes a value d_0 at r=1 arcsec and scales in inverse proportion to V_{r2}. The SiO density is taken as ρ times the CO density. Its depletion is approximately accounted for by limiting the layer radius to 1.5 arcsec instead of 3 arcsec for CO. The line width is described in an ad hoc and very approximate way with Gaussian forms of different σ values, as one would do for turbulence, although it is expected to result from shocks induced by pulsations and convective cell ejections in the innermost layers. In particular, the presence of in-falling gas is simply ignored. Beyond a distance $r=r_{turb2}$, we assume a constant $\sigma=V_{turb2}$. Below this distance, we assume a radial dependence of σ of the form $\sigma = V_{turb1}\exp[-0.5(3r/r_{turb})^2]$. Moreover, even closer to the star, at distances smaller than r_{turb}/q_r, we assume a broader value of $\sigma=q_V V_{turb1}$.

The main results are illustrated in Fig. 10 and summarised in Table 4 together with the rms differences Δ between the model flux densities and those observed. We stress that a reliable evaluation of the model parameters would require data cubes covering a larger solid angle and the observation of at least another line of each of the two molecular species; in particular the numbers quoted after the ± sign should not be understood as uncertainties but simply as measures of the sensitivity of the best fit to the values taken by each parameter when the others are kept fixed. The quality of the fits is satisfactory. The main differences between the SiO emissions of the two stars are a steeper radial dependence of the temperature (ρ=1.44±0.04 instead of 0.74±0.05) and a larger region of line broadening (r_{turb}=0.84±0.03 instead of 0.31±0.07) for R Dor than for W Hya. The value of ρ obtained for W Hya, $(10\pm3)\times10^{-3}$, corresponds to an abundance ratio $^{28}SiO/^{12}CO$~13%, consistent with values commonly adopted in the published literature (e.g. M. Van de Sande, et al. (2018) [27] quote ~30% for R Dor). In both cases, the terminal velocity is reached close to the star.

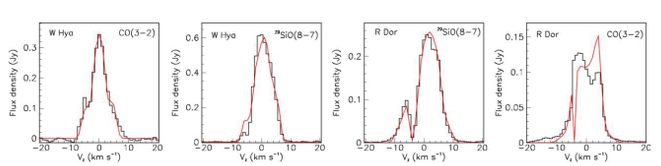


Fig. 10. Comparison between CO and SiO spectra observed in quadrant 7 with best fit model predictions. The pair of left panels is for W Hya and the pair of right panels for R Dor (properly scaled to quadrant 7). In the three leftmost panels the red curve shows the result of the best fit. In the rightmost panel, the red curve is the model prediction for the CO spectrum using the parameters obtained for the best fit to the SiO spectrum.

Table 4. Parameters of the radiative transfer fits in quadrant 7 for W Hya and R Dor, the latter being scaled as described in the text.

	T_0 [K]	p	V_{term} [km s^{-1}]	r_{term} [arcsec]	d_0 [mol/cm^3]	r_{turb} [arcsec]	V_{turb1} [km s^{-1}]	q_V	q_r	V_{turb2} [km s^{-1}]	ρ [10^{-3}]	Δ [mJy]
W Hya	1970 ±120	0.74 ±0.05	5.0 ±0.3	0.32 ±0.02	79±4	0.31 ±0.07	1.9 ±0.3	2.2 ±0.7	3.15 ±0.30	0.92 ±0.20	10±3	CO:12 SiO:33
R Dor	1730 ±75	1.44 ±0.04	4.15 ±0.10	0.19 ±0.02	47±6	0.84 ±0.03	1.83 ±0.09	2.47 ±0.15	6.3 ±0.4	<0.2	8.5±1.1	9

The values quoted after the ±sign measure the sensitivity of the fit to the particular parameter: they correspond to a 10% increase of the χ^2 per degree of freedom, the other parameters being kept at their best fit values. They should not be interpreted as uncertainties.

The fit to the R Dor SiO spectrum, although of apparently better quality than that to the W Hya spectrum, is actually less reliable, being unconstrained by the CO data. In particular, the value obtained for V_{turb2} is unreasonably low. Moreover, the values of ρ and d_0 are correlated, only their product matters (0.79 molecules cm^{-3} for W Hya and 0.40 molecules cm^{-3} for R Dor). The CO spectrum predicted for R Dor using the parameters of the best fit to the SiO spectrum is compared with observation in the rightmost panel of Fig. 10. In order to account approximately for the larger beam size, we have subtracted from the observed quadrant 7 spectrum a fraction of the observed quadrant 3 spectrum such that the resulting continuum level cancels. The high V_z wings are well described but the bulk emission is slightly red-shifted in the model instead of blue-shifted in the data. Moreover, the model predicts a sharp absorption near V_z=4 km s^{-1}, which is absent from the data. This is caused by the low V_{turb2} value obtained from the best fit to the SiO spectrum and by the known difference of V_{term} for the SiO and CO emissions, interpreted by P.T. Nhung, et al. (2021) [23] as evidence for persistence of some acceleration at large distances. Yet, these are small differences considering the crudeness of the exercise, the overall properties of the spectrum (mean and rms values of V_z, total flux density) being well predicted.

In summary, while very crude, this exercise has reliably shown that the morpho-kinematics of the CSE of W Hya, when observed outside region C, is qualitatively similar to that of R Dor and can be described by a model accounting for the main features usually displayed by oxygen-rich AGB stars.

4.3. W Hya inside the C region

In order to shed some light on the morpho-kinematics at stake in region C of W Hya, we show in the left panel of Fig. 11, for both CO and SiO emissions, the difference between the Doppler velocity spectra observed in quadrants 7 and 5. Both are in the outer ring, the former outside of region C and the latter inside. The result is very similar to that obtained in Fig. 7 for the blob region, defined as $|\omega|<30^o$ instead of $|\omega-350^o|<45^o$. It gives evidence for intense blue-shifted emission accompanied by strong line broadening. With the idea that this is the effect of a source confined to short distances from the star in the blue-shifted hemisphere, we modify the model used to produce the spectra illustrated in Fig. 10 by keeping all parameters unchanged except for a narrow layer defined as $0>z>-0.25$ arcsec ($0>z>-25$ au) where we adjust

density, temperature and velocity in order to reproduce the CO and SiO spectra observed in quadrant 5. The results are displayed in the right panels of Fig. 11 and Table 5. The quality of the fits is again satisfactory given the extreme crudeness of the model: the rms differences Δ between the model flux densities and those observed are 39 mJy for the CO spectrum and 47 mJy for the SiO spectrum. The main outstanding parameter value is the CO density, d_0=930±160 molecules cm^{-3}, an order of magnitude larger than outside region C. Also remarkably different is the line width, V_{turb1}=3.75±0.50 km s^{-1} instead of 1.9±0.3 km s^{-1} outside region C. The velocity of the layer is 6.5±0.5 km s^{-1}, and the other parameters take similar values as outside region C. Because of the high value of the density, the SiO layer is extremely opaque and the value of ρ is largely undefined: its best value is 37×10^{-3} but any value larger than 20×10^{-3} gives an acceptable fit.

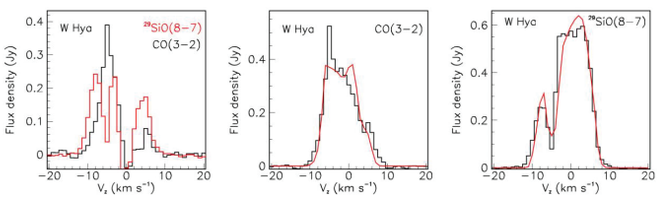


Fig. 11. Left: difference between the Doppler velocity spectra observed in quadrants 7 and 5 of W Hya for ^{12}CO(3-2) (black) and ^{29}SiO(8-7) (red) emissions. Middle and right: quadrant 5 observed spectra (black histograms, CO middle and SiO right) are compared with the results of the fit described in the text (red lines).

Table 5. Parameters of the radiative transfer fits in quadrants 5 and 7 of W Hya (same format as Table 4).

	V(layer) [km s^{-1}]	d_0 [mol/cm^3]	r_{turb} [arcsec]	V_{turb1} [km s^{-1}]	T_0 [K]	p	ρ [10^{-3}]
Quadrant 5 (inside C)	6.5±0.5	930±60	3.75±0.5	0.38±0.05	1970±200	0.71±0.09	>20
Quadrant 7 (outside C)	no layer	79±4	1.9±0.3	0.31±0.07	1970±120	0.74±0.05	10±3

5. Continuum emission

Two sources contribute to continuum emission: the star itself and hot dust surrounding it. Observations by K. Ohnaka, et al. (2016) [10], K. Ohnaka, et al. (2017) [11] using SPHERE/ZIMPOL on the VLT and bracketing the ALMA observations analysed in the present article have given evidence for anisotropic dust formation, with an enhancement in the northern hemisphere and a depression in the south-western quadrant. Observation of H_α emission has given evidence for shocks anti-correlated with dust formation. Moreover, the polarisation pattern observed using SPHERE/ZIMPOL displays the circular symmetry characteristic of light scattered from a spherical shell and the polarised intensity gives further evidence for a south-western depression of the dust density.

The level of continuum emission obtained from the analyses of SiO and CO line emissions presented in the present article, confined within ~50 mas from the centre of the star, is similar to that of R Dor (after correction for the difference in distances) in agreement with earlier evaluations of the star temperatures, ~2700 K for W Hya [16] and ~3000 K [28] for R Dor. As the continuum emission of W Hya has been studied in detail by W.H.T. Vlemmings, et al. (2017) [15], we did not attempt to add new contributions to their results. However, our results seem to be partly inconsistent with theirs and we find it useful to explain this apparent inconsistency. As illustrated in the right panel of Fig. 12, these authors show the presence of a very hot and compact spot on the south-western quadrant of the stellar disc, within 20 mas from the centre of the star, and interpret it as evidence for a shock produced by a convective cell ejection; the spot temperature is at the level of 50000 K, compared with 2500 K for the temperature of the photosphere and 2900 K for that of the heated molecular gas layer surrounding it. As illustrated in Fig. 12, we find no evidence for such a hot spot; on the contrary, we observe a small depression of continuum emission in the south-western quadrant.

We recall that W Hya was observed on three days 5 December, 3 December and 30 November 2015. The data are stored in three sets 1, 2 and 3 respectively. They are calibrated using the script provided by the ALMA staff. The maximal baseline is 7.7 km for set 1, 6.3 km for set 2 and 10.8 km for set 3. There are three spectral windows, two of 937.5 MHz and 960 channels were centred on 330.85 (spw0) and 342.85 GHz (spw1), and a window of 1.875 GHz and 3840 channels was centred on 344.95 GHz (spw2). The continuum emission displayed in the left panel of Fig. 12 uses the most extended array and is averaged over spw2, excluding lines identified by using task *uv_prev* in IMAGER/GILDAS. The image has been produced with CASA using uniform weighting. The beam is 28×20 mas^2 (FWHM) and the map is produced with a circular restoring beam of 28 mas.

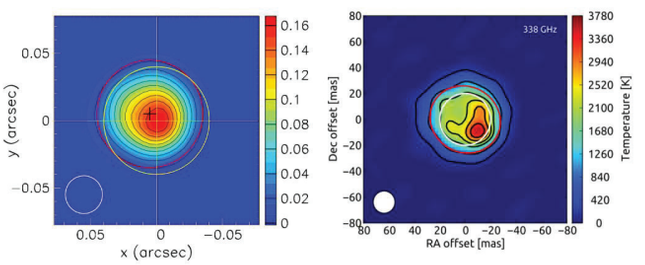


Fig. 12. Continuum emission of W Hya. Left: image obtained from data set 3. The yellow circle, centred on the origin of coordinates at RA=13h49m01.9386s and Dec=−28d22m04.489s, is centred at the maximum of the emission. The red circle is instead centred (black cross) on the boundary of the emission, as is the red circle in the right panel. It is shifted ~5 mas north and ~5 mas east. The right panel shows the image obtained by W.H.T. Vlemmings, et al. (2017) [15]. In both panels the beams are shown in the lower left corners. The colour scales are in Jy beam^{-1} for the left panel and in Kelvin for the right panel, the conversion factor between these being 13,600 K per Jy beam^{-1}.

The right panel of Fig. 12 shows the image obtained by W.H.T. Vlemmings, et al. (2017) [15]. In addition to the flagged antennas indicated in the standard script, they flagged one additional antenna which had a high system temperature. They performed two rounds

of phase self-calibration and one round of phase and amplitude self-calibration on the continuum. They obtained a beam with uniform weighting of 17×24 mas^2 and the final image is restored with a circular beam of 17 mas, smaller than the major axis.

However, the apparent contradiction between the two results boils down to what is taken as precise star location. In our case, we assume that it is the maximum of the emission, which is slightly shifted south-west with respect to the centre of the boundary of the emission, approximately by 5 mas west and 5 mas south. In such a case, we conclude that the emission is slightly depressed in the south-western quadrant. W.H.T. Vlemmings, et al. (2017) [15], who use a smaller beam than we do, obtain an image that strongly suggests that the maximum of the emission corresponds to a hot spot, and that the star location is instead at the centre of the boundary of the emission. It is therefore essentially the use of a smaller beam, possibly together with two successive self-calibrations, that makes the difference of interpretation. We note that a same comment applies to the interpretation of the images obtained by K. Ohnaka, et al. (2016) [10], K. Ohnaka, et al. (2017) [11].

6. A global view

Accounting for the observations made in the previous sections, the present section is an attempt at drawing a consistent picture of the morpho-kinematics of the CSE of W Hya.

The star is surrounded by an expanding layer of gas and dust, which, to first order, is optically thick for ^{29}SiO(8-7) emission and optically thin for ^{12}CO(3-2) emission. In contradiction with earlier estimates [6], the terminal velocity is ~5 km s^{-1} instead of 7.5 km s^{-1}. It has been reached early, at probably some ~12 R_* (0.3 arcsec) from the centre of the star, where the escape velocity is still 7.7 km s^{-1}: this implies that the acceleration of the gas by collisions with dust grains manages to keep the velocity approximately constant over a broad radial range. Such early acceleration to the terminal velocity is consistent with the conclusion reached by A. Takigawa, et al. (2017) [19], who claim that aluminium-rich grains have formed within some 3 R_* (75 mas) from the centre of the star and are efficient at boosting rapidly the gas to radial velocities at the 5 km s^{-1} scale. It is supported by the narrow width of the peaks observed in emission for CO and in absorption for SiO in Figs. 2 and 5: if acceleration were progressive these peaks would tend to trail in the red-shifted direction. Note that masers can be invoked to explain the narrow CO emission peak, but they would not explain the narrow SiO absorption peaks. Moreover, the observation of a peak at +5 km s^{-1} in some panels of Figs. 2 and 5 supports the picture of a spherical shell. For such narrow peaks to be present, this layer must have been stable for many years, a century at least. Comparison with R Dor has shown remarkable similarities and has given evidence for the validity of a picture accounting for the main features usually displayed by oxygen-rich AGB stars to properly describe the morpho-kinematics of the CSE of W Hya outside region C.

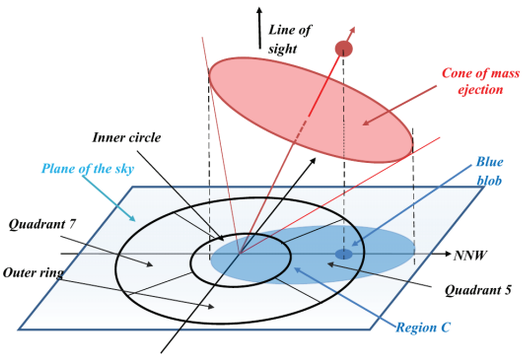


Fig. 13. Simplified schematic of a possible geometry.

The intense blue-shifted emission observed in region C displays features of a mass ejection that occurred recently, at the scale of a few years, and is associated with the blue blob. Fig. 13 shows a simplified schematic of a possible geometry: based on the confinement of region C on the sky plane and on the crude radiative transfer considerations made in this paper, it assumes that the impact of the mass ejection covers a broad cone with an axis making an angle of some 30° with the line of sight, implying that the source associated with the blue blob is located some 0.2 arcsec (20 au) above the plane of the sky. At a velocity of ~10 km s^{-1}, such distance is covered in ~10 years. Within the cone, and within radial distances confined to less than 25 au from the centre of the star, the number density is increased by an order of magnitude, with a SiO/CO ratio larger by at least a factor 2, and the effective line width is also about twice as large. The mean radial velocity inside the cone is 6-7 km s^{-1}. The cone geometry matches the dust distribution observed by K. Ohnaka, et al. (2017) [11]. However, it evolves at a time scale of a few years while the variability observed by these authors is at the scale of a few months. Such a picture impacts significantly the evidence for CO maser emission claimed by W.H.T. Vlemmings, et al. (2021) [21] but does not exclude some possible contribution.

Within the spectral interval of terminal velocity, −6 to −4 km s^{-1}, both over region C and outside, the expanding spherical layer discussed under a) is essentially optically thick for the SiO line and the observed emission probes exclusively its external shell: it makes therefore no difference whether one looks over region C or out of it. What makes the difference in the depth of the absorption peak illustrated in the central panel of Fig. 3 is instead the emission outside the spectral interval of terminal velocity. There, the opacity of the SiO line is much less and the emission over region C is enhanced by the additional contribution of the emission of region C. It is this enhancement, rather than the intensity of the absorption, that explains the correlation illustrated in the right panel of Fig. 3.

Outside the spectral interval of terminal velocity, we observe the emissions of the CO and SiO lines from gas having not yet

reached the terminal velocity and/or from gas expanding in a direction significantly different from the line of sight. Such gas is confined within at most 30 au from the centre of the star (at such a distance, gas having deviated from the line of sight by a projected distance R=150 mas makes an angle of 30° with the line of sight and its Doppler velocity is 87% of its radial velocity). It is seen in emission for both the CO and SiO lines with the exception of some red-shifted absorption of SiO line emission over the inner quadrants, which we interpret as evidence for in-falling gas. Important line broadening is observed in this component of the CSE, witness of the turbulent regime induced by shocks from pulsations and cell ejections. It reaches typically ±10 km s^{-1}, however covering a smaller radial distance in W Hya than it does in R Dor as had been already noted by D.T. Hoai, et al. (2021) [17] and has been confirmed by the analysis presented earlier. The spectra are significantly narrower outside region C, where a depression of dust formation has been reported by K. Ohnaka, et al. (2017) [11].

7. Conclusions

In summary, the picture that emerges from all above observations is made of two components, one stable over time, at the scale of at least a century, the other variable, at the scale of months and years. It differs significantly from the picture that was suggested by earlier analyses. In particular, it gives evidence for a recent mass ejection, which had been overlooked by all former studies and it underscores relations between the dust emission observed at the VLT and the gas emission observed at ALMA.

The stable component consists of an approximately spherical envelope of gas and dust expanding radially and reaching rapidly, at a radius of ~0.3 arcsec=30 au=12R_*, a terminal velocity of ~5 km s^{-1}. The close environment of the star hosts a complex chemistry of gas and dust, displaying both stable and variable features. In particular, W.H.T. Vlemmings, et al. (2017) [15] have given evidence for gas having to stay close to the star photosphere for at least 10^3 years and K. Ohnaka, et al. (2017) [11] have demonstrated the variability of dust formation.

The variable component covers a broad cone that projects on the plane of the sky over region C, elongated and shifted with respect to the star in the NNW direction (~10° west of north). It is characterized by a high density and displays features of a mass ejection that started a few years ago. It is somehow related to the variable features observed earlier: dust formation enhanced in the same region [11, 19] and evidence for shocks in the south-western quadrant [11, 15]. The details of this relation are still in need of being clarified.

While consistent with, and inspired by many of the conclusions reached by earlier authors from studies of the close vicinity of the star, in particular W.H.T. Vlemmings, et al. (2017) [15], K. Ohnaka, et al. (2017) [11] and A. Takigawa, et al. (2017) [19], this picture contradicts the model proposed by T. Khouri, et al. (2014a) [6] that implies a progressive acceleration to a terminal velocity of 7.5 km s^{-1}. It also impacts significantly the evidence for CO maser obtained by W.H.T. Vlemmings, et al. (2021) [21] and provides useful clarifications on the presence of a hot spot south-west of the stellar disc discovered by W.H.T. Vlemmings, et al. (2017) [15].

As first argued by W.H.T. Vlemmings, et al. (2017) [15] this picture favours convective cell ejections as playing a major role in the generation of the nascent wind, the stable component of the CSE being seen as the result of many successive such events occurring in different directions at short time intervals. Major progress in modelling the relevant dynamics has been recently reported [29]. If confirmed by future studies, such a picture raises several questions; in particular, the detailed relation between the features observed by W.H.T. Vlemmings, et al. (2017) [15] and K. Ohnaka, et al. (2017) [11], including the observed anti-correlation between shocks and dust formation, needs to be clarified. Moreover, the nearly exact compensation of the star gravity by the acceleration produced by the stellar light on dust grains, required to make the gas velocity stay constant well before having escaped the star gravity, needs to be understood.

While making use of many observations and analyses reported earlier, the present article offers a new and significantly improved description of the morpho-kinematics of the wind of W Hya. We failed to think of another sensible interpretation than presented here; in particular, invoking the presence of an evaporating planet in relation with the blue blob and the high Doppler velocity wings, as described in a recent publication [30], cannot be seriously pursued in the present case. Most of our knowledge of the morpho-kinematics of the CSE of W Hya is based on observations confined to the very close neighbourhood of the star, less than 0.2 arcsec projected distance.

To confirm the validity of the picture inferred from these observations, it is essential to observe the molecular line emission of several lines, such as CO, SiO, SO, SO_2, HCN and some of their isotopologues, over angular distances at the scale of a few arcsec. Each individual star has its own specific identity and progress in understanding the dynamics at stake in the generation of the nascent wind requires a detailed exploration of its properties. Moreover, new observations of the close neighbourhood of the star, of a same quality as those made at or near the end of 2015, and possibly including line emissions probing higher temperatures, such as $SO_2(34_{3,31}\text{-}34_{2,32})$, would secure and refine our understanding of the variable dynamics.

APPENDICES

Table A1 lists some of the most relevant observations and analyses of the close neighbourhood of the star available in the published literature and Fig. A1 displays spectral maps for ^{12}CO(3-2) and ^{29}SiO(8-7) emissions in both W Hya and R Dor. Figure A2 displays Doppler velocity spectra of the ^{12}CO(3-2) and ^{29}SiO(8-7) emissions within 150 mas projected distance from the centre of W Hya, separated in twelve sextants. Tables A2 and A3 list the parameters describing these spectra, as defined in the caption of Fig. A2.

Table A1. Some of the most relevant observations and analyses of the close neighbourhood of the star available in the published literature.

R [R*]	~1	1-2	2-3	3-6
D.T. Hoai, et al. (2021) [17]				
^{29}SiO(8-7)	Inner quadrants (sextants)			Outer quadrants (sextants)
^{12}CO(3-2)	Broad linewidth			Blue blob
W.H.T. Vlemmings, et al. (2017) [15]				
Continuum ALMA	Photosphere 2500 K			
	Hot spot SW			
CO v=1 ALMA	Warm molecular gas layer 2900 K	Cool gas in-fall and outflow 900 K		
K. Ohnaka, et al. (2017) [11]				
VLT SPHERE ZIMPOL		Clumpy dust clouds Depressed SW	Polarization on dust Shell structure	
		Hα: shocks anti-correlated with dust clouds	Weaker H_α	
		TiO emission		
VLTI/AMBER		CO lines		
A. Takigawa, et al. (2017) [19]				
AlO ALMA		Depletion AlO gas Formation AlO dust SW depressed		
^{29}SiO ALMA		Blob revealing acceleration		Progressive depletion

Table A2. Parameters describing the Doppler velocity spectra observed in the CO sextants, as described in the caption of Fig. A2.

Sextant	1	2	3	4	5	6	7	8	9	10	11	12
a (mJy)	74	60	82	47	35	70	0	0	–4	2	1	1
b(mJy)	110	100	120	110	110	120	210	230	280	190	220	130
c	3.71	4.63	4.15	3.61	2.93	2.52	4.13	4.44	2.69	2.80	3.08	4.13
d(mJy)	170	110	120	79	69	190	220	77	70	71	89	210
e	–5.23	–5.20	–5.05	–5.00	–4.89	–4.88	–5.10	–5.03	–5.43	–5.01	–4.90	–3.38
f	1.18	0.74	0.84	0.79	1.43	1.65	1.48	0.39	0.78	0.52	0.60	2.40

Table A3. Parameters describing the Doppler velocity spectra observed in the SiO sextants, as described in the caption of Fig. A2.

Sextant	1	2	3	4	5	6	7	8	9	10	11	12
a(mJy)	79	66	93	52	39	76	–7	–7	1	–3	–9	–3
b(mJy)	81	76	86	112	124	109	420	430	440	420	540	510
c	4.17	4.85	4.38	3.58	3.25	3.26	4.96	4.74	3.86	3.43	3.27	3.96
d(mJy)	–110	–110	–140	–90	–74	–112	–247	–220	–200	–150	–160	–250
e	–4.93	–4.84	–4.82	–4.75	–4.76	–4.88	–4.91	–4.82	–4.59	–4.33	–4.40	–4.80
f	0.76	0.90	0.88	0.91	0.92	0.66	0.60	0.97	1.03	1.11	1.00	0.58

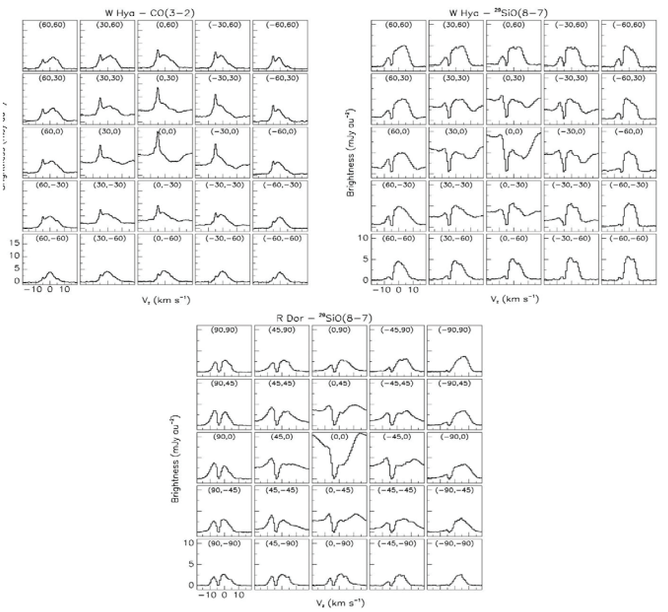


Fig. A1. Spectral maps of W Hya and R Dor molecular line emissions. The first set of 5×5 maps is for the ^{12}CO(3-2) emission of W Hya. The second and third sets are for the ^{29}SiO(8-7) emissions of W Hya and R Dor, respectively. Each spectrum is averaged over a squared aperture of 30×30 mas^2 for W Hya and of 45×45 mas^2 for R Dor centred at (*x, y*) position indicated in the inserts (in mas) of each panel. The beam (FWHM) is 52×40 mas^2 for W Hya and 41×35 mas^2 for R Dor. The R Dor spectra have been divided by a factor 3.1 to account for the different distances (104 pc for W Hya and 59 pc for R Dor).

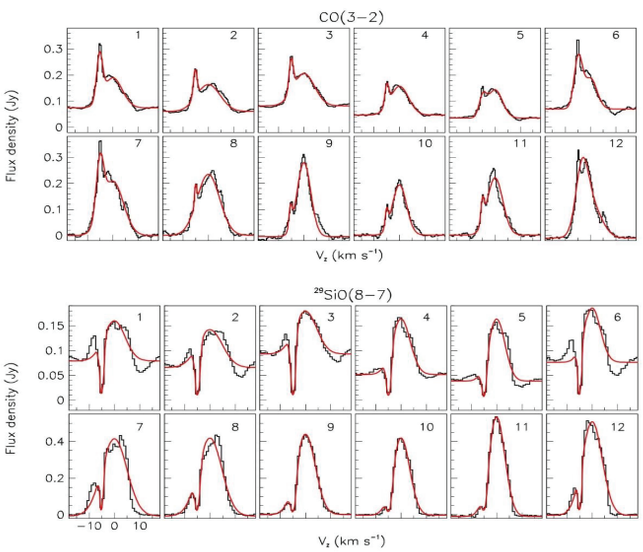


Fig. A2. Doppler velocity spectra observed in 12 sextants for the ^{12}CO(3-2) emission (upper pair of rows) and ^{29}SiO(8-7) emission (lower pair of rows) of W Hya. The sextants cover the projected radial ranges 0<*R*<75 mas (upper row of each pair) and 75<*R*<150 mas (lower row of each pair), respectively. In each row the sextants are numbered in increasing order starting from north counter-clockwise, from 1 to 6 for the inner sextants and from 7 to 12 for the outer sextants. Sextants 1 and 7 cover position angles $|\omega|<30^o$. The red lines show the results of fits of the form: $f=a+b\exp(-½V_z^2/c^2)+d\exp[-½(V_z-e)^2/f^2]$. The best fit values of the parameters are listed in Tables A2 and A3.

CRediT author statement

D.T. Hoai, P.T. Nhung: Conceptualisation, Methodology, Data curation, Formal analysis, Writing - Reviewing and Editing; P. Darriulat: Conceptualisation, Methodology, Writing original draft preparation, Writing - Reviewing and Editing; P.N. Diep: Funding acquisition, Writing - Reviewing and Editing; N.B. Ngoc, T.T. Thai, P. Tuan-Anh: Writing - Reviewing and Editing.

ACKNOWLEDGEMENTS

We thank Pr. Wouter Vlemmings and Pr. Aki Takigawa for useful discussions. We are grateful to Drs Jan Martin Winters and Pierre Leasaffre for very careful reading of the manuscript and for pertinent comments that helped improving the presentation of the results. This paper uses ALMA data ADS/JAO.ALMA#2015.1.01446.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The data are retrieved from the JVO/NAOJ portal. We are deeply indebted to the ALMA partnership, whose open access policy means invaluable support and encouragement for Vietnamese astrophysics. Financial support from the World Laboratory, the Odon Vallet Foundation and the Vietnam National Space Center is gratefully acknowledged. This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 103.99-2019.368.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] F.V. Leeuwen (2007), "Validation of the new Hipparcos reduction", *Astronomy Astrophysics*, **474(2)**, pp.653-664, DOI: 10.1051/0004-6361:20078357.

[2] G.R. Knapp, D. Pourbaix, I. Platais, et al. (2003), "Reprocessing the Hipparcos data for evolved stars III revised Hipparcos period-luminosity relationship for galactic long-period variable stars", *Astronomy & Astrophysics*, **403(3)**, pp.993-1002, DOI: 10.1051/0004-6361:20030429.

[3] N.N. Samus, E.V. Kazarovets, O.V. Durlevich, et al. (2017), "General catalogue of variable stars: Version GCVS 5.1", *Astronomy Reports*, **61(1)**, pp.80-88, DOI: 10.1134/S1063772917010085.

[4] T. Lebzelter, K.H. Hinkle, P.R. Wood, et al. (2005), "A study of bright Southern long period variables", *Astronomy & Astrophysics*, **431(2)**, pp.623-634, DOI: 10.1051/0004-6361:20041575.

[5] M. Maercker, F.L. Schöier, H. Olofsson, et al. (2008), "Circumstellar water vapour in M-type AGB stars: Radiative transfer models, abundances, and predictions for HIFI", *Astronomy & Astrophysics*, **479(3)**, pp.779-791, DOI: 10.1051/0004-6361:20078680.

[6] T. Khouri, A.D. Koter, L. Decin, et al. (2014a), "The wind of W Hydrae as seen by Herschel. I. The CO envelope", *Astronomy & Astrophysics*, **561**, DOI: 10.1051/0004-6361/201322578.

[7] T. Khouri, A.D. Koter, L. Decin, et al. (2014b), "The wind of W Hydrae as seen by Herschel. II. The molecular envelope of W Hydrae", *Astronomy & Astrophysics*, DOI: 10.1051/0004-6361/201424298.

[8] T. Danilovich, R. Lombaert, L. Decin, et al. (2017), "Water isotopologues in the circumstellar envelopes of M-type AGB stars", *Astronomy & Astrophysics*, DOI: 10.1051/0004-6361/201630055.

[9] B.R.M. Norris, P.G. Tuthill, M.J. Ireland, et al. (2012), "A close halo of large transparent grains around extreme red giant stars", *Nature*, **484(3793)**, pp.220-222, DOI: 10.1038/nature10935.

[10] K. Ohnaka, G. Weigelt, K.H. Hofmann (2016), "Clumpy dust clouds and extended atmosphere of the AGB star W Hydrae revealed with VLT/SPHERE-ZIMPOL and VLTI/AMBER", *Astronomy & Astrophysics*, **589**, DOI: 10.1051/0004-6361/201628229.

[11] K. Ohnaka, G. Weigelt, K.H. Hofmann (2017), "Clumpy dust clouds and extended atmosphere of the AGB star W Hydrae revealed with VLT/SPHERE-ZIMPOL and VLTI/AMBER. II. Time variations between pre-maximum and minimum light", *Astronomy & Astrophysics*, **597**, DOI: 10.1051/0004-6361/201629761.

[12] T. Khouri, W.H.T. Vlemmings, C. Paladini, et al. (2020), "Inner dusty envelope of the AGB stars W Hydrae, SW Virginis, and R Crateris using SPHERE/ZIMPOL", *Astronomy & Astrophysics*, **635**, DOI: 10.1051/0004-6361/201834618.

[13] M. Hadjara, P. Cruzalèbes, C. Nitschel, et al. (2019), "A CO-multilayer outer atmosphere for eight evolved stars revealed with VLTI/AMBER", *MNRAS*, **489(2)**, pp.2595-2614, DOI: 10.1093/mnras/stz2240.

[14] R.Z. Geisler, R. Köhler, F. Kemper, et al. (2015), "Spectro-Imaging of the asymmetric inner molecular and dust shell region of the Mira variable W Hya with MIDI/VLTI", *PASP*, **127(954)**, pp.732-741, DOI: 10.1086/682261.

[15] W.H.T. Vlemmings, T. Khouri, E. O'Gorman, et al. (2017), "The shock-heated atmosphere of an asymptotic giant branch star resolved by ALMA", *Nature Astronomy*, **1**, pp.848-853, DOI: 10.1038/s41550-017-0288-9.

[16] W.H.T. Vlemmings, T. Khouri, H. Olofsson (2019), "Resolving the extended stellar atmospheres of asymptotic giant branch stars at (sub-)millimetre wavelengths", *Astronomy & Astrophysics*, **626**, DOI: 10.1051/0004-6361/201935329.

[17] D.T. Hoai, P.T. Anh, P.T. Nhung, et al. (2021), "ALMA observations of W Hya: Impact of missing baselines", *Journal of Korean Astronomical Society*, **54(6)**, pp.171-182.

[18] T. Khouri, L.B.F.M. Waters, A.D. Koter, et al. (2015), "Dusty wind of W Hydrae. Multi-wavelength modelling of the present-day and recent mass loss", *Astronomy & Astrophysics*, **577**, DOI: 10.1051/0004-6361/201425092.

[19] A. Takigawa, T. Kamizuka, S. Tachibana, et al. (2017), "Dust formation and wind acceleration around the aluminum oxide-rich AGB star W Hydrae", *Science Advances*, **3(11)**, DOI: 10.1126/sciadv.aao2149.

[20] A. Takigawa, T.H. Kim, Y. Igami, et al. (2019), "Formation of transition Alumina dust around Asymptotic giant branch stars: Condensation experiments using induction thermal plasma systems", *The Astrophysical Journal Letters*, **878(1)**, DOI: 10.3847/2041-8213/ab1f80.

[21] W.H.T. Vlemmings, T. Khouri, D. Tafoya (2021), "Maser emission from the CO envelope of the asymptotic giant branch star W Hydrae", *Astronomy & Astrophysics*, **654**, DOI: 10.1051/0004-6361/202141656.

[22] T. Danilovich, A.M.S. Richards, A.I. Karakas, et al. (2019), "An ALMA view of CS and SiS around oxygen-rich AGB stars", *MNRAS*, **484(1)**, pp.494-509, DOI: 10.1093/mnras/stz002.

[23] P.T. Nhung, D.T. Hoai, P.T. Anh, et al. (2021), "Morpho-kinematics of the circumstellar envelope of the AGB star R Dor: A global view", *MNRAS*, **504(2)**, pp.2687-2706, DOI: 10.1093/mnras/stab954.

[24] G.A. Mamon, A.E. Glassgold, P.J. Huggins (1988), "The photodissociation of CO in circumstellar envelopes", *The Astrophysical Journal Letters*, **328**, pp.797-808, DOI: 10.1051/0004-6361/201731639.

[25] S. Etoka, L. Błaszkiewicz, M. Szymczak, et al. (2001), "Monitoring of long term behaviour of OH masers in semiregular variables: R Crt, W Hya and RT Vir", *Astronomy & Astrophysics*, **378(2)**, pp.522-538, DOI: 10.1051/0004-6361:20011184.

[26] H.S.P. Müller, F. Schlöder, J. Stutzki, et al. (2005), "The Cologne database for molecular spectroscopy, CDMS: A useful tool for astronomers and spectroscopists", *Journal of Molecular Structure*, **742(3)**, pp.215-227, DOI: 10.1016/j.molstruc.2005.01.027.

[27] M.V.D. Sande, L. Decin, R. Lombaert, et al. (2018), "Chemical content of the circumstellar envelope of the oxygen-rich AGB star R Doradus. Non-LTE abundance analysis of CO, SiO, and HCN", *Astronomy & Astrophysics*, **609**, DOI: 10.1051/0004-6361/201731298.

[28] T. Dumm, H. Schild (1998), "Stellar radii of M giants", *New Astronomy*, **3(3)**, pp.137-156, DOI: 10.1016/S1384-1076(98)00003-7.

[29] S. Höfner, B. Freytag (2019), "Exploring the origin of clumpy dust clouds around cool giants. A global 3D RHD model of a dust-forming M-type AGB star", *Astronomy & Astrophysics*, **623**, DOI: 10.1051/0004-6361/201834799.

[30] C.A. Gottlieb, L. Decin, A.M.S. Richards, et al. (2022), "Atomium: Alma tracing the origins of molecules in dust forming oxygen rich M-type stars: Motivation, sample, calibration, and initial results", *Astronomy & Astrophysics*, **660**, DOI: 10.1051/0004-6361/202140431.

DOI: 10.31276/VJSTE.64(4).28-31

DFT studies of the copper active site in AA13 polysaccharide monooxygenases

Chinh N. Le[1], Cuong X. Luu[1], Son Tung Ngo[2, 3], Van V. Vu[1, 4*]

[1]*NTT Hi-Tech Institute, Nguyen Tat Thanh University, 300A Nguyen Tat Thanh Street, Ward 13, District 4, Ho Chi Minh City, Vietnam*

[2]*Laboratory of Theoretical and Computational Biophysics, Ton Duc Thang University, 19 Nguyen Huu Tho Street, Tan Phong Ward, District 7, Ho Chi Minh City, Vietnam*

[3]*Faculty of Applied Sciences, Ton Duc Thang University, 19 Nguyen Huu Tho Street, Tan Phong Ward, District 7, Ho Chi Minh City, Vietnam*

[4]*Faculty of Biotechnology, Nguyen Tat Thanh University, 300A Nguyen Tat Thanh Street, Ward 13, District 4, Ho Chi Minh City, Vietnam*

Received 19 August 2021; revised 29 October 2021; accepted 15 November 2021

Abstract:

AA13 polysaccharide monooxygenases (AA13 PMOs) are novel enzymes that break down starch using a copper active site in a substrate binding groove on a solvent-exposed surface. The structure of the copper active site is influenced by the residues in the groove, while the crystal structure of Cu(II)-AA13 was damaged by photoreduction and lacked two exogenous ligands. We utilized density functional theory (DFT) calculations to obtain insights into the structure of Cu(II)-AA13 in the presence and absence of a key residue (G89) of the groove that interferes with the distal coordination site. Results show that the copper active site of AA13 PMOs can exhibit both 6-coordinate and a 5-coordinate structures depending on position of G89. The active site features are intermediate to those in AA9 and AA10 PMOs, which are the most abundant and well characterized PMO families. In addition, the superoxo species of AA13 has structural parameters halfway between those in AA9 and AA10 PMOs. The structural relationship between the active site and intermediates of AA13 with AA9 and AA10 PMOs is also consistent with their evolutionary relationship.

Keywords: **biofuel, DFT calculations, oxygen activation, polysaccharide monooxygenases.**

Classification number: **2.1**

1. Introduction

Oxygen activation at the copper active site of PMOs have gained significant attention in the past decade (Fig. 1) [1-9]. This active site is located on a solvent-exposed protein surface (Fig. 1A) in which the copper centre is coordinated by two absolute conserved histidine residues forming a histidine brace on the equatorial coordination plane (Figs. 1B and 1C). The distal site is occupied by a tyrosine residue in some PMO families or by a hydrophobic residue in other PMO families. It is responsible for selective hydroxylation of one of the two C-H bonds of the glycosidic linkage in polysaccharides (Fig. 1D). This hydroxylation is followed by an elimination step leading to the cleavage of the glycosidic linkage. PMOs can work in an endo fashion by creating new chain ends that are accessible by canonical glycoside hydrolases. PMOs can thus significantly boost the activity of glycoside hydrolases in degrading recalcitrant polysaccharides.

Several PMOs families have been reported thus far. Most of these families are active on the β(1→4) glycosidic linkages found in cellulose, hemicellulose, and chitin. AA13 is the only family of PMOs active on α(1→4) glycosidic linkages in starch. There are two major types of starch substrates termed as amylose and amylose pectin. Amylose only contains α(1→4) linkages and forms single and double helices that in turn form a microcrystalline structure. Amylopectin, on the other hand, also contains α(1→6) linkages that form branches. Amylopectin is thus more amorphous and is much more amenable to hydrolysis by the canonical glycoside hydrolases (amylases). In contrast, amylose is considered a resistant starch that is digested more slowly by amylases. Recent work has shown that AA13 PMOs are remarkably more active on amylose than on amylopectin [10]. Unlike other PMOs, AA13 PMOs use a shallow surface groove to bind amylose double helices. The active site of AA13 PMOs is modulated by residues in this groove including a flexible loop that could interfere with the distal coordination site (Fig. 1B). In addition, due to the effect of radiation damage during data collection, possibly both equatorial and distal ligands are lacking and the structure of AA13 PMOs in the Cu(II) oxidation

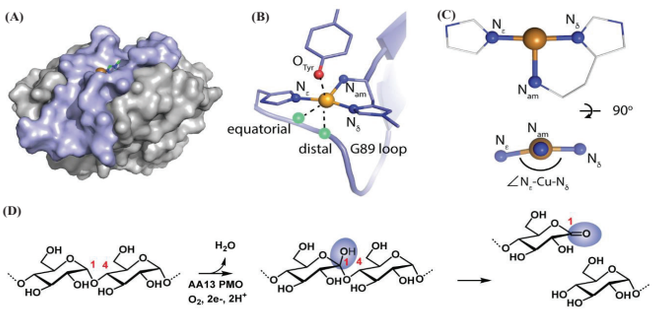


Fig. 1. Structure and reaction of AA13 PMOs. (A) Overall structure of AA13 PMOs with highlighted substrate binding groove and copper active site; **(B)** The copper active site; **(C)** The histidine brace motif; **(D)** Hydroxylation of the C1 position leading glycosidic linkage cleavage.

*Corresponding author: Email: anhvan.vu@gmail.com

state remains unclear. In this work, we utilized DFT calculations to investigate the active site structure and the influence of the flexible loop on the inner-coordination sphere of Cu(II)-AA13.

2. Methods

The input structures for AA9 and AA13 PMOs were obtained from the PDB files 5TKI [11] (Fig. 2A) and 4OPB [12] (Fig. 2B), respectively. Residues surrounding the copper centre, which are conserved throughout each family, were taken into account. Water molecules were added to the equatorial (O_{eq}) and distal (O_{dis}) coordination sites of AA13 PMO to create the input structure. Geometries of the species were optimized with the Gaussian 09 package [13] using the B3LYP functional at 6-31g(d) basis set [14] as previously described for PMOs [15, 16]. The solvation model was the dielectric polarized continuum model (D-PCM).

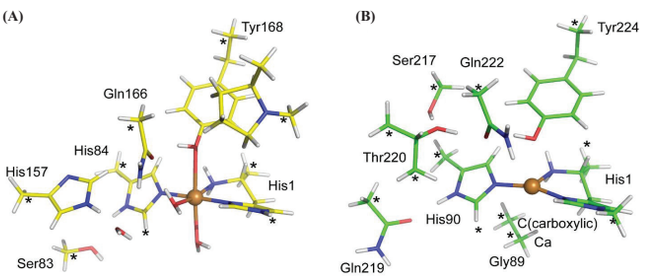


Fig. 2. (A) Starting structures for AA9 (5TKI) and (B) AA13 (PDB ID 4OPB). Asterisks denote frozen atoms during the DFT optimization procedure. The copper centres are shown as orange spheres.

3. Results and discussion

3.1. Effect of permittivity on the DFT calculation results

We initially carried out DFT calculation for the copper active site in NCU01050, a PMO belonging to the AA9 family. This PMO has a very high-resolution structure obtained with both X-ray and neutron diffraction (Fig. 3A). When DFT calculations were carried out in vacuum, the optimized structure deviated significantly from the crystal structure (Table 1 and Fig. 3B). The Cu-O_{Tyr} decreased from 2.683 to 2.360 Å, while the Cu-O_{dis} distance increased from 2.430 to 2.980 Å. The $\angle N_{\varepsilon}$-Cu-N_{δ} angle decreased from 178.0° to 164.5° resulting in the distortion of the equatorial plane (Fig. 3A). These three calculated metrics approach experimental values when the calculations were performed in a medium with higher permittivity (diethyl ether, ε=4.33) (Fig. 3C). When carried out in water (ε=80.1), these three calculated metrics matched experimental values. This result indicates that the medium polarization has strong effect on the DFT calculations of the copper active site in PMOs.

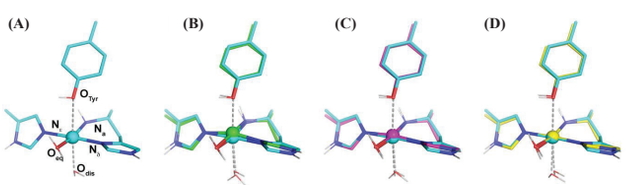


Fig. 3. The coordination structure of the copper active site in an AA9 PMO. (A) The crystal structure of AA9 PMO (5TKI). **(B-D)** DFT structures of AA9 PMO obtained in vacuum, diethyl ether, and water, respectively, overlaid on the crystal structure.

Table 1. The effect of the medium permittivity on DFT calculation results of the copper active site in Cu(II)-AA9 (all distances reported in Å).

NCU01050	Cu-N_{δ}	Cu-N_{ε}	Cu-N_{am}	Cu-O_{Tyr}	Cu-O_{eq}	Cu-O_{dis}	$\angle N_{\varepsilon}$-Cu-N_{δ} (°)
XRD, 5TKI	1.979	1.974	2.062	2.683	1.981	2.430	178.0
Vacuum, ε=1	1.932	1.923	2.028	2.360	2.074	2.980	164.5
Diethyl ether, ε=4.33	1.924	1.922	2.029	2.569	2.070	2.542	172.5
Water, ε=80.1	1.923	1.929	2.031	2.710	2.064	2.422	177.0

3.2. Optimised structures of Cu(II)-AA13

We added two water molecules to the equatorial and distal coordination sites of AA13 to create two input models for Cu(II)-AA13: Model 1 without G89 backbone (Fig. 4A) and Model 2 with G89 backbone (Fig. 4B). DFT optimization was then carried out for these models in vacuum, diethyl ether, and water.

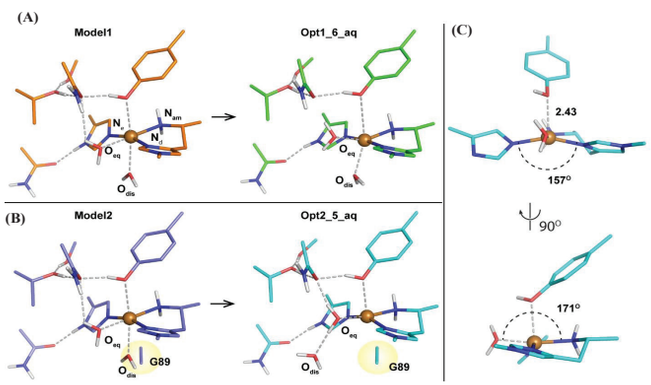


Fig. 4. DFT optimized geometries of the copper active site in AA13. (A) Without G89 backbone; **(B)** With G89 backbone; **(C)** Close-up view of the inner sphere of Opt2_5_aq. Oeq and Odis represent the aqueous ligands added to the equatorial and distal coordination sites in the input structures.

In the absence of the G89 backbone, DFT optimization resulted in a 5-coordinate copper species in vacuum (Opt1_5_vc), and a 6-coordinate copper species in both diethyl ether (Opt1_6_et) and water (Opt1_6_aq) (Table 2). These results indicate that the medium with higher permittivity better stabilizes the distal axial aqueous ligand in AA13 as also observed in the calculations with Cu(II)-AA9 presented above.

Table 2. Structural parameters obtained from DFT optimizations for Cu(II)-AA13 (all distances reported in Å).

	Cu-N_{δ}	Cu-N_{ε}	Cu-N_{am}	Cu-O_{Tyr}	Cu-O_{eq}	Cu-O_{dis}	$\angle N_{\varepsilon}$-Cu-N_{δ} (°)
Opt1_5_vc	1.959	1.972	2.072	2.358	2.068		155.9
Opt1_6_et	1.941	1.959	2.069	2.823	2.128	2.368	170.8
Opt1_6_aq	1.941	1.958	2.072	2.972	2.137	2.302	173.2
Opt2_5_vc	1.955	1.970	2.080	2.349	2.067		156.2
Opt2_5_et	1.953	1.974	2.081	2.404	2.048		156.1
Opt2_5_aq	1.951	1.979	2.077	2.425	2.029		157.1

In the presence of the G89 backbone, geometry optimization resulted in a 5-coordinate species in all three media investigated, where the distal aqueous ligand was forced out of the binding position. Consequently, the copper centre moved toward the proximal O_{Tyr} ligand resulting in a bent histidine brace (Figs. 4B and 4C). The structural parameters of all 5-coordinate species (Opt1_5_vc, Opt2_6_vc, Opt2_5_et, and Opt2_5_aq) are very similar to one another (Table 2).

3.3. Energy difference between 5-coordinate and 6-coordinate copper(II) species in AA13

In vacuum, the optimization of Model1 converged to a 5-coordinate species after a long "lagging" period of about 30 steps (step 20-50) (Fig. 5) in which the copper centre was 6-coordinate. In diethyl ether and water, our calculations stopped at 6-coordinate species and did not optimize any further. We thus took a 5-coordinate intermediate species at step 75 of the optimization in vacuum and further optimized it using the same functional and basis set in diethyl ether and water and obtained two optimized structures, namely, Opt1_5_et and Opt1_6_aq, respectively. This allowed us to compare single point energies of 5-coordinate and 6-coordinate species (Table 3). The energy difference between the 5-coordinate species and 6-coordinate species decreases as the dielectric constant of the medium increases. Nevertheless, the energy difference between the 5- and 6-coordinate species in water is significantly large (~-2.9 kcal/mol). This result indicates that 5-cooridnate species is likely preferred by Cu(II)-AA13 PMOs.

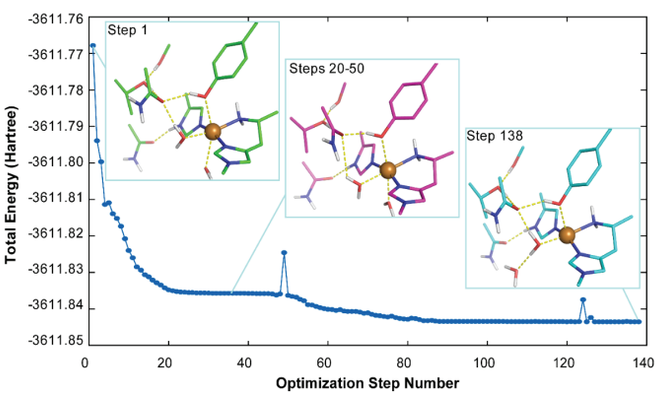


Fig. 5. DFT optimization process of Model 1 in vacuum.

Table 3. Energy difference between 5- and 6-coordinate species obtained with single point energy calculation.

Medium	ε	5-coordinate (Hartree)	6-coordinate (Hartree)	Energy difference (Hartree)	Energy difference (kcal/mol)
Vacuum*	1	-3612.65626130	-3612.64819820	-0.00806310	-5.06
Diethyl ether	~4.33	-3612.80146385	-3612.79488220	-0.00658156	-4.13
Water	~80.1	-3612.84860538	-3612.84397050	-0.00463488	-2.91

*6-coordinate species in vacuum was taken at the optimization step 50.

3.4. Optimized structure of a key intermediate of AA13 PMOs

One of the first key intermediates of the PMO reaction is the O_2-adduct, namely, a Cu(II)-superoxo intermediate. We carried out DFT calculations for Cu(II)-AA13-superoxo in the presence of G89 and absence of O_{dis}. Both singlet and triplet spin states were investigated. The optimized structures of these two states overlay well with one another (Fig. 6). The optimized structural parameters are presented in Table 4. These parameters are similar to one another although Cu-O_{Tyr} and ∠Nε-Cu-Nδ change slightly as observed for Cu(II)-AA9. Single point energy was also calculated for Cu(II)-AA13-superoxo (Table 5). The triplet state is remarkably more stable than the singlet state (-19 kcal/mol), which is consistent with the literature [15, 16].

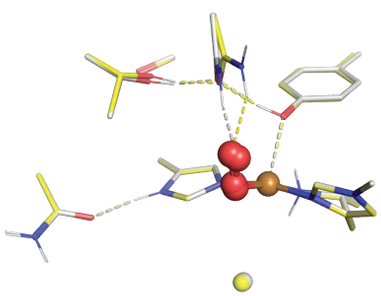


Fig. 6. Overlaid optimized structures of Cu(II)-AA13-superoxo intermediate in singlet (carbon atoms shown as white sticks) and triplet (carbon atoms shown as yellow sticks) spin states.

Table 4. Structural parameters obtained from DFT optimizations for Cu(II)-AA13-superoxo (all distances reported in Å).

	Cu-N_δ	Cu-N_ε	Cu-Na	Cu-O_{Tyr}	Cu-$O1_{oo}$	Cu-$O2_{oo}$	∠Nε-Cu-N_δ (°)
OptO2_S_vc	1.941	1.957	2.137	2.442	1.892	2.762	163.6
OptO2_S_et	1.93	1.955	2.126	2.487	1.888	2.760	166.8
OptO2_S_aq	1.927	1.959	2.114	2.509	1.886	2.768	167.5
OptO2_T_vc	1.967	1.977	2.134	2.378	1.981	2.913	158.7
OptO2_T_et	1.957	1.974	2.128	2.413	1.987	2.915	161.7
OptO2_T_aq	1.954	1.978	2.119	2.423	1.990	2.917	162.2

Table 5. Energy difference between triplet and singlet states of Cu(II)-AA13-superoxo obtained with single point energy calculation.

Medium	ε	Singlet (Hartree)	Triplet (Hartree)	Energy difference (Hartree)	Energy difference (kcal/mol)
Vacuum	1	-3690.1971719	-3690.2276186	-0.0304467	-19.11
Diethyl ether	~4.33	-3690.2462955	-3690.2766351	-0.0303396	-19.04
Water	~80.1	-3690.2652720	-3690.2957090	-0.0304370	-19.10

3.5. Implication in the evolution of PMOs

The structures of Cu(II)-AA13 and Cu(II)-AA13-superoxo correlate well with the phylogenetic relationship of AA13 with AA10 and AA9 PMOs, which are the two most abundant and well characterized PMO families. Phylogenetic analysis showed that the

AA13 clade is placed between the AA9 and AA10 clades (Fig. 7) [17]. Cu(II)-AA9 has an elongated octahedral copper centre, while Cu(II)-AA10 has a trigonal bipyramidal copper centre (Fig. 7A) [17]. The copper centre in Cu(II)-AA13 exhibit both 5- and 6-coordinate features, in which the 5-coordinate structure is preferred. This 5-coordinate structure is halfway between those in AA9 and AA10.

In addition, the O_2 moiety in the Cu(II)-AA9-O_2 intermediate binds in an end-on mode with the terminal O atom pointing down the distal space forming H-bonds with a water molecule and two outer sphere residues (PDB ID 5TKH) (Fig. 7B) [11]. On the other hand, the O_2 moiety of the Cu(II)-AA10-O_2 intermediate binds in a bidentate mode with the terminal O atom oriented toward the proximal space (PDB ID 5VG0) [18]. The O_2 moiety of Cu(II)-AA13-O_2 binds in an end-on mode, but, unlike in AA9, the terminal O atom is oriented toward the proximal space. This binding configuration is also halfway between the superoxo intermediate in AA9 and AA10 PMOs. These results show that as the PMO family evolved, the active site structure and key reaction intermediates also gradually changed.

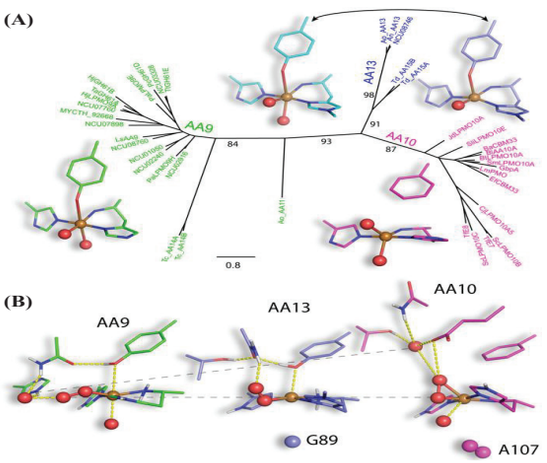


Fig. 7. The relationship between the copper active sites in PMO families. (A) The phylogenetic tree of structurally characterized PMOs. The active site core structures are shown for AA9, AA10, and AA13 PMOs; **(B)** Comparison of the structure of the dioxygen intermediate of AA13 (derived with DFT) with the crystal structures of those in AA9 (5TKH) and AA10 (5VG0).

4. Conclusions

The copper(II) active site of AA13 PMOs and their superoxo intermediates were optimized for the first time. The preferred structure of Cu(II)-AA13 is a distorted 5-coordinate species. This structure is halfway between those of the most abundant and well characterized AA9 and AA10 PMO families. Likewise, the structure of the superoxo intermediate is also halfway between that of an end-on intermediate in AA9 PMOs and a side-on intermediate in AA10 PMOs. The structural features of AA13 are consistent with their evolutional relationship with AA9 and AA10 PMOs.

CRediT author statements

Chinh N. Le, Cuong X. Luu, Son Tung Ngo, Van V. Vu: Data curation, Software, Visualisation, Investigation, Writing original draft preparation.

ACKNOWLEDGEMENT

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) grant # 103.01-2017.368 (to CXL).

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

RERERENCES

[1] G.V. Kolstad, B. Westereng, S.J. Horn, et. al (2010), "An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides", *Science*, **330(6001)**, pp.219-222, DOI: 10.1126/science.1192231.

[2] P.V. Harris, D. Welner, K.C. McFarland, et al. (2010), "Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: Structure and function of a large, enigmatic family", *Biochemistry*, **49(15)**, pp.3305-3316, DOI: 10.1021/bi100009p.

[3] W.T. Beeson, C.M. Phillips, J.H.D. Cate, et al. (2012), "Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases", *J. Am. Chem. Soc.*, **134(2)**, pp.890-892, DOI: 10.1021/ja210657t.

[4] C.M. Phillips, W.T. Beeson, J.H.D. Cate, et al. (2011), "Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by neurospora crassa", *ACS Chem. Biol.*, **6(12)**, pp.1399-1406, DOI: 10.1021/cb200351y.

[5] R.J. Quinlan, M.D.Sweeney, L.L. leggio, et al. (2011), "Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components", *PNAS*, **108(37)**, pp.15079-15084, DOI: 10.1073/pnas.1105776108.

[6] V.V. Vu, W.T. Beeson, E.A. Span, et al. (2014), "A family of starch-active polysacchride monooxygenases", *PNAS*, **111(38)**, pp.13822-13827, DOI: 10.1073/pnas.1408090111.

[7] R. Hemsworth, B. Henrissat, G.J. Davies, et al. (2014), "Discovery and characterization of a new family of lytic polysaccharide monooxygenases", *Nat. Chem. Biol.*, **10(2)**, pp.122-126, DOI: 10.1038/nchembio.1417.

[8] F. Sabbadin, G.R. Hemsworth, L. Ciano, et al. (2018), "An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion", *Nat. Comm.*, **9(1)**, DOI: 10.1038/s41467-018-03142-x.

[9] M. Couturier, S. Ladevèze, G. Sulzenbacher, et al. (2018), "Lytic xylan oxidases from wood-decay fungi unlock biomass degradation", *Nat. Chem. Biol.*, **14**, pp.306-310.

[10] V.V. Vu, J.A. Hangasky, T.C. Detomasi, et al. (2019), "Substrate selectivity in starch polysaccharide onooxygenases", *Journal of Biological Chemistry*, **294(32)**, pp.12157-12166, DOI: 10.1074/jbc.RA119.009509.

[11] W.B. O'Dell, P.K. Agarwal, F. Meilleur (2017), "Oxygen activation at the active site of a fungal lytic polysaccharide monooxygenase", *Angew. Chem. Intl. Ed.*, **56(3)**, pp.767-770, DOI: 10.1002/anie.201610502.

[12] L.L. Leggio, T.J. Simmons, J.C.N. Poulsen, et al. (2015), "Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase", *Nat. Commun.*, **6**, DOI: 10.1038/ncomms6961.

[13] M.J. Frisch, G.W. Trucks, H.B. Schlegel (2019), *Gaussian 16*, https://gaussian.com/citation/, accessed 12 March 2021.

[14] C.I. Bayly, P. Cieplak, W. Cornell, et al. (1993), "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model", *The Journal of Physical Chemistry*, **97(40)**, pp.10269-10280, DOI: 10.1021/j100142a004.

[15] S. Kim, J. Ståhlberg, M. Sandgren, et al. (2014), "Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism", *PNAS*, **111(1)**, pp.149-154, DOI: 10.1073/pnas.1316609111.

[16] C.H. Kjaergaard, M.F. Qayyum, S.D. Wong, et al. (2014), "Spectroscopic and computational insight into the activation of O_2 by the mononuclear Cu centre in polysaccharide monooxygenases", *PNAS*, **111(24)**, pp.8797-8802, DOI: 10.1073/pnas.1408115111.

[17] V.V. Vu, S.T. Ngo (2018), "Copper active site in polysaccharide monooxygenases", *Coord. Chem. Rev.*, **368**, pp.134-157, DOI: 10.1016/j.ccr.2018.04.005.

[18] J.P. Bacik, S. Mekasha, Z. Forsberg, et al. (2017), "Neutron and atomic resolution X-ray structures of a lytic polysaccharide monooxygenase reveal copper-mediated dioxygen binding and evidence for N-terminal deprotonation", *Biochemistry*, **56(20)**, pp.2529-2532, DOI: 10.1021/acs.biochem.7b00019.

DOI: 10.31276/VJSTE.64(4).32-35

Study on the dispersion of radon (^{222}Rn) in geological objects in Bat Xat district, Lao Cai province, Northern Vietnam

Van Dung Nguyen[1*], Thi Lan Anh Vu[1, 2], Thi Ha Thu Dang[3], Ha Phuong Vu[3], Thi Cuc Nguyen[1]

[1]*Faculty of Environment, Hanoi University of Mining and Geology, 18 Vien Street, Duc Thang Ward, Bac Tu Liem District, Hanoi, Vietnam*

[2]*Faculty of Environmental Science, University of Science, Vietnam National University - Hanoi, 334 Nguyen Trai Street, Thanh Xuân Trung Ward, Thanh Xuan District, Hanoi, Vietnam*

[3]*Center for Life Science Research, Faculty of Biology, University of Science, Vietnam National University - Hanoi, 334 Nguyen Trai Street, Thanh Xuân Trung Ward, Thanh Xuan District, Hanoi, Vietnam*

Received 4 January 2021; revised 20 March 2021; accepted 1 April 2021

Abstract:

The Bat Xat district in the Lao Cai province of Vietnam has enormous potential for resources of copper and rare earth minerals. In fact, exploration and exploitation activities for these minerals have brought about positive economic benefits to this area in recent years. However, this activity also brings about negative impacts on the environment and the surrounding population. The mineral deposits in this area contain a significant content of radioactive substances like uranium and thorium that radiate radon (^{222}Rn) and thoron (^{220}Rn). Therefore, the aim of this study is to investigate the distribution of radon as a function of depth of geological objects in the study areas. Results of this study showed that radon concentration has an exponential distribution with soil depth with R^2 ranging from 0.87 to 0.97. It was found that radon concentration in sand was higher than that in clay and soils mixed with gravels. It was also shown that the density and moisture content of the soil is inversely proportional to the diffusion length of radon in the soil.

Keywords: **dispersion, Lao Cai, RAD-7, radon concentration, radon in soil.**

Classification numbers: **2.1, 4.3**

1. Introduction

The radioactive decay of uranium and thorium chains and their progenies are widely distributed on Earth's surface and exist in all soil, rock, and mineral types at different concentrations. These radioisotopes affect living organisms. Once released into the environment, radioisotopes can enter the body through food, water, and air. In particular, radon exists in the form of a noble gas, so its dispensability into the environments could be on a wider scale than once thought. According to research from scientists around the world, radon is one of the leading carcinogens causing lung cancer [1-4]. Approximately 59% of the annual effective radiation dose that people receive is derived from radon. In the air, at least 70% of radon is released from the soil [5-8].

Radon migrates to the soil surface by different mechanisms and then emanates into the atmosphere [9]. The diffusion of radon in soil is a complex process. This process depends on many factors such as physical, chemical, geological, climate, and uranium content in the soil.

Measurement of the emission of radon from soil into the air has been used to assess the risk of radon exposure to the population. Further, the distribution of radon concentrations in soil has been successfully used in predicting Earth's kinetic phenomena such as earthquakes, karst caves, geological faults, and eruptions [10, 11].

The aim of this study is to investigate the dispersion of radon concentrations within geological objects into the environment in the Bat Xat district, Lao Cai province of Vietnam.

The Bat Xat district is delimited by the coordinates 22°51'÷22°55' N latitude and 103°68'÷103°74' E longitude. It is a complex terrain area as the central part contains low mountainous terrain surrounded by two sides of high mountains that are strongly dissected. The elevation change from 500-2,000 m creates many walls that are separated by river systems. The population distribution in the area is uneven. Most residents are concentrated in Muong Hum, including the Kinh and Day ethnic groups. Most of the Dao and H'mong converge into scattered villages in the high part of the terrain in the area and a few Dao people live in the low and fairly flat valley of Pieng Lao. In Sin Chai village, the Ha Nhi live right on the ore bodies of the radioactive Muong Hum rare earth mine and their houses resemble cottages with only one door; thus, homes are not ventilated and the accumulation of radon gas remains a threat to local residents [12].

2. Materials and methods

2.1. Geological characteristics of the study area

According to the survey results of the Geological Division for Radioactive and Rare Elements [12, 13], the geological features of the study area include the following formations: 1) Suoi Chieng (PP*sc1*), for which the main components are two-mica-granite shale, quartz-feldspar-mica schist, and thin layers of quartzite; 2) Sin Quyen (PP-MP*sq*), which contains metamorphic sedimentary formations, and the main components of biotite schist, mica schist, amphibolite, and quartzite; 3) Cha Pa Formation (NP*cp*),

*Corresponding author: Email: nguyenvandung@humg.edu.vn

which is distributed in a small area in the southeast corner of the study area. The main components of the NPsp are quartz-sericite schist and marble; 4) Ban Nguon formation (D*1bn*) for which the main lithological compositions are shale, sericite, carbonate shale, quartzite, calcareous sandstone, occasionally scapolite, cordierite, and biotite horn; 5) Ban Pap formation (D*1-2bp*) with a petrographic composition including a thin layer of limestone, clay limestone, dark gray limestone, fine to medium grain, rock with block structure or thin layers, silicate limestone, silicic shale, and in some places alternating layers of quartz - sericite - chlorite schist; and 6) the quaternary formation (Q) in which the main compositions are sand, gravel, and clay (Fig. 1) [12, 13].

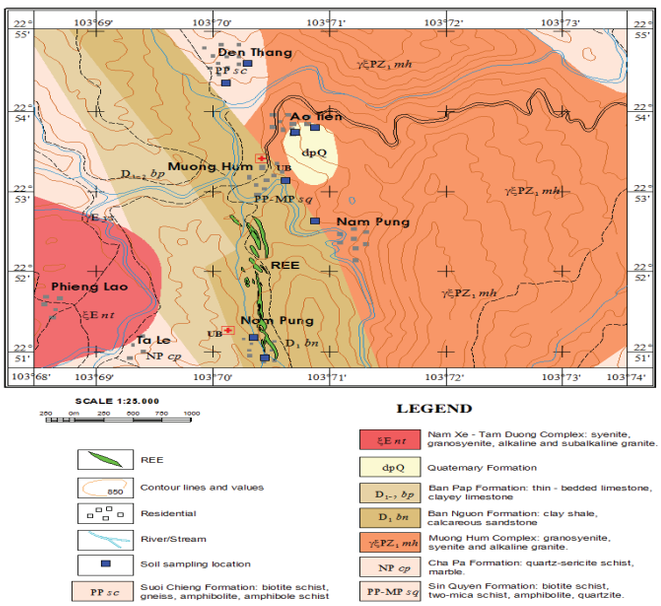


Fig. 1. Geological characteristics of the study area.

In the study area, there are copper and rare earth minerals that contain radioactive substances like uranium and thorium at high levels. The Muong Hum rare earth mine contains highly radioactive substances with uranium content ranging from 0.005 to 0.265% U_3O_8 and thorium content ranging from 0.106 to 0.188% ThO_2. The exploration and reserve assessment has been completed and will be exploited and processed in the near future [13]. The process of exploiting and processing radioactive minerals, excavating ores, exploiting land, collecting ore, crushing, and enriching strongly dispersing radioactive elements into the surrounding environment. There is a concentration of radioactive radon gas that affects the health of people living in rare earth mines and surrounding areas.

Therefore, in this study, an investigation of the diffusion of radon within the soil depth of different geological formations was conducted. The survey was carried out at 8 different locations in the Bat Xat district of Lao Cai province. These sites were chosen because they are typical geological formations in the region and where residents and farming land with food crops reside. The marks of 4 locations on the map in Fig. 1 (blue colour points) are described in Table 1.

Table 1. Survey locations of radon diffusion in soil.

No.	Geological object	Location	Symbol	X	Y
1	Suoi Chieng formation (PP_{sc})	Muong Hum	SC.1	103°70'76"	22°54'19"
2		Muong Hum	SC.2	103°70'39"	22°54'20"
3	Sin Quyen formation (PP-MPsq)	Nam Pung	SQ.1	103°71'27"	22°53'10"
4		Nam Pung	SQ.2	103°71'52"	22°52'65"
5	Ban Nguon formation (D_1bn)	Den Sang	BN.1	103°71'11"	22°51'10"
6		Den Sang	BN.2	103°71'01"	22°51'33"
7	Quaternary formation (Q)	Sang Ma Sao	DT.1	103°71'54"	22°53'70"
8		Sang Ma Sao	DT.2	103°71'34"	22°53'64'

2.2. Methods

Determination of radon distribution in soil: The radon concentration was determined at each location using punching equipment at depths ranging from 30 cm to 120 cm. Measurements were carried out in the field using the dedicated radon measuring system RAD-7 shown in Fig. 2. The device consists of a hemispherical chamber for air coated with a conductive layer. A silicon semiconductor detector was installed at the centre of the chamber. The measuring system consists of a gas pump that was connected to a gas receiver made with holes that allow gas to be taken in from underground. At the beginning of the measurement, the gas pump takes the gas containing radon at the position where the gas receiver is inserted into the RAD-7 chamber. Decaying radon produces the radionuclides ^{218}Po and ^{214}Po inside the chamber. Under an electric field, ^{218}Po and ^{214}Po are led to the detector in the centre of the hemisphere. Then, the RAD-7 converts the alpha energy imparted from these radionuclides into electrical signals [14-16]. The time taken for each measurement was 30 min.

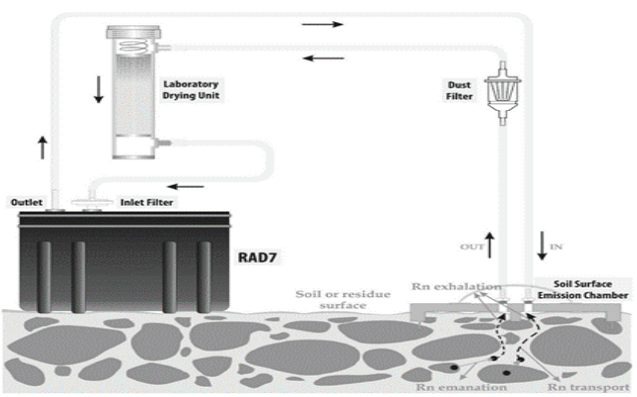


Fig. 2. A schematic layout of the installation used for radon distribution within soil depth.

In order to obtain the radon concentration as a function of soil depths, the diffusion of radon in soil was assumed to obey an exponential law as follows [5, 7]:

$$D\frac{d^2C(z)}{d^2z} - \lambda C(z) + \lambda C_\infty = 0 \quad (1)$$

where C(z)(Bq.m^{-3}) is the radon concentration in soil pores at depth z (cm); C_∞(Bq.m^{-3}) is the radon concentration at infinite soil depth; D is the radon diffusion coefficient in soil (m^2.s^{-1}); λ is the decay constant of radon (s^{-1}).

The solution of Eq.(1) is

$$C(z) = C_\infty + Be^{-\sqrt{\frac{\lambda}{D}}z} + Ae^{\sqrt{\frac{\lambda}{D}}z} \quad (2)$$

where A and B are constants; the radon concentration at an infinite depth of soil (C_∞) is constant for a single homogeneous soil layer and is determined by the following equation:

$$\frac{dC}{dz} = 0 \ (z = \infty) \quad (3)$$

Applying boundary conditions to Eq.(2), one obtains an expression for C(z) as follows:

$$C(z) = C_\infty + Be^{-\sqrt{\frac{\lambda}{D}}z} = C_\infty + Be^{-\frac{z}{L}} \quad (4)$$

where L (cm) is the diffusion length of radon in soil.

Radon concentration in soil at each site was determined at different depths (C(z)). The variation of radon concentration (C(z)) by depth (z) was fit according to the exponential law (Eq.4) and the coefficients C, B, and L were derived from this simulation in Table 3.

Determination of soil properties: Soil samples collected in the field were brought to the laboratory to determine some soil properties such as density, water content, and particle size.

The samples were weighed for mass determination, then dried for 6 h at 300°C. Soil moisture and soil density are determined by the following Eqs. (5) and (6):

$$w = \frac{m_w - m_d}{m_d}(\%) \quad (5)$$

$$\rho = \frac{m_d}{V}(g/cm^3) \quad (6)$$

in which w (%) is the moisture content of the sample; m_w(g) is the wet weight of the soil sample; m_d (g) is the dry weight of the soil sample; and V (cm^3) is the volume of the soil sample.

Soil particle size, moisture content, and soil density were determined at the laboratory of the Soil and Fertilizers Research Institute. Soil particle size directly influences radon emission and radon distribution in soil. If uranium is evenly distributed within the spaces in between soil grains, then the particle size is smaller and the radon emission is higher and vice versa [10]. However, when the soil particle size is greater than 0.1 mm, the emission is low and nearly equivalent to other particle sizes [10].

In this study, we use the following QA/QC procedures: sampling (sampling location, date), sample transportation (keeping sample storage and storage requirements), handling and laboratory sample preparation, analysis, control and calibration of analytical instruments, and sample control analysis.

3. Results and discussion

The profile of radon concentrations by soil depth is presented in Table 2 and depicted in Figs. 3 and 4. For each geological formation (object), the profile was experimentally determined at two points.

Table 2. Radon concentrations within soil depth.

Depth (cm)	Radon concentrations (Bq.m^{-3})							
	Suoi Chieng formation		*Sin Quyen formation*		*Ban Nguon formation*		*Quaternary formation*	
	SC.1	*SC.2*	*SQ.1*	*SQ.2*	*BN.1*	*BN.2*	*DT.1*	*DT.2*
20	259	313	498	641	399	433	367	531
40	726	802	1201	986	969	1151	1094	1301
60	1078	1139	1648	1256	1432	1636	1694	1895
80	1344	1372	1931	1468	1808	1964	2190	2353
100	1545	1533	2111	1634	2115	2185	2600	2706
120	1696	1644	2225	1764	2365	2334	2939	2979
140	1810	1721	2297	1865	2568	2436	3219	3189

With the results of Table 2, the diffusion expressions (Eq. 4) with C_∞, B, and L parameters for different geological formations are derived and presented in Table 3. In Table 3, the soil density and soil moisture in each formation are also presented.

Table 3. Diffusive equation constants of radon for different geological formations in Bat Xat along with soil density and moisture of the formation.

No.	Geological object	Diffusive equation	Diffusive length, L (cm)	Soil density, ρ (g.cm^{-3})	Soil moisture, w (%)
1	Suoi Chieng formation ($PPPP_{sc1}$)	$C(z)=2{,}161-2{,}521e^{-z/71}$	71	2.64	15.9
2		$C(z)=1{,}892-2{,}287e^{-z/54}$	54	2.66	15.1
3	Sin Quyen formation (PP-MPsq)	$C(z)=2{,}423-3{,}132e^{-z/44}$	44	2.54	17.8
4		$C(z)=2{,}134-2{,}032e^{-z/82}$	82	2.61	16.9
5	Ban Nguon formation (D_1bn)	$C(z)=3{,}454-3{,}754e^{-z/97}$	97	2.08	10.1
6		$C(z)=2{,}546-3{,}376e^{-z/51}$	51	2.11	12.5
7	Quaternary formation (Q)	$C(z)=4{,}554-5{,}065e^{-z/98}$	105	1.49	11.7
8		$C(z)=3{,}897-4{,}364e^{-z/77}$	77	1.54	10.1

As seen in Table 2, the diffusion length of radon in the soil at the survey locations ranged from 44 cm to 105 cm and averaged 72.4 cm. This result was in good agreement with those found for soils in Japan and Iran [7, 8].

The radon diffusion length of a maximum of 98 cm in soil of the study area implies that the amount of harmful radon emission into the air near the local population could mainly originate from depths less than 100 cm. Radon concentrations at the depth of 100 cm ranged from 591 Bq.m^{-3} to 3,219 Bq.m^{-3}. The International Atomic Energy Agency (IAEA) has recommended that the radon concentration in soil could affect human health with a high risk if it exceeds 30,000 Bq.m^{-3} [17]. The results of this study show that in the survey locations, radon emissions are within the permissible limit.

The concentration distribution by depth fits an exponential law as the correlation value R2 ranged from 0.87 to 0.97 (Figs. 3 and 4). The discrepancy between experimental data and theoretical calculation is due to the fact that almost all surveyed sites were influenced by human activities and the soil was no longer primitive. Soil samples from the Ban Nguon and quaternary formations had a higher concentration of radon compared to those in soil samples from the Suoi Chieng and Sin Quyen formations. The reason for this is that the soil from the Ban Nguon and quaternary formations was weathered from rocks containing a high content of uranium. Results of this study also

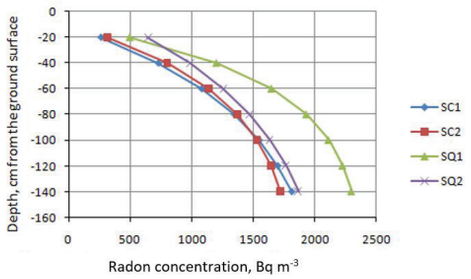


Fig. 3. The profile of radon concentration in Suoi Chieng and Sin Quyen formations.

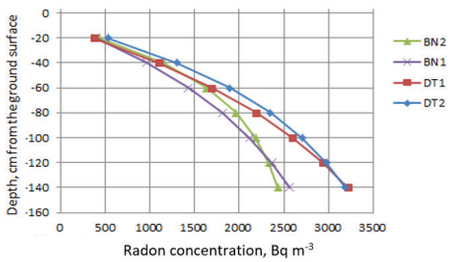


Fig. 4. The profile of radon concentration in Ban Nguon and quaternary formations.

show that radon in the soil from the study area has a small diffusion length. This means that the diffusion capacity of radon into the air is low such that the concentration of radon in the soil is high. In addition, the distribution of radon concentration in the soil is highly dependent on the distribution of uranium content in the soil particles [10].

The dependence of the diffusion length on soil density and soil moisture are shown in Table 2. Results in Table 2 show that the diffusion length of radon in soil is inversely proportional to the density and moisture content of the soil (R^2<0.5) due to high soil density and moisture hinders the diffusion of radon (Figs. 5A and 5B). These preliminary findings could be explained by the fact that the high density and moisture of the soil hinder the diffusion of radon molecules. However, the diffusion of radon in soil also depends on several other factors such as soil composition, temperature, weather, etc., which need to be further investigated in the future.

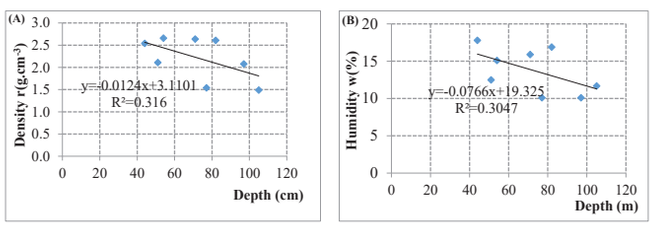


Fig. 5. The dependence of radon diffusion lengths on the (A) density and (B) moisture in the soil.

4. Conclusions

The distribution of radon concentration by depth in the surveyed soil samples obeyed the exponential law with values of correlation coefficient R2 ranging from 0.87 to 0.97. This shows that the survey sites have no risk of geological variation. The diffusion length of radon in clay is quite small compared to sandy soils and soils containing a lot of gravel. This is because clay samples have a relatively large soil density and water content.

The study results also show that the density and moisture content of the soil is inversely proportional to the diffusion distance of radon in the soil.

The radon concentration in the surveyed soil samples is within safe limits set by IAEA. The diffusion length of radon in the soil samples from the study area are smaller than that from other parts of the world. The radon diffusion length is greater for soil samples with low density and moisture content. Different soil samples have radon concentrations and diffusion lengths that depend on the origin of the formation, its composition, soil structure, temperature, humidity, and weather, etc.

CRediT author statements

Van Dung Nguyen, Thi Lan Anh Vu, Thi Ha Thu Dang, Ha Phuong Vu, Thi Cuc Nguyen: Data curation, Software, Visualisation, Investigation, Writing original draft preparation.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] P. Bossew (2010), "Radon: Exploring the log-normal mystery", *Journal of Environmental Radioactivity*, **101(10)**, pp.826-834, DOI: 10.1016/j.jenvrad.2010.05.005.

[2] F.P. Carvalho, M.C. Reis (2006), "Radon in Portuguese houses and workplaces", *International Conference Healthy Buildings*, II, pp.507-511.

[3] World Health Organization (2009), *Handbook on Indoor Radon: A Public Health Perspective*, World Health Organization Press, https://www.who.int/publications/i/item/9789241547673, accessed 21 August 2020.

[4] United States Environmental Protection Agency (2003), *EPA Assessment of Risks from Radon in Homes*, https://www.epa.gov/sites/default/files/2015-05/documents/402-r-03-003.pdf, accessed 21 August 2020.

[5] M.A. Domis, S. Xanthos, A. Clouvas, et al. (2009), "Experimental and theoretical study of radon distribution in soil", *Health Physics*, **97(4)**, pp.322-331, DOI: 10.1097/HP.0b013e3181adc157.

[6] G. Akerblom, P. Andersson, B. Clevensjo (1984), "Soil gas radon - A source for indoor radon daughters", *Radiation Protection Dosimetry*, **7**, pp.49-54, DOI: 10.1093/oxfordjournals.rpd.a082961.

[7] F. Jamadi, A.H. Ranjbar, H. Abbaslou (2015), "Radon gas diffusion coefficient in moisturized soil samples", *International Journal on Technical and Physical Problems of Engineering*, **7(1-4)**, pp.15-20.

[8] G. Prasad, T. Ishikawa, M. Hosoda, et al. (2012), "Estimation of radon diffusion coefficients in soil using an updated experimental system", *Scientific Instruments*, **83(9)**, pp.93-100, DOI: 10.1063/1.4752221.

[9] G. Imme, R. Catalano, G. Mangano, et al. (2012),"Radon exhalation measurements for environmental and geophysics study", *Radiation Physics and Chemistry*, **95**, pp.349-351, DOI: 10.1016/j.radphyschem.2013.02.033.

[10] Y. Ishimori, K. Lange, P. Martin (2013), *Measurement and Calculation of Radon Releases from NORM Residues*, Technical Reports Series No. 474, accessed 15 August 2020.

[11] C.A. Roba, D. Niță, C. Cosma, et al. (2012), "Correlations between radium and radon occurrence and hydrogeochemical features for various geothermal aquifers in Northwestern Romania", *Geothermics*, **42**, pp.32-46, DOI: 10.1016/j.geothermics.2011.12.001.

[12] B.T. Tran, D.H. Trinh, P. Nguyen (2007), "Investigating the current state of radioactive environment on the deposits of Dong Pao, Than Sin-Tam Duong (Lai Chau), Muong Hum (Lao Cai), Yen Phu (Yen Bai), Thanh Son (Phu Tho), An Diem, Ngoc Kinh - Suon Giua (Quang Nam)", *Journal of Geology*, **298**, pp.41-47.

[13] L.K. Phon (2001), *Research on Determining The Amount of Radiation, Their Pollution Level on The Environment, Public Health, Proposing Solutions to Prevent and Mitigate Losses: Damage in some Residential and Mining Areas in Lao Cai Province*, Provincial Project, Lao Cai Department of Science and Technology (in Vietnamese).

[14] Directorate for Standards, Metrology and Quality (2012), *TCVN 9416:2012: Investigation and Assessment of Environmental Geology by Radioactive Gas Method* (in Vietnamese).

[15] Durridge, *RAD7 Radon Detector*, https://durridge.com/products/rad7-radon-detector/, accessed 21 August 2020.

[16] J. Seo, M.M. Nirwono, S.J. Park, et al. (2018), "Standard measurement procedure for soil radon exhalation rate and its uncertainty", *Journal of Radiation Protection and Research*, **43(1)**, pp.29-38, DOI: 10.14407/jrpr.2018.43.1.29.

[17] United Nations (2000), *Sources to Effects Assessment for Rn in Homes and Workplaces*, II(E), 138pp.

DOI: 10.31276/VJSTE.64(4).36-44

An open-architecture integration solution for the research and development of smart robots

Tran Thang Do[1*], Van Phong Dinh[2], Quang Hoang Nguyen[2], Duc Hoang Chu[3]

[1]Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Nghia Do Ward, Cau Giay District, Hanoi, Vietnam
[2]Hanoi University of Science and Technology, 1 Dai Co Viet Street, Bach Khoa Ward, Hai Ba Trung District, Hanoi, Vietnam
[3]Ministry of Science and Technology, 113 Tran Duy Hung Street, Trung Hoa Ward, Cau Giay District, Hanoi, Vietnam

Received 20 August 2021; revised 2 November 2021; accepted 16 November 2021

Abstract:

Many key factors driving the Fourth Industrial Revolution (Industry 4.0) are to construct industrial robots to become increasingly intelligent, multi-functional, flexible, and efficient, as well as a complementing platform for collaboration with people. The latest generation of intelligent robots, known as Cobots (collaborative robots), can now cooperate and stand side-by-side with humans in a safe way and perform many tasks that require the same skills as humans. This paper firstly overviews the new generation of Cobots used for manufacturing and other applications to see forward into the future. Secondly, we briefly mention innovation technologies that should be integrated and extended deeper into smart robots to develop their applications in smart manufacturing, science and technology, high education, and so on. As a result, it is necessary to build new prototype robot systems that possess open hardware and software architectures for the development of both their smartness and dynamic skills. The main content of this paper will propose an open-architecture system for smart robots as well as like-robot machines using a CompactDAQ platform from National Instruments (NI). In the end, some information about the realized system implemented at the Institute of Mechanics of Vietnam Academy of Science and Technology will be shown for far joint discussion.

Keywords: **cobots, DAQ systems, integrated systems, motion control, smart manufacturing, smart robots.**

Classification number: **2.3**

1. Introduction

Industry 4.0 is an inevitable trend that is changing the way people live and work. This revolution brings opportunities and challenges to development in every country around the world. The manufacturing industry is a major part of the economy where people are increasingly resorting to mechanization and automation. Industry 4.0 is shaping smart factories based on smart manufacturing systems with highly flexible automation (autonomous systems) where machines, operators, and other resources can communicate together to create products in smarter and more organic ways. In a smart factory, a software-controlled system monitors physical processing and replicates the physical environment by creating a virtual replication. Based on the self-organization mechanism, the software-controlled system will make decentralized decisions for production processes. Next, Industry 4.0 will not only focus on individual factories, but also include the integration of many production facilities, service systems, and supply chains to achieve value-added networks.

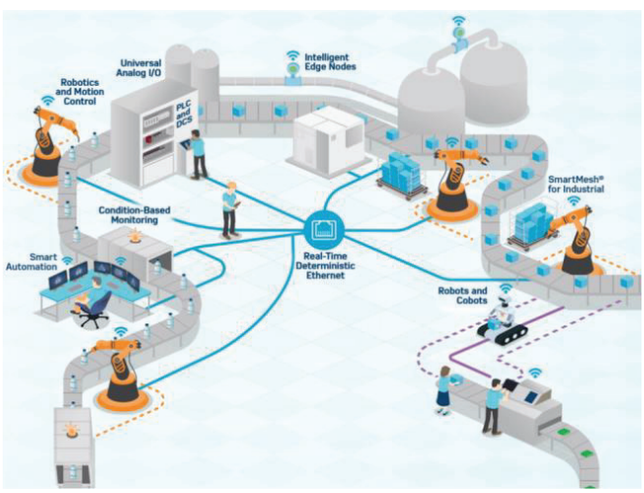


Fig. 1. The secure connected enterprise.

Industry 4.0 refers to a comprehensive cyberspace physical integration of corporate operations using the internet (industry) and everything integrated with operational technology (OT) and business systems (IT). This revolution will allow ecosystems that

[*]Corresponding author: Email: dtthang.mechatronics@gmail.com

link vendors, partners, distributors, and users in a robust value chain driven by simultaneous advances in cloud computing, communication infrastructure, and advanced technologies that help narrow the digital world (Fig. 1). However, technology at the edge of the network (where data is generated) is becoming increasingly important to ensure the integrity and value of information across the system. They need to be designed at a systemic scale with many technological challenges. This technology itself is not enough, but it is necessary to realize the vision of Industry 4.0 with innovative approaches, technical models, and deep expertise. The old convention "who does what" in the industrial ecosystem will be challenged and new partnerships will emerge. In general, the basic structure of a smart factory model includes intelligent machines and flexible automation systems; robots and smart actuators; application of artificial intelligence (AI) technology; all connected devices - Internet of things (IoT); and big data.

To this day, the automation strategy model of A. Granlund, et al. (2012) [1] is used for the process of integrating traditional industrial robots in the production system to increase production efficiency and replace operators who perform heavy, repetitive, and insecure work [2]. Traditional robots require an isolated robot cell to avoid direct human contact and are limited to a fixed position with a preprogramed task [3]. Compared with traditional robots, the new generation of intelligent robots, or Cobots, can collaborate with people to complete co-tasks [4], thus demonstrating interactions between people and robots. The technology also introduces a new concept of workplace design where a human-robot collaborative work cell can be described as a co-working and coordinated workspace encompassing Cobots performing tasks to suit the operator. Communication between Cobots and humans must be accomplished to fulfil certain requirements (friendliness) [3].

Recent years have witnessed an explosive expansion of the global robotics market, bringing in significant progress and maturity in robot technology. The leaders in robotics technology around the world include MIT Institute of Computer Science and Artificial Intelligence; Stanford AI Laboratory; the Robotics Institute of Carnegie Mellon University (CMU); Georgia Institute of Technology's Human-Computer Interaction Research Facility; Waseda University's Human-Robot Research Institute; The University of Tsukuba Smart Robotics; The Robotics Center and Mechatronics Center (DLR), and the German Aerospace Institute, etc. These institutes have been contributing to the research and development of intelligent robots, creating a treasure trove of human intelligence about robots. At the same time, there are many well-known robot companies including ABB (Switzerland), KUKA (Germany), Yaskawa Electrics and FANUC (Japan), Northrop Grumman, iRobot, Visual Surgical Robot (USA), ABP (UK), Saab Seaeye Underwater Robot, Reis Robot Group (Germany), Pesco (Canada) and Aldebaran (France), which have formed many pilot industries in their respective countries or regions, and have made remarkable contributions to the application and marketization of robots. In addition, several large businesses have held robotics competitions to accumulate solutions to industry-specific or robot-related problems with well-known titles include the Amazon Item Pick Challenge with Edge Logistics [5], the Unmanned Driving Urban Challenge [6], the DARPA Robot Challenge [7], the NASA Return Robot Challenge [8], etc. These robot events have wide attention around the world and encourage effectively promoted robot technology, but also provided a good opportunity to conceptualize and build a large playing field that helps enhance communication between robot development teams and applications. Many research institutes and robotics companies have made emerging breakthroughs in many related fields such as dual-arm collaborative robots, AGV intelligent logistics technology, unmanned driving technology, medical operation and rehabilitation robots, intelligent service robots, and dedicated robots.

Cobots, with the ability to work collaboratively with people, are typical representatives of the next generation of intelligent industrial robotics. Fig. 2 shows some famous collaborative robots. The intelligent collaborative delivery robot of the Rethink Company (USA) can either complete work independently on the work floor or cooperate with workers to control, pack, and process materials. This Cobot uses infrared sensors to detect approaching objects and cleverly avoid collisions [9]. In 2015, the ABB Company developed a dual-arm structured Yumi robot integrated with vision technology, instructional programming, precision servo control, and collision avoidance mechanism. These technologies help the robot complete smart assembly operations based on visual guides and force control and responds to demand from the consumer electronics industry for both flexible and traditional manufacturing [10].

Fig. 2. Famous human-robot collaborative robots.

Likewise, there are many other robots, including the UR series from Universal Robots; CR-35iA from FANUC (with a maximum payload of 35 kg); and the LBRiiwa and Automatic Pickup (AIP) technology from KUKA and Swisslog. In the field of medical surgery, intelligent robots can guarantee lower blood loss, higher resilience, more accuracy and agility; and possess great commercial potential. In 1985, the Puma 560, which was the first robot used in surgery, participated in a brain biopsy by providing brain locator services. In 1997, AESOP performed the

first laparoscopic surgery. The first version of the da Vinci surgical robot was born in 1999 (Fig. 3). Currently, there are nearly 4000 fourth generation da Vinci robots in the world. This robotic product has performed over 4 million surgeries and is known to be the most successful medical surgical robot in the world to date [11].

Fig. 3. The da Vinci surgical robot.

The Miro robot developed by the DLR Company (Germany) for cosmetic and orthopaedic surgery has a weight of only 10 kg [12]. The ViRob, a microrobot for automatic information collection from the medical company Microbot (Israel), appearing in 2015, can be controlled remotely via electromagnetic waves. This robot can provide cameras, drugs, or devices to narrow and twisted areas of the human body such as the blood vessels and the spinal canal, thus assisting doctors in minimally invasive surgeries. In 2003, a robot was designed to aid brain surgery, and its fifth generation has been introduced into clinical practice [13]. In 2013, the "Minimal Invasive Laparoscopic Surgical Robot System" was successfully built by the Robot Institute of Harbin Institute of Technology. This robot achieved emerging breakthroughs in technologies including master-slave control algorithm, machine design, and 3D peritonitis [14]. There are many other medical robots around the world, for example, the SpineAssist of Mazor Robotics (Israel) is used to regenerate and heal the spine. Carnegie Mellon University's master-slave CardioArm S-shaped robot consists of more than 100 joints and can be applied for minimally invasive heart disease surgeries. Rosa Brain and Rosa Spine robots of Medtech (France) are used as surgical support platforms. Most of these robots take the form of a single-arm structure or a combination of two arms or dual-arm robots.

Despite moment capabilities and advantages in certain circumstances, single-arm robotic systems have limited manipulations of dexterity, approach, and lifting capabilities. Continued advances in dual-arm robot technology offer a significant advantage over single-arm applications. With the dual-arm robot system, skilful and precise manoeuvring is sure, as the robot acts as an extension of the operator. A dual-arm robotic manipulation system provides the skill to stabilize an object with one arm while allowing the other arm to manipulate the object. This advance will help robots perform tasks that are particularly difficult and time-consuming for single-arm robot systems, such as removing bottle caps, unpacking, ruck sacking, and pulling items apart. In contrast, a dual-arm robot system can mimic operator movements to perform dangerous, complex tasks with human-like skills that neutralizes threats from a safe distance.

In the long term, an intelligent robot is a very modern, interdisciplinary, and multi-field integrated system capable of connecting IoT based on big data, cloud computing, and high-speed communication networks making them suitable for the current digital transformation trend. The process ranging from mastering the art of design, integration, manufacturing, and development of robot applications to construction in the robotics industry is essential to creating a smart manufacturing ecosystem and other values (Fig. 4).

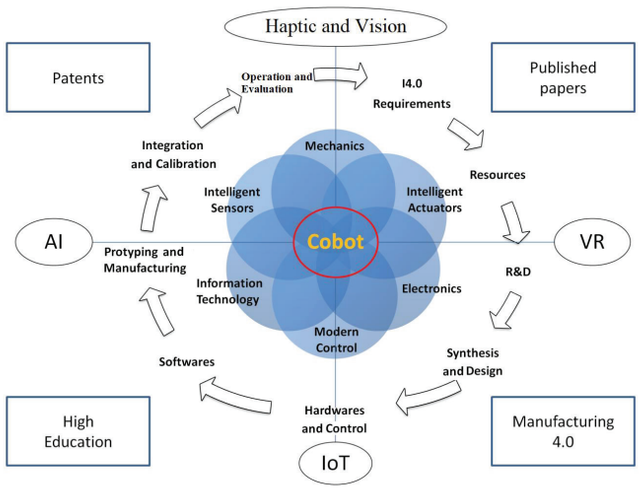


Fig. 4. The state-of-the-art of Cobot integration.

2. Innovative technology related to robotics

Compared to the most important and leading technologies involved, at present, intelligent/smart robots have non-complicated intelligence and relatively basic functions. They behave hardly in circumstances requiring human-robot cooperation under complex situations and have difficulty responding to user needs. It is needs to go over these technical drawbacks for the next steps. First, human-robot collaboration, and R&D on key technologies such as multi-modal awareness, environmental modeling, and decision-making optimization will be pushed to enhance the collaborative relationship between humans and robots. Second, robotic technology will be deeply integrated with the IoTs, big data, and cloud computing. Similarly, large sharing of data and computing resources will be fully utilized, which in turn will enhance the ability of intelligent robots to serve. Furthermore, AI technology such as emotional interaction, image recognition, deep learning, and brain-like intelligence must be deep developed and used to build robots capable of intelligent decision-making at a high level, ensuring safety and reliability. In addition to the above technological drawback, there are several new challenges emerging in the way to innovate robot technology: How to effectively utilize the industrial chain to conduct R&D cooperation; How to improve operation system for robots; and how to find innovative solutions to integrate multiple fields smoothly need to be answered. Research and innovation on leading technologies are critical to the formation

and development of a nation's future emerging industries along with data and cloud-based network decision-making mechanisms and big data.

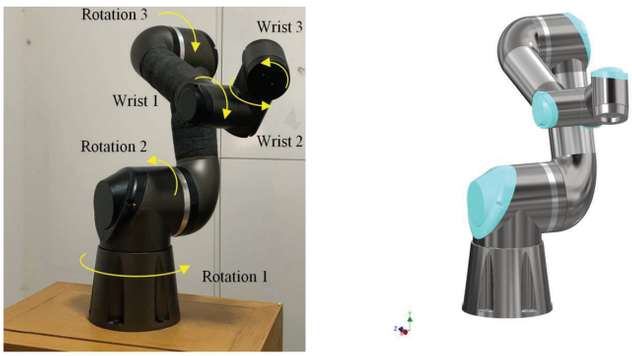


Fig. 5. The SM6 Cobot: (A) real robot and (B) 3D CAD model.

Figure 5 shows two pictures of the SM6 Cobot developed by the Mechatronics Department, Institute of Mechanics, Vietnam Academy of Science and Technology. In the figure, from left to right, is the image of a real robot, then the 3D simulation model, respectively. This robot reaches some features in Table 1 below.

Table 1. Specifications of the SM6 collaborative robot system.

Type			Articulated type
Degree of freedom (DOF = number of axes)			06
Max. payload (kg)			approximate to 2.5
Positional repeatability (mm)			approximate to ±0.1
Motion range (°)	Arm rotation 1 (°)		-360 to +360
	Arm rotation 2 (°)		-360 to +360
	Arm rotation 3 (°)		-150 to +150
	Wrist 1 (°)		-360 to +360
	Wrist 2 (°)		-360 to +360
	Wrist 3 (°)		-720 to +720
Controller	Number of controlled axes		6 to 8
	Drive system		full digital servo system
	Type of motion control	Manual mode	coordinated movement, individual movement, motion tracking, [Interpolation mode] joint, base, tool
		Auto mode	coordinated movement, individual movement, motion tracking, [Interpolation mode] joint, linear interpolated motion
	Force Control		yes
	Programming		direct teaching method through tablet
			hand by hand
			haptic device
			vision
	Safety mode		stop or speed reduction to specific situations
Connection			ethernet
	Gripper		two figures and other tools
Mass (kg)			approximately 20

In the next phase, this robot system will be continuously developed by extended and deep integration processes with innovative technologies aimed to be more intelligent.

2.1. Technology for designing and manufacturing robots

Robot technology is a complex and advanced technology involving multiple disciplines and interdisciplinary industries including mechanics - electronics, automatic control, sensor technology, computer - information technology, new materials, biotechnology, and artificial intelligence (increasingly to become more integrated and expanded). Robot manufacturing technology is “an advanced technology” of the human race, closely related to a multi-stage process from theoretical research and technology development to technical applications using interdisciplinary integrated thinking. The mechanism between "software" and "hardware" of the system includes many physical components, technologies, and knowledge from many diverse fields. Fig. 4 shows the backbone process for creating general robots like the SM6 robot. Now, it is time to refresh this process for intelligent robots for a new robot generation integrated with novel technologies.

2.2. The technology of cooperation between humans and robots

Human-robot collaboration plays an unlack role in the next generation of intelligent robotics. ISO issued the latest robot production standards in 2016 [15]. To realize a coexistence of robots and men, human collaboration robots need to be safe, comfortable, adaptable, and easy to program. Safety means protecting people from dangerous injuries that could result from robots during an interaction. Comfortability means that a robot's actions are subject to human cognitive habits and thus humans can predict a robot's intentions. Adaptation is when robots can be able to understand human needs and precisely adapt to human movement and various tasks. Easy programming means that people can easily program, learn methods of operation, and control robots. Therefore, cooperative robots with the four following activities are required:

Safe stop: This operation forces the robot to stop before humans enter the robot's workspace. The aim here is to keep people secure and prevent exposure to risks or dangers in the workspace.

Manual human guide: This will give humans access to control the robot by leading the robot’s moves and operations.

Speed monitoring and separation: This aims to reduce speed when people enter the work area (a safety scanner can be used [16]). The distance between the robot and person is always measured such that the robot either stops, slows down, or backs away from a person to maintain safety [17]. The speed of the robot will decrease if a person moves close to a robot, and, if they move too close, the robot will automatically stop for safety reasons.

Force limitation and force control: A focus on avoiding injury and pain that robots can cause to humans. To avoid injury and pain, it is necessary to limit the robot's speed as well as force. Therefore, the ability to injure humans will decrease [16].

The main technologies to be achieved include the design of a hard-soft-flexible coupling mechanisms [18-20], intelligent drive technology, safe decision-making mechanisms geared towards human-robot cooperation [21, 22], 3D modelling of environments [23], interactive force sensing and control, a distributed collaboration of multi-agent systems, vision-based image processing [24-26], furthermore emotional recognition and interaction mechanisms toward the friendly cooperation between humans and robots.

2.3. Technology for multi-robot systems

A multi-robot system refers to an identified number of independent robots that take moving and cooperate as a self-organizing group. Such collaborative action can help the team realize complex functions and demonstrate a clear trend towards the "in projects" or "goals" collective [27]. Studies of multi-robot systems will focus on the aspects such as increasing the convergence speed of coordinated control and dealing with finite-time control [28], transforming the topology of system changes over time, and organizing the multi-agent network in a more proper way [29], designing a program that estimates according to the global non-linear collaboration state, and realizing differential co-operative control groups of robots based on a heuristic algorithm [30].

Compared with the traditional single-robot system, there is no global control goal however there is distributed control issue in a multi-robot system. Multi-robot collaboration improves efficiency in the execution of tasks and its lifetime enhances the robustness of the system can complete distributed tasks that a single-robot system cannot do. Furthermore, the multi-robot system is easily expanded and updated. Each robot has relatively basic functions and a limited ability to collect, process, and communicate information. Through transmission and interaction between robots, the whole system will be highly effective in collaboration and higher intelligence performance, thus achieving many difficult and complex jobs. Impurity requires high precision and is beyond the capabilities of a single-robot system [31]. Multi-robot systems demonstrate great power in multi-sensor collaborative information processing, multi-robot collaboration, drone fleets, multi-operation control, and other fields [32-34].

2.4. The technology of control and decision-making is like the brain

With robots' deeper and broader involvement in industrial production and the social life next generation, thus it will be required to increase their performance. Robots usually serve in complex and interchangeable environments with unpredicted uncertainties, and missions are of a high hermaphroditic and nonlinear nature, performance requirements go beyond the validity and plausibility of common control algorithms. Capable of rich sensor data collection capabilities, the robot can perform sorting, synthesizing, and extracting data efficiently and reliably in the future (Fig. 6).

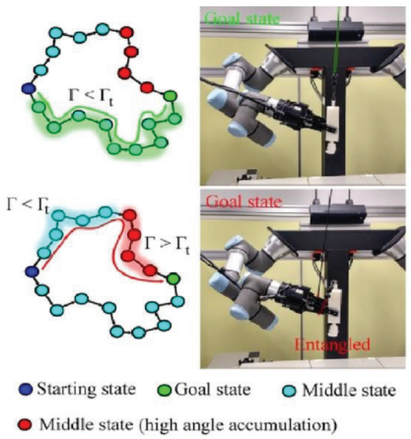


Fig. 6. Training a robot in object grasping based on deep learning.

Robots possess strong learning abilities including supervised or unsupervised learning. Supervised learning will train logical model parameters based on artificially labeled sample sets from which manipulation control decisions are made. In contrast, unsupervised learning will train optimal control laws from the unlabelled sample sets [35]. Supervised learning meanly focuses on imitation and performance [36, 37]. A. Paraschos, et al. (2013) [38] proposed the primitive probability of motion (ProMP), which describes primitive motion using probabilistic modelling that allows the robot to adapt to changing tasks. In several studies [39, 40], the sub-primitive motion constructs a compact expression of robot control laws used to perform actions such as collision, impact, and capture. Robots can mimic human actions and adapt to different circumstances by adjusting their primitive parameters of motion. Unsupervised learning is mainly focused on building complex neural networks and deep learning. Deep learning was proposed by Hinton, which simulates brain function learning by applying neural networks. S. Levine, et al. (2016) [41], C. Finn, et al. (2017) [42] deployed unsupervised learning by establishing a larger neural network. The robot can plan strategies while being trained as well as build a mapping relationship between robot movement, sensory information, and mission, thus helping the robot complete the task without human intervention such as closing and opening doors and holding items (Fig. 6). Based on the multi-layered expression mechanism in the brain, such an unattended feature learns and recognizes deeply abstract expressions from the original game and avoids artificial interference in this process to export geometric objects.

Moreover, deep learning theory, to some extent, helps solves the problem of local convergence and is more adaptive in previous artificial neural networks.

2.5. Virtual reality technology for robotics

The new generation of intelligent robots will feature high connectivity, virtual-real linkage, intelligent software, and human-robot collaboration. More specifically, the robot can collect a lot of rich data from diverse sensors and upload them to the cloud to process and share information. Virtual and real devices

are deeply integrated forming a closed process of collecting, processing, analysing, responding to data, as well as executing and implementing "real-virtual-real" transformations. The smart algorithm to analyze big data based on advanced software applications enables the new generation of intelligent robots to thrive in a softwareization, content-depended, centralized platform API. Robots can be able to realize human-robot interaction using images and videos, even sense man's psychological activities, and exchange emotions applicating deep learning. The smart robot number is gradually expanding and becoming more systematic, essentially building an all-round chain and welcoming impressed technological innovation.

2.6. The technology of a robot's network based on cloud computing and big data

During the 2010 International Conference on Humanoids, for the first time, J.J. Kuffner, et al. (2011) [43] presented the concept of a "cloud robot". Cloud robots are an integration of cloud computing and robotics. Cloud computing is an internet-based computer mode. Software, resources, and information can be shared and made available to endpoints according to business needs. Like a terminal network, the robot does not need to store all the data or possess powerful computation but must link up with relevant servers to obtain the information it needs. Cloud robots can not only upload complex computational issues to the cloud but can also get a lot of data and exchange both information and skills. It is more powerful than storage, computation, and learning, and makes it more convenient to re-share resources between robots. Additionally, the robot will bear a smaller burden under similar circumstances, thus saving developers time by eliminating repetitive work [44]. Many companies like Google, Microsoft, and Baidu are involved in a computing cloud and big data technology research. Various cloud robot service architectures have been presented including a cloud service platform for networked robots (ex: ROS platform [45] and RoboEarth platform), a cloud service platform for sensor network (e.g., Sensor Cloud [46] and X-sensor), and a cloud service platform for the RSNP Model (e.g., Jeeves platform) [47]. The development of big data and cloud computing will drive development in the realm of robotics such as object capture, SLAM, and emotional perception. Notably, AS ORO Laboratory from Singapore has built a cloud-computing system that allows robots to build 3D maps of the environment at a much faster rate than the computer connected to the robot [48]. B. Kehoe, et al. (2013) [49] from the University of California, Berkeley, completed the task of capturing via a cloud-based 3D robot using a PR2 robot from Willow Garage Company and a target recognition tool from Google. Microsoft has developed an articulate Cortana personal assistant for conferences, which is dependent on cloud technology and can create a comfortable conversation with everyone. It can examine if it is the same talker or not during the conversation. Google's DeepMind developed and improved the AI architecture AlphaGo, which overcome many Weiqi experts around the world [50]. Cloud robot technology, as well as its unique advantages integrated with big data and cloud computing technology, are going to spark a huge transformation in the field of smart robots.

With ever-deepening industrial reform, changing social needs, and technological advancement, robots in the world are undergoing never seen growth. Countries around the world are competing in pushing robot innovation strategies. Third backbone technologies represented by big data, cloud computing, mobility, and social interaction drive the global robot industry growth towards intelligence, innovation, and digitalization. In the future, along with the development of smart hardware and AI technology, intelligent robots will surely make great progress and be broadly used as collaborative robots in unmanned vehicles, medical care, logistics, education, and entertainment. The diversity of user needs and the increasing emergence of never seen technologies push robots to evolve in highly intelligent, highly adaptable, and network-based ways.

3. A proposed open-architecture employed NI technologies for robotic development (the SM6 Cobot)

In essence, the general architecture of a robotic system needs to meet the requirements of implementation from basic to advanced skills and learning abilities. In traditional manufacturing, the main skill of robots is to perform motion tasks considering interactions of forces with the environment, while intelligent robots will develop the self-learning ability to collaborate directly with humans. To adapt to this requirement, there have been various solutions developed based on so many different software and hardware platforms throughout two phases including expanded-dimension integration (open sources) and deep-dimension integration (packed sources).

In this paper, we propose a hardware and software architecture employing the NI CompactDAQ platform aimed to integrate and expand innovative technologies into robotics for the development of smarter robots. The latest CompactDAQ system can perform both control and acquisition functions at the same time, providing data measurement that meets inspection requirements without limitations in terms of distance and environmental conditions. This device typically pairs with I/O modules that provide communication to the sensor using software optimized for both industrial and R&D applications. There will be solutions made of data acquisition modules that can synchronize multiple measurements over a network bringing the data digitization system closer to the sensor thereby reducing noise and simplifying wiring connections.

A CompactDAQ system can be easily connected to your PC via USB or Ethernet and then integrate one or more I/O modules for direct communication with sensors. Users can choose from a built-in controller that works with Windows or a real-time operating system like Linux to run independently. The CompactDAQ controller is a tightly integrated, reliable, high-performance industry standard. It is an ideal solution for collecting and analysing data in built-in or portable memory. Obviously, this unit also allows for many standard connectivity

and expansion options including USB, Ethernet, CAN/LIN, and RS232. In addition, the integrated solution can be expanded with high-quality input and output modules (C Series) to provide signal conditioning and analogue-to-digital conversion for CompactDAQ systems (NI currently offers up to 60 C Series I/O modules to interface with most sensors on the market, in a quick and easy way for a size-optimized, cost-optimized custom hardware setup with good performance). Note that these modules can be hot-swappable and plugged directly into the main machine, enabling customers to easily build a system to suit specific test requirements. The aforementioned allows processes from data collection and analysis to measurement system programming to be fully automated and seamless, so that tests can be flexibly tailored to specific needs.

CompactDAQ hardware is built to work well with DAQExpress™ software, FlexLogger™, and visual language programming environments such as LabVIEW to meet extensive development requirements and even deep integration. Not only that, but the CompactDAQs also use the powerful NI-DAQmx driver, making it possible to write programs in many other languages.

For motion control purposes, NI strongly supports testing, measurement, and control platform of mechanisms, machines, robots, and so on. In this application, the system will further integrate C Series and interface modules for CompactRIO and EthernetRIO, or PCI and PXI plug-in motion controllers. Combining this hardware system with the use of LabVIEW software, a multi-purpose controller, and a complete range of motors and drives, customers can build a variety of advanced multi-utility motion applications quickly and with lower costs.

In practice, motion systems are often built upon application software, motion control, actuators, and other mechanical components. The lower part of the motion control architecture is implemented at a hardware level close to what is required by the application, for example, initial setup, supervisory control, trajectory generation, trajectory interpolation, and control loops (position, velocity, torque).

NI currently offers two different motion architectures programmed with different software (Fig. 7). One architecture uses the NI plug-in motion controller in conjunction with motion-enabled NI-Motion controller programming, while the other uses a real-time controller and FPGA for programming with the LabVIEW SoftMotion Module (Fig. 8).

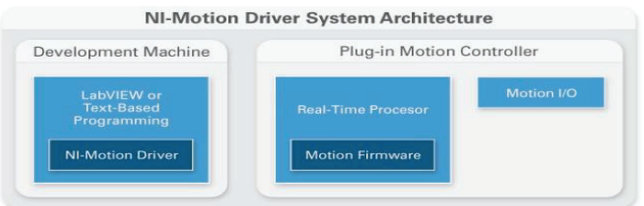


Fig. 7. Plug-in motion controller architecture.

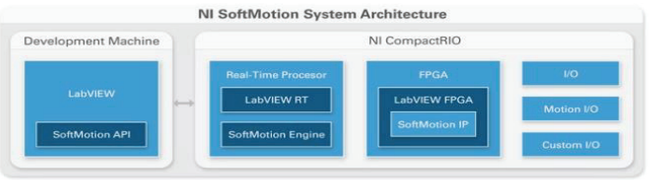


Fig. 8. Motion system architectures.

SoftMotion is a solution for software-based motion control, where various motion control components are modularized such as supervisory control, trajectory generation, and control loops (Fig. 8). Each component is then routed to a central motion system (developing machine, real-time processor, or FPGA) tailored to a specific application purpose. Customers can fully use SoftMotion for advanced application development from scratch, or customize it at the application code level, real-time SoftMotion Engine, or IP FPGA level. When using SoftMotion in conjunction with LabVIEW, it helps to quickly configure motion setup, simulate, link drivers to hardware, and test and adjust motion setup before application development. NI cRIO platforms are used to build an open architecture for integrating intelligent robots in several types of constructs: single-arm Cobots (Fig. 9) and dual-arm Cobots with a proposed systematic diagram as shown in Fig. 10.

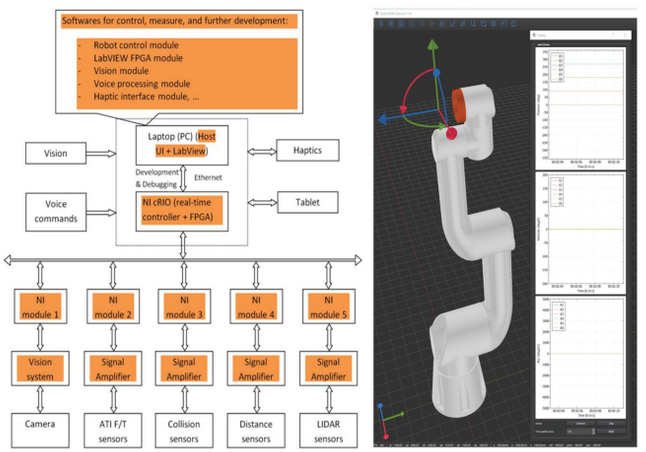


Fig. 9. A systematic diagram for open robotic architectures.

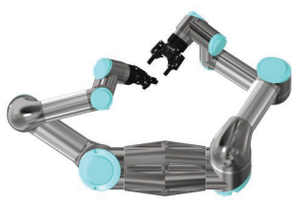


Fig. 10. 3D Simulation of the dual-arm SM6 Cobot.

Alternatively, the NI-Motion driver can be inherited to program the motion controllers on PCI and PXI (e.g., NI 7330, NI 7340, NI 7350, and NI 7390). This driver includes LabVIEW VIs (with examples) that will help users quickly create motion control

applications using LabVIEW for Windows plug-in controllers or real-time LabVIEW. In addition, NI offers a wide range of easy-to-connect, reliable driver solutions to its embedded controllers that perform well in a variety of power ranges and are compatible with many other third-party devices.

Figure 11 shows a hardware and software system based on the solution of this paper that has been implemented for directly developing SM6 Cobot at the Automation and Mechatronics Laboratory of the Institute of Mechanics, Vietnam Academy of Science and Technology, as mentioned above. Also, this system is expected to help develop intelligent systems like robots.

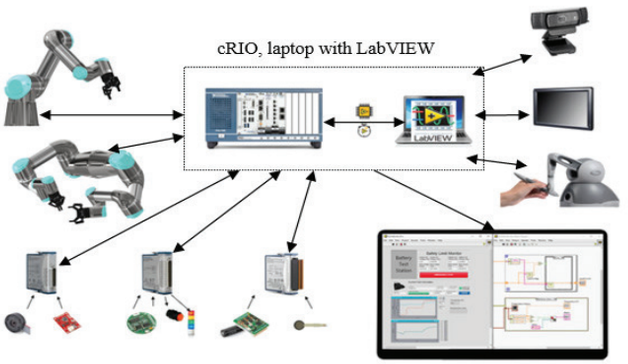


Fig. 11. A hardware and software system using NI devices for robot development.

4. Conclusions

Intelligent robots are predicted to liberate the manufacturing industry not only within the Fourth Industrial Revolution, which will root from automotive, electric-electronics, metal-mechanical, biomedical, and other industries. Developing a novel integrating solution will help capture this future opportunity for Vietnam by the intensive focus on automotive, electric-electronics, medical, other advanced technologies, and will step-by-step create a smart manufacturing ecosystem with the contribution of intelligent robots. First of all, it will develop in-depth knowledge, design, and prototypes of intelligent robots to help raise society's awareness of robots, artificial intelligence, Internet of things, virtual reality technologies, and other related fields in Vietnam. After that, based on the previous results, it will prepare the technology database to transfer, consult, support, and take joint development with enterprises aimed to commercialize products and services successfully. These goals will be achieved by concurrent streams of supporting activities like scientific publications and collaborative workshops, engagement with the manufacturing industry, especially in smart manufacturing, and continuous training generations of engineers at various levels.

The SM6 robot introduced in this paper is the second version and was just completed in 2021 (the first one was completed in 2018). The latest version is expected to integrate with more expanded and deep innovative technologies including many other techniques for both dynamics skill improvement and smartness. To realize this expectation, the authors propose a new hardware and software architecture using the cRIO platform from NI opened for integration development of new robots.

Interdisciplinary integrating solutions help develop in-depth research in related fields like robotics, mechatronics systems, control, AI (machine learning, deep learning), IoT, VR, cloud computing, big data, simulation, and embedded systems. Furthermore, it contributes to close links between domestic and foreign research groups. All these aspects help build and develop a deep and wide cooperative relationship between research institutes, universities, and enterprises in the innovation progressing trend.

CRediT author statements

Tran Thang Do: Conceptualization, Methodology, Supervision; Van Phong Dinh, Quang Hoang Nguyen, Duc Hoang Chu: Data curation, Software, Visualisation, Investigation, Writing original draft preparation.

ACKNOWLEDGEMENTS

This work was simultaneously supported by the Vietnam Academy of Science and Technology for developing robotic technology in microbiological fertilizer production lines; and in part by the National Program: Support For Research, Development, and Technology Application of Industry 4.0 (KC-4.0/19-25), under the grant for the project: Research, design, and manufacture of cobot applied in industry and some other fields with Human-Robot Interaction (code: KC-4.0-35/19-35).

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] A. Granlund, N. Friedler (2012), "A model for the formulation of an automation strategy", *4th World Conference P&OM/19th EUROMA Conference*.

[2] M. Huber, M. Rickert, A. Knoll, et al. (2008), "Human-robot interaction in handing-over tasks", *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, DOI: 10.1109/ROMAN.2008.4600651.

[3] M. Gualtieri, R. Platt (2018), "Learning 6-DoF grasping and pick-place using attention focus", *2nd Conference on Robot Learning (CoRL 2018), Zurich, Switzerland*, pp.1-10

[4] N. Wang, D.V. Pynadath, S.G. Hill (2015), "Building trust in a human-robot team with automatically generated explanations", *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 12pp.

[5] N. Correll, K.E. Bekris, D. Berenson, et al. (2018), "Analysis and observations from the first amazon picking challenge", *IEEE Transactions on Automation Science and Engineering*, **15(1)**, pp.172-188, DOI: 10.1109/TASE.2016.2600527.

[6] S. Kammel, J. Ziegler, B. Pitzer, et al. (2008), "Team AnnieWAY's autonomous system for the 2007 DARPA Urban Challenge", *Journal of Field Robotics*, **25(9)**, pp.615-639, DOI: 10.1007/978-3-642-03991-1_9.

[7] S.Y. Feng, E. Whitman, X. Xinjilefu, et al. (2015), "Optimization-based full body control for the DARPA robotics challenge", *Journal of Field Robotics*, **32(2)**, pp.293-312, DOI: 10.1109/HUMANOIDS.2014.7041347.

[8] The National Aeronautics and Space Administration, *Past Challenges,* https://www.nasa.gov/directorates/spacetech/centennial_challenges/sample_return_robot/index.html, accessed 4 April 2021.

[9] Rethink Robotics, *Sawyer Black Edition* http://www.rethinkrobotics.com/baxter/, accessed 4 April 2021.

[10] ABB, *Detailed Information for: IRB 14000,* https://new.abb.com/products/robotics/collaborative-robots/yumi/irb-14000-yumi, accessed 4 April 2021.

[11] I.A.M.J. Broeders, J. Ruurda (2001), "Robotics revolutionizing surgery: The intuitive surgical 'Da Vinci' system", *Industrial Robot*, **28(5)**, pp.387-392, DOI: 10.1108/EUM0000000005845.

[12] U. Hagn, M. Nickl, S. Jörg, et al. (2008), "The DLR MIRO: A versatile lightweight robot for surgical applications", *Industrial Robot*, **35(4)**, pp.324-336, DOI: 10.1108/01439910810876427.

[13] Z.M. Tian, W.S. Lu, T.M. Wang, et al. (2008), "Application of a robotic telemanipulation system in stereotactic surgery", *Stereotactic and Functional Neurosurgery*, **86(1)**, pp.54-61, DOI: 10.1159/000110742.

[14] Y.S. Sun, D.M. Wu, Z.J. Du, et al. (2001), "Robot-assisted needle insertion strategies based on liver force model", *Robot*, **33(1)**, pp.66-70, DOI: 10.3724/SP.J.1218.2011.00066.

[15] International Organization for Standardization (2016), ISO/TS 15066:2016, https://www.iso.org/standard/62996.html, accessed 4 April 2021.

[16] S.S. Chang, K.I. Baek, T. Hsiai, et al. (2016), *What Optimization Principle Explains The Zebrafish Vasculature?*, APS Division of Fluid Dynamics.

[17] F. Ore, L. Hanson, N. Delfs, et al. (2015), "Human - Industrial robot collaboration-development and application of simulation software", *International Journal of Human Factors Modelling and Simulation*, **5(2)**, pp.164-185, DOI: 10.1016/j.procir.2016.02.002.

[18] S. Wolf, G. Hirzinger (2008), "A new variable stiffness design: Matching requirements of the next robot generation", *IEEE International Conference on Robotics and Automation*, DOI: 10.1109/ROBOT.2008.4543452.

[19] J. Choi, S. Hong, W. Lee, et al. (2011), "A robot joint with variable stiffness using leaf springs", *IEEE Transactions on Robotics*, **27(2)**, pp.229-238, DOI: 10.1109/TRO.2010.2100450.

[20] S. Wolf, O. Eiberger, G. Hirzinger (2011), "The DLR FSJ: Energy based design of a variable stiffness joint", *IEEE International Conference on Robotics and Automation*, DOI: 10.1109/ICRA.2011.5980303.

[21] A.M. Zanchettin, N.M. Ceriani, P. Rocco, et al. (2016), "Safety in human-robot collaborative manufacturing environments: Metrics and control", *IEEE Transactions on Automation Science and Engineering*, **13(2)**, pp.882-893, DOI: 10.1109/TASE.2015.2412256.

[22] M. Zinn, O. Khatib, B. Roth (2004), "A new actuation approach for human friendly robot design", *IEEE International Conference on Robotics and Automation*, DOI: 10.1109/ROBOT.2004.1307159.

[23] J.S. Gutmann, M. Fukuchi, M. Fujita (2008), "3D perception and environment map generation for humanoid robot navigation", *International Journal of Robotics Research*, **27(10)**, pp.1117-1134, DOI: 10.1177/0278364908096316.

[24] A. Schmitz, P. Maiolino, M. Maggiali, et al. (2011), "Methods and technologies for the implementation of large-scale robot tactile sensors", *IEEE Transactions on Robotics*, **27(3)**, pp.389-400, DOI: 10.1109/TRO.2011.2132930.

[25] A. Fanaei, M. Farrokhi (2006), "Robust adaptive neuro-fuzzy controller for hybrid position/force control of robot manipulators in contact with unknown environment", *Journal of Intelligent & Fuzzy Systems*, **17(2)**, pp.125-144.

[26] H. Masuta, N. Kubota (2010), "Information reduction for environment perception of an intelligent robot arm equipped with a 3D range camera", *Proceedings of SICE Annual Conference 2010*, pp.392-397.

[27] R.O. Saber, J.A. Fax, R.M. Murray (2007), "Consensus and cooperation in networked multi-agent systems", *Proceedings of The IEEE*, **95(1)**, pp.215-233, DOI: 10.1109/JPROC.2006.887293.

[28] G.D. Shi, K.H. Johansson (2011), "Multi-agent robust consensus - Part Convergence analysis", *50th IEEE Conference on Decision and Control and European Control Conference*, DOI: 10.1109/CDC.2011.6160957.

[29] Y.G. Sun, L. Wang, G.M. Xie (2008), "Average consensus in networks of dynamic agents with switching topologies and multiple time-varying delays", *Systems & Control Letters*, **57(2)**, pp.175-183, DOI: 10.1016/j.sysconle.2007.08.009.

[30] M. Brambilla, E. Ferrante, M. Birattari, et al. (2013), "Swarm robotics: A review from the swarm engineering perspective", *Swarm Intelligence*, **7(1)**, pp.1-41, DOI: 10.1007/s11721-012-0075-2.

[31] H.T. Xue, Y.Y. Ye, L.C. Shen, et al. (2001), "A roadmap of multi-agent system architecture and coordination research", *Robot*, **23(1)**, pp.85-90, DOI: 10.3321/j.issn:1002-0446.2001.01.017 (in Chinese).

[32] M. Flint, M. Polycarpou, E.F. Gaucherand (2002), "Cooperative control for multiple autonomous UAV's searching for targets", *Proceedings of The 41st IEEE Conference on Decision and Control*, DOI: 10.1109/CDC.2002.1184272.

[33] A.T. Hafez, A.J. Marasco, S.N. Givigi, et al. (2015), "Solving multi-UAV dynamic encirclement via model predictive control", *IEEE Transactions on Control Systems Technology*, **23(6)**, pp.2251-2265, DOI: 10.1109/TCST.2015.2411632.

[34] A.L. Yang, W. Naeem, M.R. Fei, et al. (2017), "Multiple robots formation manoeuvring and collision avoidance strategy", *International Journal of Automation and Computing*, **14(6)**, pp.696-705, DOI: 10.1007/s11633-016-1030-2.

[35] X.J. Zhu, A.B. Goldberg, R. Brachman, et al. (2009*)*, *Introduction to Semi-supervised Learning*, Springer, DOI: 10.1007/978-3-031-01548-9.

[36] P. Englert, A. Paraschos, M.P. Deisenroth, et al. (2013), "Probabilistic model-based imitation learning", *Adaptive Behavior*, **21(5)**, pp.388-403, DOI: 10.1177/1059712313491614.

[37] B.D. Argall, S. Chernova, M. Veloso, et al. (2009), "A survey of robot learning from demonstration", *Robotics and Autonomous Systems*, **57(5)**, pp.469-483, DOI: 10.1016/j.robot.2008.10.024.

[38] A. Paraschos, C. Daniel, J. Peters, et al. (2013), *Probabilistic Movement Primitives*, https://core.ac.uk/download/pdf/80684028.pdf.

[39] S.M.K. Zadeh, A. Billard (2011), "Learning stable nonlinear dynamical systems with Gaussian mixture models", *IEEE Transactions on Robotics*, **27(5)**, pp.943-957, DOI: 10.1109/TRO.2011.2159412.

[40] J. Kober, K. Mölling, O. Krömer, et al. (2010), "Movement templates for learning of hitting and batting", *IEEE International Conference on Robotics and Automation*, DOI: 10.1109/ROBOT.2010.5509672.

[41] S. Levine, P. Pastor, A. Krizhevsky, et al. (2016), "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection", *The International Journal of Robotics Research*, **37(10)**, DOI: 10.1177/0278364917710318.

[42] C. Finn, S. Levine (2017), "Deep visual foresight for planning robot motion", *IEEE International Conference on Robotics and Automation*, DOI: 10.1109/ICRA.2017.7989324.

[43] J.J. Kuffner, S.M. LaValle (2011), "Space-filling trees: A new perspective on incremental search for motion planning", *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, DOI: 10.1109/IROS.2011.6094740.

[44] G.H. Tian, Y.W. Xu (2014), "Cloud robotics: Concept, architectures and key technologies", *Journal of Shandong University (Engineering Science)*, **44(6)**, pp.47-54.

[45] M. Quigley, K. Conley, B. Gerkey, et al. (2009), "ROS: An opensource robot operating system", *ICRA Workshop on Open Source Software*, 6pp.

[46] M. Yuriyama, T. Kushida (2010), "Sensor-cloud infrastructure physical sensor management with virtualized sensors on cloud computing", *13th International Conference on Network-Based Information Systems*, DOI: 10.1109/NBiS.2010.32.

[47] S. Nakagawa, N. Igarashi, Y. Tsuchiya, et al. (2012), "An implementation of a distributed service framework for cloud-based robot services", *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*, DOI: 10.1109/IECON.2012.6389225.

[48] L. Turnbull, B. Samanta (2013), "Cloud robotics: Formation control of a multi robot system utilizing cloud infrastructure", *Proceedings of IEEE Southeastcon*, DOI: 10.1109/SECON.2013.6567422.

[49] B. Kehoe, A. Matsukawa, S. Candido, et al. (2013), "Cloud-based robot grasping with the Google object recognition engine", *IEEE International Conference on Robotics and Automation*, DOI: 10.1109/ ICRA.2013.6631180.

[50] D. Silver, A. Huang, C.J. Maddison, et al. (2016), "Mastering the game of Go with deep neural networks and tree search", *Nature*, **529(7587)**, pp.484-489, DOI: 10.1038/nature16961

DOI: 10.31276/VJSTE.64(4).45-50

Fuzzy finite element analysis based on the transformation between fuzzy and random variables

Tuan Hung Nguyen[1*], Huynh Xuan Le[2]

[1]*Faculty of Civil Engineering, Thuyloi University, 175 Tay Son Street, Trung Liet Ward, Dong Da District, Hanoi, Vietnam*
[2]*Hanoi University of Civil Engineering, 55 Giai Phong Street, Dong Tam Ward, Hai Ba Trung District, Hanoi, Vietnam*

Received 5 May 2021; revised 29 June 2021; accepted 14 July 2021

Abstract:

In this paper, a fuzzy finite element method (FFEM) for determining the responses of structures is proposed by using the transformation between fuzzy and random variables. Firstly, the formulae for establishing normal random variables equivalent to symmetric triangular fuzzy numbers are presented based on the combination of the principle of insufficient reason and that of maximum specificity. As a result, fuzzy finite element analysis is transferred into stochastic finite element analysis. To solve this problem, the response surface method with the aid of standard normal random variables is utilized to approximate the real responses of structures. Then, the errors between the training and test sets are estimated to select a suitable response surface model amongst the regression models. Lastly, the formulae for determining the mean and the standard deviation values of the responses of structures are established. The accuracy and effectiveness of the proposed method are verified via an illustrative example.

Keywords: **fuzzy finite element, fuzzy sets theory, possibility-probability transformations, response surface method, stochastic finite element, surrogate model.**

Classification number: **2.3**

1. Introduction

FFEM is the combination of finite element method (FEM) and fuzzy sets theory [1], which is used to define the responses of structures in cases where the input quantities contain incomplete information such as loads, material and geometric properties, stiffness of supports, etc., and is described in the form of fuzzy numbers. In recent years, a large number of FFEMs have been proposed beginning with static analysis, then extending into the dynamic analysis of structures. Fundamental strategies of FFEM can be categorized into main groups as follows: the interval arithmetic approach for static analysis of structures [2-8], the optimization strategy for static and dynamic analysis of structures [9-20], the combination of interval arithmetic and optimization strategy for dynamic analysis of structures [21-23], and applying stochastic finite element methods (SFEM) [24] for fuzzy analysis of structures [25-27]. Amongst these approaches, the α-optimization method [9] has been gradually acknowledged as the standard procedure for FFEM. In this method, the search process is performed in the input domain to seek the exact bounds on the objective function by iteratively evaluating the objective function at designated points. Therefore, this method is very time-consuming because it must be accomplished with a large amount of finite element analysis.

To overcome this limitation, this study focuses on a novel approach for analysing fuzzy finite elements by using the transformation between fuzzy and random variables. Firstly, based on the combination of the principle of insufficient reason and that of maximum specificity, normal random variables equivalent to symmetric triangular fuzzy numbers of the input data are presented and explored in detail. As a result, these fuzzy numbers are replaced by normal random variables so SFEM such as Monte Carlo simulations, the perturbation method, the spectral stochastic finite element method, and so on can be used for the analysis of structures. In this study, the response surface method (RSM) is utilized to approximate the real responses of structures. Then, the least-squares error criterion between the training and the test sets is proposed to select the suitable response surface model amongst the regression models. Lastly, the mean and the standard deviation values of the responses of structures are directly calculated from the response surface model. The proposed method is verified via a two-bar truss structure with spring supports.

2. Methods

2.1. The overview of transformation principles and the formulae for determining the deviation of the equivalent normal variables

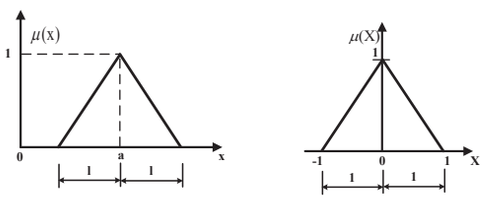
The transformation from fuzzy numbers into random quantities and its converse should be taken into account in any problem where heterogeneous uncertain and imprecise data appear together (e.g., information deficit, linguistic variables, statistical data). Representative transformation principles such as insufficient reason, maximum specificity, and uncertainty invariance are proposed by D. Dubois, et al. (1993) [28], D. Dubois, et al. (2004) [29], D. Dubois (2006) [30], and G.J. Klir (2005) [31], respectively. The advantage of Klir's principle is the information preservation

*Corresponding author: Email: hungtuan@tlu.edu.vn

of fuzzy measure and equivalent probability. However, in Ref. [1], the results of using one may conflict with the probability/possibility consistency principle. Besides, three assumptions must be used in Klir's approach while Dubois's approach does not make any assumptions. Nevertheless, an initial fuzzy number will differ from the final fuzzy number after a transform forward to probability measure by the principle of insufficient reason and then a transform back to fuzzy measure by the principle of maximum specificity. Hence, transformations (from possibility to probability and converse) based on these two principles make non-conservative information. To overcome this drawback, T.H. Nguyen, et al. (2019) [32] proposed an innovation transformation to calculate the deviation of the equivalent normal random variable. This approach is presented and detailed explored below.

Consider a symmetric triangular fuzzy number $\tilde{X}_i = (a, l)_{LR}$ (Fig. 1A) where a is the belief value (at the membership level μ=1) of the fuzzy number; l is the spread of the fuzzy number; and the standardized fuzzy variable $\tilde{x}_i = (0,1)_{LR}$ is defined as follows [19]:

$$\tilde{x}_i = \frac{\tilde{X}_i - a}{l} \quad (1)$$



(A) Original fuzzy variable (B) Standardized fuzzy variable

Fig. 1. Transformation to a standardized fuzzy variable.

For the transformation from a standardized symmetric triangular fuzzy number (Fig. 1B) into a random quantity, the error of probability measure between the equivalent probability density function $p(x)$ of the standardized symmetric triangular fuzzy number obtained by the principle of insufficient reason and the probability density function of the normal random variable $p_1(x)$ is expressed by the following formula:

$$(P(A) - P_1(A))^2 \to min, \forall x \in [-1,0) \quad (2)$$

where p(x), p_1(x), P(A), and P1(A) are determined as follows:

$$p(x) = \begin{cases} -\frac{1}{2} ln(-x) ; x \in [-1,0) \\ -\frac{1}{2} ln(x) ; x \in 0,1] \end{cases} \quad (3)$$

$$p_1(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{x^2}{2\sigma^2}\right)} \quad (4)$$

$$P(A) = \frac{1}{2}[x - xln(-x) + 1] \quad (5)$$

$$P_1(A) = \int_{-1}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{x^2}{2\sigma^2}\right)} dx \quad (6)$$

with σ being the deviation of the normal random variable, {A} is the event that has $1 \le x_o \le x$, and P(A) and P_1(A) are probabilities of event A for the density distribution functions p(x) and p_1(x).

From Eq. (2), we obtain:

$$F_1(\sigma) = \int_{-1}^{0} (P(A) - P_1(A))^2 \, dx \to min \quad (7)$$

The probability density function p(x) and normal random variable p_1(x) have different domains. Therefore, to ensure that the probability of the density function p_1(x) in (-∞,1) is insignificant, one needs

$$F_2(\sigma) = P_1[A *: x_o \in (-\infty, -1)] = \int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{x^2}{2\sigma^2}\right)} dx \to min \quad (8)$$

Combining Eqs. (7) and (8), we have:

$$F(\sigma) = F_1(\sigma) + F_2(\sigma) = \int_{-1}^{0} (P(A) - P_1(A))^2 \, dx + \int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{x^2}{2\sigma^2}\right)} dx \to min \quad (9)$$

For the reverse transformation, that is, from the normal random variable into an equivalent fuzzy number, the error of possibility measure between the equivalent fuzzy number of the normal random variable obtained by the principle of maximum specificity and the standardized fuzzy number is expressed by the following formula:

$$G(\sigma) = \int_{-1}^{0} (\pi_1(x) - \pi(x))^2 \, dx + \int_{-\infty}^{-1} {\pi_1}^2(x) \, dx = \int_{-1}^{0} (\pi_1(x) - 1 - x)^2 \, dx + \int_{-\infty}^{-1} {\pi_1}^2(x) \, dx \to min \quad (10)$$

where π(x) is the membership function of the standardized fuzzy number in [-1,0] as:

$$\pi(x) = 1 + x \quad (11)$$

Then, π_1(x) is the equivalent fuzzy number of a normal random variable determined as follows (with limits −∞ and +∞ replaced by -6σ and 6σ, respectively):

$$\pi_1(x) = \pi_1(-x) = \int_{-6\sigma}^{x} p_1(y) dy + \int_{-x}^{6\sigma} p_1(y) dy \quad (12)$$

To solve the multi-objective optimization problem of Eqs. (9) and (10), one transforms multiple objectives into a scalar objective function by multiplying each objective function by a weighting factor and summing up all contributors as in:

$$H(\sigma) = \gamma F(\sigma) + (1 - \gamma) G(\sigma) \to \min \quad (13)$$

where γ ∈[0,1].

The mathematical meaning of Eq. (13) is an extension that modifies the equivalent characteristic according to two principles: the principle of insufficient reason when going from a fuzzy number to random variable and the principle of maximum specificity when going from a random variable to fuzzy number. To solve Eq. (13), a genetic algorithm (GA) is applied using built-in functions in Matlab. The relation between the weighting factor γ and deviation σ is depicted in detail in Fig. 2.

One realizes that the proposed transformation encodes a family of normal random variables, including the result of Klir's method, from the standardized fuzzy number. Indeed, the deviation of a normal random variable using the uncertainty invariance principle is 0.3989 [25] and the same result is attained according to the proposed transformation when the weighting factor is 0.728. Hence, the constraints generated in the proposed transformation are more flexible than Klir's method.

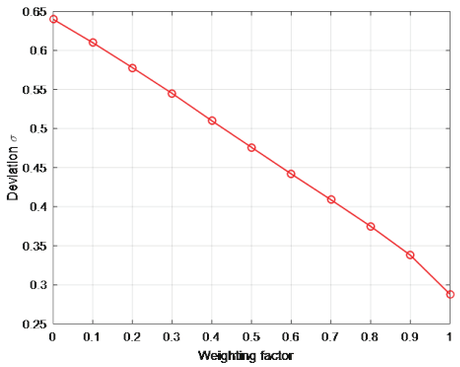


Fig. 2. Representation of the relation between deviation and weighting factor.

2.2. The calculation for the responses of structures

After a transformation from fuzzy to random variables, fuzzy finite element analysis becomes stochastic finite element analysis. Then, RSM is utilized to define the responses of the structures. The basic idea of RSM is to approximate an implicit function by an equivalent polynomial function. The calculation procedure is expressed in the next sub-sections.

Design of experiments: The design of the experiments is based on the sampling plan in design variable space (i.e., input data space). The design plays a very valuable role in determining the regression coefficients of surrogate models. The important problem is how we evaluate the goodness of such designs with a limited number of samples. In the present paper, the Box-Behnken design [33], which has good results in actual problems, is selected to be the design of the experiments. In the Box-Behnken design, sample points are chosen at all possible combinations of the mean values (μ_i=0) and $\mu_i \pm 3\sigma_i$ (0±3) as shown in Fig. 3.

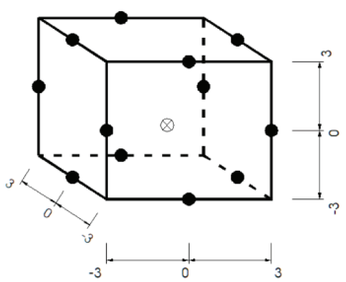


Fig. 3. The Box-Behnken design with three standard normal random variables.

Surrogate models: In statistical theory, surrogate models are often used that include the polynomial regression model, Kringing model, and radial basis functions [34]. The first and the second models are parametric models based on the assumed functional form of the response in terms of the design variables. The last model is non-parametric and uses different types of local models in different regions of the data to build up an overall model. Among these models, the polynomial regression model is often used to build a response surface function due to its calculation simplicity. In this study, the complete quadratic polynomial regression model is used for the responses of structure, in which all random variables X_{ci} are standard normal and assumed to be uncorrelated.

The complete quadratic polynomial regression model (The CQP model):

$$y(\mathbf{X}) = a_o + \sum_{i=1}^{n} a_i X_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij} X_i X_j + \sum_{i=1}^{n} a_{ii} X_i^2 \quad (14)$$

According to the extension principle [1], the belief values of the fuzzy response of structures are received from the combination of belief values of input data. Hence, the surrogate model needs to respond to this constrained condition, such as in:

$$a_o = \hat{y}(\mathbf{x} = \mathbf{a}) \quad (15)$$

where ŷ(x=a) are displacements at the membership level μ=1 of the input data and are determined by classical FEM.

The remaining regression coefficients in Eq. (14) are determined by the least-squares method.

Error estimation and selecting reasonable design: Due to their ease of calculation, split sample and cross-validation methods [34] are always used to select a rational design. To reduce the variance of the error estimation, which usually appears in the split sample method, the cross-validation method is proposed. Differing from the split sample method, this method makes no distinction between the training and test data sets. In this study, leave-one-out cross-validation is applied where each response point is tested once and trained *k-1* times because the centre point was computed from Eq. (15). The error estimation of the j^{th} design (using $X^{(j)}$ as the test set) is determined by the formula:

$$GSE_j = (y_j - \hat{y}_j^{(-j)})^2 \quad (16)$$

where GSE_j is the square error estimation of the j^{th} design; y_j is output value at $X^{(j)}$ determined by classical FEM; and $\hat{y}_j^{(-j)}$ is the estimated value at $X^{(j)}$ of the j^{th} design.

The design uses the least-squares error estimation method to calculate the responses of the structures.

Determination for the responses of structures: The parameters of the responses of structures, including the mean and the deviation values, are directly calculated from the surrogate model. After using the properties for the n^{th} moments of the standard normal variable, we obtain:

- The mean value of the responses of structures:

$$m_r = E(y) = a_o + \sum_{i=1}^{n} a_{ii} \quad (17)$$

- The standard deviation value of the responses of structures:

$$\sigma_r = \sigma(y) = \sqrt{E(y^2) - (E(y))^2} \quad (18)$$

where $E(y^2)$ is determined as follows:

$$E(y^2) = (a_o)^2 + \sum_{i=1}^{n} (a_i)^2 + 3\sum_{i=1}^{n} (a_{ii})^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (a_{ij})^2 + 2\sum_{i=1}^{n-1} \sum_{k=i+1}^{n} a_{ii} a_{kk} + 2a_o \sum_{i=1}^{n} a_{ii} \quad (19)$$

3. Results and discussion

Consider a two-bar truss with a spring support in Fig. 4, where the elastic modulus of the material is $\tilde{E}$, the loads are $\tilde{P}_1$

and $\tilde{P}_2$, and the spring stiffness $\tilde{k}$ are symmetric triangular fuzzy numbers: $\tilde{E}=(210,20)_{LR}$ GPa; $\tilde{P}_1=(30,3)_{LR}$ kN; $\tilde{P}_2=(25,3)_{LR}$ kN; $\tilde{k}=(2000,400)_{LR}$ kN/m.

The cross-sectional area of both bars is $A_1=A_2=5.10^{-4}$ m^2.

Required: Determine the horizontal displacement u_1 and the vertical displacement v_1.

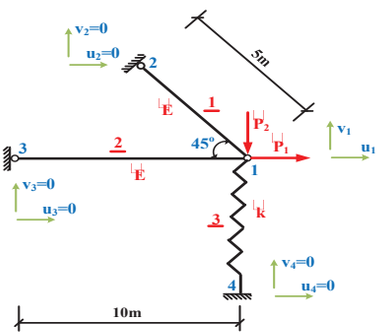


Fig. 4. Two-bar truss with spring support.

One realizes that the weighting factor of 0.5 represents the equivalent characteristic according to two principles: The principle of the insufficient reason when going from fuzzy numbers to random variables and the principle of maximum specificity when going from random variable to fuzzy number. Hence, in this study, we utilize the weighting factor to calculate the responses of the structure. The other weighting factor values are used to survey the variation of the responses of the structure.

The mean values m_r and the standard deviation values σ_r of the horizontal displacement u_1 and the vertical displacement v_1 for each value of the weighting factor γ are shown in Table 1.

Table 1. The mean values m_r and the standard deviation values σ_r of the horizontal displacement u_1 and the vertical displacement v_1.

The weighting factor γ	The horizontal displacement u_1 (m)		The vertical displacement v_1 (m)	
	The mean value m_r	*The standard deviation value σ_r*	*The mean value m_r*	*The standard deviation value σ_r*
0	0.0007425	0.000218	-0.0013835	0.000310
0.1	0.0007395	0.000036	-0.0013735	0.000248
0.2	0.0007397	0.000068	-0.0013793	0.000096
0.3	0.000740	0.000102	-0.0013799	0.000144
0.4	0.0007406	0.000136	-0.0013807	0.000192
0.5	0.0007411	0.000161	-0.0013815	0.000229
0.6	0.0007421	0.000204	-0.0013830	0.000290
0.7	0.0007431	0.000239	-0.0013840	0.000343
0.8	0.0007451	0.000273	-0.0013862	0.000391
0.9	0.0007464	0.000310	-0.0013878	0.000439
1	0.000740	0.000098	-0.0013798	0.000138

Tables 2 and 3 show the results of fuzzy horizontal displacement $\tilde{u}_1$ and fuzzy vertical displacement $\tilde{v}_1$ using the α-optimization method, respectively. Membership functions of fuzzy horizontal displacement $\tilde{u}_1$ and fuzzy vertical displacement $\tilde{v}_1$ are depicted in Figs. 5 and 6, respectively.

Table 2. The result of fuzzy horizontal displacement $\tilde{u}_1$ using the α-optimization method.

α-cuts	The α-optimization method	
	Lower u_1 (m)	*Upper u_1 (m)*
0.0	0.000188	0.001352
0.2	0.000294	0.001224
0.4	0.000402	0.001098
0.6	0.000512	0.000976
0.8	0.000624	0.000856
1.0	0.000739	0.000739

Table 3. The result of fuzzy vertical displacement $\tilde{v}_1$ using the α-optimization method.

α-cuts	The α-optimization method	
	Lower v_1 (m)	*Upper v_1 (m)*
0.0	-0.002283	-0.000675
0.2	-0.002083	-0.000803
0.4	-0.001893	-0.000937
0.6	-0.001713	-0.001078
0.8	-0.001542	-0.001225
1.0	-0.001379	-0.001379

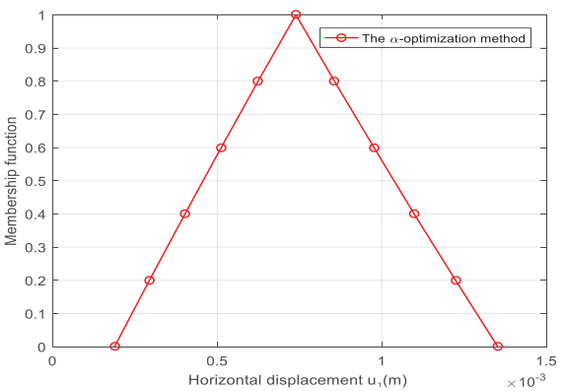


Fig. 5. Fuzzy horizontal displacement $\tilde{u}_1$.

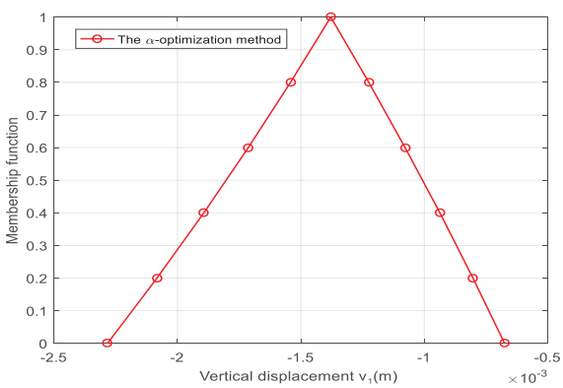


Fig. 6. Fuzzy vertical displacement $\tilde{v}_1$.

To verify the accuracy of the proposed methods, the mean value m_r and the confidence interval $[m_r\text{-}h.\sigma_r, m_r+h.\sigma_r]$ are compared with the belief value and the support of fuzzy response of the structure using the α-optimization method, respectively. Because the non-linearity terms are comprised of the responses of structure, the load effects (the output data) such as displacements and internal forces exhibit non-Gaussian probabilistic characteristics. Hence, the interval $[m_r\text{-}3.2\sigma_r, m_r+3.2\sigma_r]$, which corresponds to the largest 99% confidence interval for the Laplace, the uniform, and the triangular probability distributions [29] is chosen. Tables 4 and 5 show the percentage errors between the mean values at each value of the weighting factor γ and the belief values for the horizontal displacement u_1 and the vertical displacement v_1, respectively. The differences between the results of the confidence intervals and that of the support of fuzzy displacements $\tilde{u}_1$ and $\tilde{v}_1$ are calculated in Tables 6 and 7, respectively, including percentage errors in the lower bound, the upper bound, and the width of solutions. The formulae for calculating these percentage errors are presented in [4].

Table 4. The percentage errors between the mean and belief values for the horizontal displacement u_1.

The weighting factor γ	The mean value m_r (m)	The belief value (m)	Error (%)
0	0.0007425	0.000739	-0.48864
0.1	0.0007395	0.000739	-0.07360
0.2	0.0007397	0.000739	-0.10331
0.3	0.000740	0.000739	-0.15292
0.4	0.0007406	0.000739	-0.22886
0.5	0.0007411	0.000739	-0.29201
0.6	0.0007421	0.000739	-0.43162
0.7	0.0007431	0.000739	-0.57072
0.8	0.0007451	0.000739	-0.84072
0.9	0.0007464	0.000739	-1.00942
1	0.000740	0.000739	-0.14791

Table 5. The percentage errors between the mean and belief values for the vertical displacement v_1.

The weighting factor γ	The mean value m_r (m)	The belief value (m)	Error (%)
0	-0.0013835	-0.001379	-0.30406
0.1	-0.0013735	-0.001379	0.41886
0.2	-0.0013793	-0.001379	-0.00017
0.3	-0.0013799	0.000739	-0.04119
0.4	-0.0013807	-0.001379	-0.09851
0.5	-0.0013815	-0.001379	-0.26862
0.6	-0.0013830	-0.001379	-0.34153
0.7	-0.0013840	-0.001379	-0.40643
0.8	-0.0013862	-0.001379	-0.50283
0.9	-0.0013878	-0.001379	-0.61848
1	-0.0013798	-0.001379	-0.03628

Table 6. The percentage errors between the confidence intervals and support of fuzzy horizontal displacement $\tilde{u}_1$.

The weighting factor γ	The confidence interval $[m_r\text{–}3.2\sigma_r, m_r+3.2\sigma_r]$		The support of fuzzy displacement $\tilde{u}_1$		Error LB (%)	Error UB (%)	Error width (%)
	Lower (m)	*Upper (m)*	*Lower (m)*	*Upper (m)*			
0	0.000044	0.001441	0.000188	0.001352	76.57	6.57	19.96
0.1	0.000625	0.000853	0.000188	0.001352	233.53	36.88	80.42
0.2	0.000522	0.000958	0.000214	0.001320	143.91	27.43	60.56
0.3	0.000414	0.001066	0.000214	0.001320	93.70	19.24	41.08
0.4	0.000307	0.001175	0.000240	0.001288	27.60	8.78	17.12
0.5	*0.000224*	*0.001258*	*0.000240*	*0.001288*	*6.66*	*2.31*	*1.31*
0.6	0.000088	0.001396	0.000267	0.001256	66.93	11.18	32.26
0.7	-0.000023	0.001509	0.000267	0.001256	108.62	20.20	54.97
0.8	-0.000128	0.001618	0.000294	0.001224	143.53	32.21	87.67
0.9	-0.000247	0.001740	0.000320	0.001192	177.15	45.95	127.93
1	0.000427	0.001053	0.000320	0.001192	33.37	11.69	28.25

Table 7. The percentage errors between the confidence intervals and the support of fuzzy vertical displacement $\tilde{v}_1$.

The weighting factor γ	The confidence interval $[m_r\text{-}3.2\sigma_r, m_r+3.2\sigma_r]$		The support of fuzzy displacement $\tilde{v}_1$		Error LB (%)	Error UB (%)	Error width (%)
	Lower (m)	*Upper (m)*	*Lower (m)*	*Upper (m)*			
0	-0.002376	-0.000391	-0.002283	-0.000675	4.04	42.03	23.37
0.1	-0.002167	-0.000580	-0.002283	-0.000675	5.09	14.09	1.31
0.2	-0.001685	-0.001073	-0.002283	-0.000675	26.20	59.07	61.97
0.3	-0.001841	-0.000919	-0.002283	-0.000675	19.39	36.20	42.71
0.4	-0.001996	-0.000765	-0.002283	-0.000675	12.58	13.38	23.47
0.5	*-0.002115*	*-0.000648*	*-0.002283*	*-0.000675*	*7.37*	*4.02*	*8.77*
0.6	-0.002312	-0.000454	-0.002283	-0.000675	1.24	32.67	15.46
0.7	-0.002481	-0.000287	-0.002283	-0.000675	8.64	57.41	36.34
0.8	-0.002636	-0.000137	-0.002283	-0.000675	15.44	79.77	55.38
0.9	-0.002793	0.000017	-0.002283	-0.000675	22.31	102.57	74.70
1	-0.001822	-0.000938	-0.002283	-0.000675	20.23	39.00	45.08

4. Discussion

Based on the results of the above example, the following discussions are presented:

- The mean values of the displacements obtained by the proposed method are close to the belief values of fuzzy displacements by the α-optimization method. Due to the difference between the mathematics of fuzzy sets and that of random parameters, this demonstrates that the complete quadratic polynomial in which all equivalent random variables are standard normal random variables using Eqs. (14) and (15) is a reasonable regression model to determine the displacements of the structure.

- The confidence intervals of the displacements with the weighting factor γ=0.5 approximate the supports of fuzzy displacements (the percentage errors are less than 10%). One realizes that a deviation of 0.476 for the equivalent normal variable, which corresponds to the weighting factor γ=0.5, is a plausible selection for calculating the responses of structures. This also points out that the proposed method is reliable. Indeed, when the weighting factor γ=0.728 corresponding to the method [25], the average percentage of errors

is about 50%, which is much more than the largest percentage errors with the weighting factor γ=0.5. Once the spread of the fuzzy spring stiffness k=200 kN/m, which is the same as the example presented in [35], the percentage errors between the confidence intervals of the displacements corresponding to the weighting factor γ=0.5 and the support of fuzzy displacements are also less than 10%.

- The number of computations obtained by the proposed method is less than that of the vertex method [12]. Indeed, in the illustrative example, the proposed method requires 25 deterministic finite element problems while 81 deterministic finite element problems are needed for analysis by the vertex method. This especially reveals that besides achieving reasonable accuracy, the proposed method is also a solution to reduce the number of computations.

5. Conclusions

This paper presents a new method for analysing FFE by using the transformation between fuzzy and random variables. The novel formulae for determining equivalent normal random variables are established and explored in detail. By using the standard normal random variables in the quadratic polynomial regression models and selecting the suitable response surface model amongst the regression models, reasonable accuracy can be achieved for the responses of structures. Simultaneously, by applying suitable experimental designs, a reduction in the number of computations is demonstrated. A weighting factor of 0.5 is a reasonable selection to calculate the responses of structures. Thorough surveys of more complex examples will be presented in future research. Additionally, the present study will be extended to asymmetric triangular fuzzy numbers.

CRediT author statement

Tuan Hung Nguyen, Huynh Xuan Le: Data curation, Software, Visualisation, Investigation, Writing original draft preparation, Conceptualization, Methodology, Supervision.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] D. Dubois, H. Prade (1980), *Fuzzy Sets and Systems*, Academic Press, 393pp.

[2] R.L. Muhanna, R.L. Mullen, H. Zang (2004), "Interval finite element as a basis for generalized models of uncertainty in engineering mechanics", *Proceedings of The NSF Workshop on Reliable Engineering Computing*, 18pp.

[3] H. Zang (2005), *Nondeterministic Linear Static Finite Element Analysis: An Interval Approach*, Ph.D. Dissertation, Georgia Institute of Technology, Atlanta.

[4] M.V.R. Rao, R.L. Mullen, R.L. Muhanna (2011), "A new interval finite element formulation with the same accuracy in primary and derived variables", *Int. J. Reliability and Safety*, **5(3/4)**, pp.336-357, DOI: 10.1504/IJRS.2011.041184.

[5] D. Degrauwe, G. Lombaert, G.D. Roeck (2010), "Improving interval analysis in finite element calculation by means of affine arithmetic", *Computers and Structures*, **88(3-4)**, pp.247-254, DOI: 10.1016/j.compstruc.2009.11.003.

[6] S. Adhikari, H.H. Khodaparast (2014), "A spectral approach for fuzzy uncertaint propagation in finite element analysis", *Fuzzy Sets and Systems*, **243**, pp.1-24, DOI: 10.1016/j.fss.2013.10.005.

[7] D. Behera, S. Chakraverty, H.Z. Huang (2016), "Non-probabilistic uncertain static responses of imprecisely defined structures with fuzzy parameters", *Journal of Intelligent and Fuzzy Systems*, **30**, pp.3177-3189, DOI: 10.3233/IFS-152061.

[8] J. Su, Y. Zhu, J. Wang, et al. (2018), "An improved interval finite element method based on the element by element technique for large truss system and plane problems", *Advances in Mechanical Engineering*, **10(4)**, pp.1-10, DOI: 10.1177/1687814018769159.

[9] B. Möller, M. Beer (2004), *Fuzzy Randomness - Uncertainty in Civil Engineering And Computational Mechanics*, Springer, 307pp.

[10] D. Degrauwe (2007), *Uncertainty Propagation in Structural Analysis By Fuzzy Numbers*, Ph.D. Dissertation, K.U.Leuven, Leuven.

[11] L. Farkas, D. Moens, D. Vandepitte, et al. (2010), "Fuzzy finite element analysis based on reanalysis technique", *Structural Safety*, **32(6)**, pp.442-448, DOI: 10.1016/j.strusafe.2010.04.004.

[12] W. Dong, H. Shan (1987), "Vertex method for computing functions of fuzzy variables", *Fuzzy Sets and Systems*, **24(1)**, pp.65-78, DOI: 10.1016/0165-0114(87)90114-X.

[13] M. Hanss (2005), *Applied Fuzzy Arithmetic - An Introduction With Engineering Applications*, Springer, Berlin, 244pp.

[14] S. Donders, D. Vandepite, J.V.D. Peer, et al. (2005), "Assessment of uncertainty on structural dynamic responses with the short transformation method", *Journal of Sound and Vibration*, **288**, pp.523-549, DOI: 10.1016/j.jsv.2005.07.003.

[15] O. Giannnini, M. Hanss (2008), "The component mode transformation method: A fast implementation of fuzzy arithmetic for uncertainty management in structural dynamics", *Journal of Sound and Vibration*, **311(3-5)**, pp.1340-1357, DOI: 10.1016/j.jsv.2007.10.029.

[16] U.O. Akpan, T.S. Koko, I.R. Orisamolu, et al. (2001), "Practical fuzzy finite element analysis of structures", *Finite Elements in Analysis and Design*, **38**, pp.93-111, DOI: 10.1016/S0168-874X(01)00052-X.

[17] M.D. Munck, D. Moens, W. Desmet, et al. (2008), "A response surface based optimisation algorithm for the calculation of fuzzy envelope FRFS of models with uncertain properties", *Computers and Structures*, **86(10)**, pp.1080-1092, DOI: 10.1016/j.compstruc.2007.07.006.

[18] A.S. Balu, B.N. Rao (2012), "High dimensional model representation based formulation for fuzzy finite element analysis of structures", *Finite Elements in Analysis and Design*, **50**, pp.217-230, DOI: 10.1016/j.finel.2011.09.012.

[19] N.H. Tuan, L.X. Huynh, P.H. Anh (2015), "A fuzzy finite element algorithm based on response surface method for free vibration analysis of structure", *Vietnam Journal of Mechanics*, **37(1)**, pp.17-27, DOI: 10.15625/0866-7136/37/1/3923.

[20] T.T. Viet, L.X. Huynh, V.Q. Anh (2017), "Fuzzy analysis for stability of steel frame with fixity factor modeled as triangular fuzzy number", *Advances in Computational Design*, **2(1)**, pp.29-42, DOI: 10.12989/acd.2017.2.1.029.

[21] H.D. Gersem, D. Moens, W. Desmet, et al. (2004), "Interval and Fuzzy finite element analysis of mechanical structures with uncertain parameters", *Proceedings of ISMA 2004*, pp.3009-3021.

[22] D. Moens, D. Vandepitte (2005), "A fuzzy finite element procedure for the calculation of uncertain frequency - response functions of damped structures: Part 1 - procedure", *Journal of Sound and Vibration*, **288(3)**, pp.431-462, DOI: 10.1016/j.jsv.2005.07.001.

[23] H.D. Gersem, D. Moens, W. Desmet, et al. (2005), "A fuzzy finite element procedure for the calculation of uncertain frequency - response functions of damped structures: Part 2 - numerical case studies", *Journal of Sound and Vibration*, **288(3)**, pp.463-486, DOI: 10.1016/j.jsv.2005.07.002.

[24] G. Stefanou (2009), "The stochastic finite element method: Past, present, and future", *Comput. Methods Appl. Mech. Engrg.*, **198(9-12)**, pp.1031-1051, DOI: 10.1016/j.cma.2008.11.007.

[25] Z. Lei, Q. Chen (2002), "A new approach to fuzzy finite element analysis", *Computer Methods in Applied Mechanics and Engineering*, **191(45)**, pp.5113-5118, DOI: 10.1016/S0045-7825(02)00240-2.

[26] H.Z. Huang, H.B. Li (2005), "Perturbation finite element method of structural analysis under fuzzy environment", *Engineering Application of Artificial Intelligence*, **18(1)**, pp.83-91, DOI: 10.1016/j.engappai.2004.08.033.

[27] P.H. Anh (2014), "Fuzzy analysis of laterally - loaded pile in layered soil", *Vietnam Journal of Mechanics*, **36(3)**, pp.173-183, DOI: 10.15625/0866-7136/36/3/3739.

[28] D. Dubois, H. Prad, S. Sandri (1993), "On possibility/probability transformations", *Fuzzy Logic*, pp.103-112.

[29] D. Dubois, L. Foulloy, G. Mauris, et al. (2004), "Probability - possibility transformations, triangular fuzzy sets, and probabilistic inequalities", *Reliable Computing*, **10**, pp.273-297, DOI: 10.1023/B:REOM.0000032115.22510.b5.

[30] D. Dubois (2006), "Possibility theory and statistical reasoning", *Computational Statistics & Data Analysis*, **51**, pp.47-59, DOI: 10.1016/j.csda.2006.04.015.

[31] G.J. Klir (2005), *Uncertainty and Information: Foundations of Generalized Information Theory*, John Wiley & Sons, DOI: 10.1002/0471755575.

[32] T.H. Nguyen, H.X. Le (2019), "A practical method for calculating reliability with a mixture of random and fuzzy variables", *Structural Integrity and Life*, **19(3)**, pp.175-183.

[33] R.L. Mason, R.F. Guns, J.L. Hess (2003), *Statistical Design and Analysis of Experiment: With Applications to Engineering and Science*, John Wiley & Sons (Second Editor), DOI: 10.1002/0471458503.

[34] N.V. Queipo, R.T. Haftka, W. Shyy, et al. (2005), "Surrogate - based analysis and optimization", *Progress in Aerospace Sciences*, **41(1)**, pp.1-28, DOI: 10.1016/j.paerosci.2005.02.001.

[35] L.X. Huynh, N.H. Tuan (2016), *Reliability of Civil Structures*, Construction Publishing House, Hanoi, Vietnam, 315pp (in Vietnamese).

DOI: 10.31276/VJSTE.64(4).51-54

The clinical features, angiographic and endovascular treatment of indirect carotid cavernous fistulas

Van Tuan Nguyen[1*], Anh Tuan Tran[2], Thi Hoang Yen Vu[3]

[1]*Department of Neurology, Bach Mai Hospital, 78 Giai Phong Street, Phuong mai Ward, Dong Da District, Hanoi, Vietnam*
[2]*Department of Radiology, Bach Mai Hospital, 78 Giai Phong Street, Phuong mai Ward, Dong Da District, Hanoi, Vietnam*
[3]*Department of Neurology, E Hospital, 89 Tran Cung Street, Co Nhue 1 Ward, Cau Giay District, Hanoi, Vietnam*

Received 25 May 2021; revised 7 August 2021; accepted 19 August 2021

Abstract:

Indirect carotid cavernous fistulas (ICCFs) are rare disease entities triggered by indirect low flow between the cavernous sinus and the meningeal branches of the internal or external carotid artery or from both. To evaluate clinical characteristics, angiography, and relationship to the result of endovascular treatment of indirect carotid cavernous fistulas. A prospective study of 32 patients diagnosed to identify ICCFs through digital subtraction angiography (DSA) and intravascular intervention. Results showed that indirect carotid cavernous fistulas were more common in females of advanced age (68.8%) and occurred spontaneously (87.5%); the most common symptoms included chemosis (93.75%), headache (87.5%), proptosis (53.1%), and tinnitus (37.5%) with type D being the most common (81.2%). Thirty out of thirty-two patients with intervention had complete flow block (83.3%), near complete (13.3%), and failure in one case. Incomplete occlusion interventions and failures were of the type D complication. The complication and number of drainage veins are related (p<0.05). ICCFs have diverse symptoms that depend on the drainage veins but most commonly were red eyes, headache, and protruding eyes. The ICCFs were mostly D-type formation and endovascular treatments were safe and effective. Common complications occurred in the case of many veins draining.

Keywords: **angiography, endovascular treatment, indirect carotid cavernous fistula.**

Classification number: **3.2**

1. Introduction

An indirect carotid cavernous fistula is a rare condition that is caused by the indirect low flow between the cavernous sinus and meningeal branches of the internal or external carotid artery or from both. This disease is common in females over 50 years old. Indirect carotid cavernous fistulas have atypical symptoms like direct carotid cavernous fistulas [1]. Symptoms depend on the flow rate and drainage vein patterns [2, 3]. Currently, there are many methods to evaluate blood vessels such as ultrasound, computed tomography of multisequence cerebral vessels, brain magnetic resonance imaging, and, especially, DSA, which is the gold standard to detect CCFs. In the past, treatment of ICCFs included symptom treatment and constriction of the carotid artery, which had low success rates and many complications. Endovascular treatments were carried out all over the world in the 1970s [4], followed by further development by other authors. Therefore, the objective of our study is to analyse clinical features, cerebral angiography, and the relationship with the outcome of endovascular treatment in patients with ICCFs.

2. Subjects and methods

Subjects: 32 patients diagnosed to identify ICCFs through DSA and intravascular intervention at Bach Mai Hospital from July 2018 to August 2019.

Methods: Prospective research. All cases meeting the selected criteria were admitted to the study to collect data on age, sex, clinical features, imaging characteristics of cerebral angiography, results of endovascular treatment, and post - intervention symptom improvement. SPSS 22.0 was used to process data.

3. Results

3.1. Demographic features

The age and sex characteristics of the research's subjects are presented in Table 1.

Table 1. The age and sex features of subjects.

Age	Male	Female	p
<50	1 (14.3%)	6 (85.7%)	0.00
≥50	3 (12%)	22 (88%)	
Mean	59.4±14.6		

3.2. Clinical features

Figure 1 showed that the most common symptoms included chemosis (93.75%), headache (87.5%), progressive vision loss (53.1%), proptosis (53.1%), and tinnitus (37.5%).

*Corresponding author: Email: ngminhchau07@gmail.com

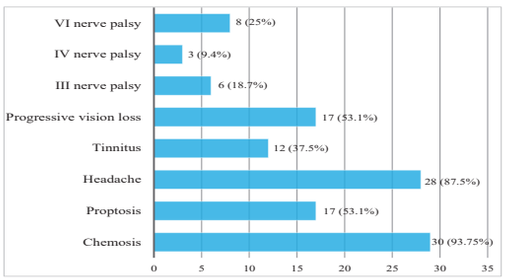


Fig. 1. Clinical features.

Imaging characteristics of cerebral angiography: In 32 cases, 7 cases had bilateral connection (21.9%), 12 cases had connection to the left cavernous sinus (37.5%), and 13 cases had connection to the right cavernous sinus (40.6%).

The features of feeding arterial branches in DSA are presented in Table 2.

Table 2. The features of feeding arterial branches in DSA.

Feeding arterial branches		Number	Rate (%)
Internal carotid artery	Right	4	12.5
	Left	5	15.6
	Bilateral	20	62.5
External carotid artery	Right	7	21.9
	Left	8	25
	Bilateral	14	43.7

In Table 2, most of cases had feeding arterial branches from both sides.

The ophthalmic vein accounted for 96.9% of draining veins, followed by the inferior petrosal sinus (34.4%) and occlusion (21.9%) (Table 3).

Table 3. Other features in DSA.

Draining veins	Number	Rate (%)
Ophthalmic vein	31	96.9
Inferior petrosal sinus	11	34.4
The cortical venous reflux	5	15.6
Contralateral drainage	6	18.7
Occlusion	7	21.9

A 60-year-old female patient was admitted due to left ptosis and tinnitus. Her DSA pictures are described in Fig. 2.

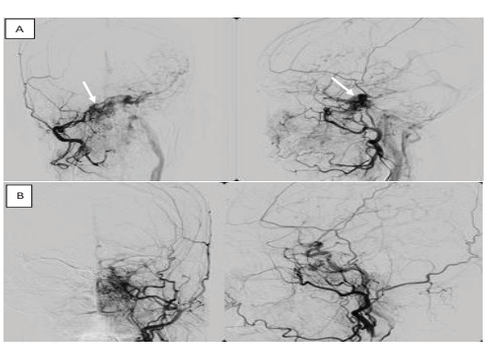


Fig. 2. DSA pictures before and after intervention. (A) ICCFs (arrow): Feeding artery from the external carotid artery bilaterally with cortical venous reflux; **(B)** After intervention: Total occlusion of the left cavernous sinus and disappearance of ICCFs.

Type D had the highest rate with 81.2% while type B and C shared the same proportion, accounting for 9.4% (Fig. 3).

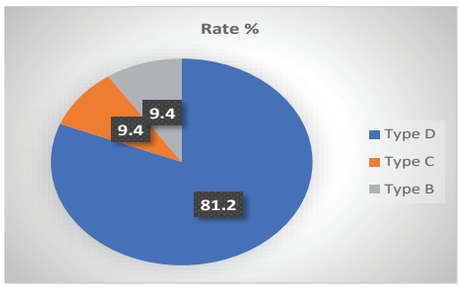


Fig. 3. Classification of flow according to Barrow.

Results of endovascular intervention

In Table 4, most patients had results of complete occlusion (83.3%). Only one case (3.4%) failed due to inaccessibility of the cavernous sinuses as the flow was too small to intervene.

Table 4. Results of endovascular intervention.

Endovascular intervention	Number	Rate (%)
Complete occlusion	25	83.3
Incomplete occlusion	4	13.3
Failed	1	3.4
Total	30	100

The ratios of complete occlusion in one side and bilateral were the same and accounted for 83.3% (Fig. 4).

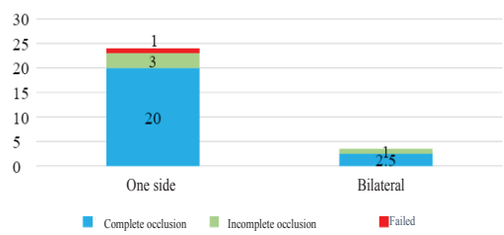


Fig. 4. The results of endovascular treatment by flow location.

Figure 5 proved that, after treatment, all symptoms decreased. Functional symptoms had a higher rate of reduction than physical signs. Chemosis reduced from 29/32 (90.6%) to 5/32 (15.6%) of cases, headache decreased from 87.5 to 12.5%, and proptosis decreased from 53.1 to 6.25% significantly with p=0.00 at a 99% confidence interval (CI).

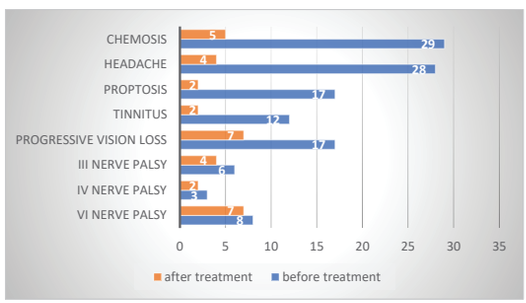


Fig. 5. Improvement of clinical symptoms after treatment.

3.4. Post - intervention complications

The most common complication after intervention was ocular palsy with 4/30 (13.33%) of cases. Three patients had VI nerve palsy, one patient had all three nerve palsies (II, IV, VI), and one patient had ischemia.

3.5. The relationship between clinical features, imaging characteristics, and results of treatment

In Table 5, symptoms of proptosis and chemosis are related to the ophthalmic venous drainage (with p=0.001 and 0.031). Tinnitus and drainage to the inferior petrosal sinus are related, and statistically significant with p=0.00, 95% confidence, OR=11.33.

Table 5. The relationship between clinical symptoms and draining veins.

Clinical symptoms		Inferior petrosal sinus		Total	p
		Yes	*No*		
Tinnitus	Yes	8 (66.7%)	4(33.3%)	12 (100%)	0.00
	No	3 (15%)	17 (85%)	20 (100%)	
	OR	11.33		2.036-63.082	
Headache	Yes	10 (35.7%)	18 (64.3%)	28 (100%)	0.573
	No	1 (25%)	3 (75%)	4 (100%)	
	OR	1.67		0.152-18.217	
Ocular palsy	Yes	5 (41.6%)	7 (58.4%)	12 (100%)	0.383
	No	6 (30%)	14 (70%)	20 (100%)	
	OR	1.67		0.347-7.424	
		Ophthalmic vein			
		Yes	No		
Proptosis	Yes	17 (100%)	0	17	0.001
	No	14 (93.3%)	1 (6.7%)	15	
Chemosis	Yes	29 (100%)	0	29	0.031
	No	2 (66.7%)	1 (33.3%)	3	
Tinnitus	Yes	11 (91.7%)	1 (8.3%)	12	0.375
	No	20 (100%)	0	20	

*p<0.05, using Fisher's Exact.

Results by types of ICCFs are provided in Table 6. All incomplete occlusion and failed cases were type D.

Table 6. The results by types of ICCFs.

	Complete occlusion	Incomplete occlusion	Failed
Type B	3	0	0
Type C	3	0	0
Type D	19	4	1
Total	25	4	1

Table 7 exhibited that there were no differences between two groups in the number of drain veins.

Table 7. The relationship between drain veins and post-intervention complication.

	Complication	No complication	Mean
The number of drain veins	4.6±0.45	1.5±0.58	2.1±1.37
p = 0.000, mean difference=3.258; 95% conference interval 2.71-3.806			

In Table 8, there was no relationship between cortical drainage veins and complications after intervention.

Table 8. The relationship between cortical drainage veins and complication.

Cortical drainage veins	Complications		Total	p
	Yes	*No*		
Yes	1 (20%)	4 (80%)	5 (100%)	0.642
No	4 (16%)	21 (84%)	25 (100%)	
Total	5	25	30	
OR	1.3		0.168-8.589	

4. Discussion

4.1. Feature of subjects

All patients were in adulthood with ages from 24 to 76 years old. The high prevalence rate of the group over 50 years old, which accounted for 78.1%, was higher than the group below 50 (p<0.05). Our results are similar to other studies around the world [5-8]. The rate of ICCFs in male and female were 12.5 and 87.5%, respectively. The female/male ratio was 7/1. The pathogenetic mechanism is unknown, but hormones may be involved [9, 10].

4.2. Clinical features

The clinical features of ICCFs are various. However, the symptoms are usually not as pronounced and typical as in direct CCFs. In our research, the most common symptoms were: chemosis (93.75%), headache (87.5%), progressive vision loss (53.1%), proptosis (53.1%), and tinnitus (37.5%). Symptoms manifesting in one eye accounted for 75% with the bilateral one being 25%. In Vietnam, chemosis was the most popular symptom. It was also the first reason for patients to see their doctor [11]. In a previous research, the rates of chemosis, tinnitus, and diplopia were 92.1, 5.3, and 2.6%, respectively [12]. Thus, clinical symptoms such as chemosis, headache, proptosis, tinnitus, vision loss, and ocular palsy were valuable suggestions for detecting the pathology of ICFFs.

4.3. Imaging characteristics of cerebral angiography

DSA was selected to diagnose cavernous carotid fistulas in all cases. This procedure helps to categorize lesions and choose the right treatment. The feeding arteries from the internal carotid artery in 20 cases was bilateral (62.5%), followed by the left side (15.6%) and the right side (12.5%). The feeding arteries from the external carotid artery in 14 cases were bilateral (43.7%), followed by the left side (25%) and the right side (21.9%). Thus, the majority of patients had the bilateral feeding arteries.

The characteristics of flow: 7/32 cases were bilateral (21.9%), right side (40.6%), and left side (37.5%). There were no differences between the three types.

The drain veins: 31/32 were from the ophthalmic vein (96.9%) and 11/32 from the inferior petrosal sinus (34.4%). The proportion of patients with cortical venous reflux on DSA was 15.6%, with contralateral drainage was 18.7%, and with occlusions in drain veins was 21.9%.

The location of the drainage veins was consistent with clinical symptoms. The most prevalent symptom was chemosis (93.75%) and DSA also showed that the majority of patients had venous drainage (96.9%). There was an association between tinnitus symptoms and inferior petrosal sinus drainage (p<0.05). In the study of M.D.

Alexander, et al. (2019) [13], the inferior petrosal sinus was observed and varied among patients (p=0.005). Therefore, we consider that the type of drainage vein will determine a patient's clinical symptoms.

Characteristics of classification of flow by D.L. Barrow, et al. (1985) [14]: Type D had the highest percentage with 26 patients (81.2%), which was followed by types B and C with equal rates (9.4%).

4.4. The relation between clinical features, imaging characteristics, and results of treatment

In our study, all 32 patients had DSA performed to diagnose ICCFs. There were two cases of type D that had no indication for intervention because of too small flow and insignificant ophthalmic vein. Thirty patients received intervention and 25/30 of said patients had complete occlusion (83.3%). One case of intervention failure was had right cavernous sinus flow, type D, ophthalmic vein, and inferior petrosal sinus obstruction. The intervention was conducted through the external carotid vein into the facial vein. It was failed when accessing to the ophthalmic vein due to small aperture, zigzag veins, and too small flow. According to some authors, for ICCFs with a low flow rate, the preferred treatment option is gamma knife radiotherapy. This is a minimally invasive method and is used when endovascular or surgical approaches are too dangerous or have failed [15-17]. The rate of complete occlusion in one side was 20/25 and bilateral was 5/6 (83.3%).

Complications after intervention were seen in five patients and accounted for 16.7%. To be specific, there were three cases of VI nerve palsy (6.7%), one case of ischemia (3.3%), and one case of ocular immobilization accompanied by an artery tear that formed aneurysm in the common femoral artery (3.3%). There were no cases of recent recurrence within a week, nor death. Sixth nerve palsy was treated by Solu Medrol and the patient recovered. In the case of the thigh aneurysm, emergency surgery to repair the artery was conducted.

Incomplete occlusion and failed cases were type D, possibly due to the type D flux received by the feeding arteries from both internal and external carotid arteries, thus the source of leakage was greater and the flow was higher.

The mean value of differences in the number of drainage veins of two groups with complication and no complications were 3.258.

The CI of the difference was 2.71-3.806 ($p<0.05$) and there was no significant difference. This result is similar to K.H. Jung, et al. (2011) [18] study with p=0.011, OR=2.189, and 95% CI from 1.198 to 3.999. The mechanism is unknown, but as can be seen in cases of multiple draining veins, when the intervention abruptly obstructs them decompensation can occur and may contribute to worsening cases.

5. Conclusions

ICFFs are common in females over 50 years old (68.8%). Clinical symptoms are varied, atypical, and depend on the pattern of draining veins with the most common symptoms including chemosis (93.75%), headache (87.5%), progressive vision loss (53.1%), proptosis (53.1%), and tinnitus (37.5%). Type D was the most common (81.2%). Thirty out of thirty-two patients with intervention had complete occlusion (83.3%), incomplete occlusion (13.3%), and failure in one case. Clinical symptoms related to drainage flow to the ophthalmic veins were proptosis ($p<0.05$) and chemosis ($p<0.05$). The occurrence of drainage to the inferior petrosal sinus increased 11.33 times when patients had tinnitus. The cases of incomplete occlusion and failure were type D, and complications were related to the number of drain veins ($p<0.05$).

CRediT author statement

Van Tuan Nguyen, Anh Tuan Tran, Thi Hoang Yen Vu: Data curation, Software, Visualisation, Investigation, Writing - Original draft preparation, Conceptualization, Methodology, Supervision.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] A.J. Ringer, L. Salud, T.A. Tomsick (2005), "Carotid cavernous fistulas: Anatomy, classification, and treatment", *Neurosurg. Clin.*, **16(2)**, pp.279-295, DOI: 10.1016/j.nec.2004.08.004.

[2] N.R. Miller (2007), "Diagnosis and management of dural carotid-cavernous sinus fistulas", *Journal of Neurosurgery*, **23(5)**, DOI: 10.3171/FOC-07/11/E13.

[3] H.S. Kalish, A. Setton, Y. Nimii, et al. (2002), "Cavernous sinus dural arteriovenous malformations: Patterns of venous drainage are related to clinical signs and symptoms", *Ophthalmology*, **109(9)**, pp.1685-1691, DOI: 10.1016/s0161-6420(02)01166-1.

[4] F.A. Serbinenko (1974), "Balloon catheterization and occlusion of major cerebral vessels", *J. Neurosurg.*, **41(2)**, pp.125-145, DOI: 10.3171/jns.1974.41.2.0125.

[5] M.K.S. Heran, D. Volders, C. Haw, et al. (2019), "Imaging-guided superior ophthalmic vein access for embolization of dural carotid cavernous fistulas: Report of 20 cases and review of the literature", *Am. J. Neuroradiol.*, **40(4)**, pp.699-702, DOI: 10.3174/ajnr.A5994.

[6] D.J. Kim, D.I. Kim, S.H. Suh, et al. (2006), "Results of transvenous embolization of cavernous dural arteriovenous fistula: A single-center experience with emphasis on complications and management", *Am. J. Neuroradiol.*, **27(10)**, pp.2078-2082.

[7] M. Théaudin, J.P.S. Maurice, R. Chapot, et al. (2007), "Diagnosis and treatment of dural carotid-cavernous fistulas: A consecutive series of 27 patients", *J. Neurol. Neurosurg. Psychiatry*, **78(2)**, pp.174-179, DOI: 10.1136/jnnp.2006.100776.

[8] K. Yoshida, M. Melake, H. Oishi, et al. (2010), "Transvenous embolization of dural carotid cavernous fistulas: A series of 44 consecutive patients", *Am. J. Neuroradiol.*, **31(4)**, pp.651-655, DOI: 10.3174/ajnr.A1882.

[9] R.M. Taniguchi, J.A. Goree, G.L. Odom (1971), "Spontaneous carotid-cavernous shunts presenting diagnostic problems", *J. Neurosurg.*, **35(4)**, pp.384-391, DOI: 10.3171/jns.1971.35.4.0384.

[10] S. Toya, R. Shiobara, J. Izumi, et al. (1981), "Spontaneous carotid-cavernous fistula during pregnancy or in the postpartum stage: Report of two cases", *J. Neurosurg.*, **54(2)**, pp.252-256, DOI: 10.3171/jns.1981.54.2.0252.

[11] V.G. Nguyen, M.T. Le (2011), "Clinical evaluation of ophthalmic outcomes of embolization intervention in carotid-cavernous fistula", *Ho Chi Minh City Journal of Medicine*, **15(1)**, pp.106-112 (in Vietnamese).

[12] D.C. Nguyen, C.C. Tran (2014), "Evaluating effectiveness of cavernous sinus dural arteriovenous fistula treatment by transvenous embolisation", *Ho Chi Minh City Journal of Medicine*, **18(6)**, pp.433-440 (in Vietnamese).

[13] M.D. Alexander, V.V. Halbach, D.K. Hallam, et al. (2019), "Relationship of clinical presentation and angiographic findings in patients with indirect cavernous carotid fistulae", *J. NeuroInterventional Surg.*, **11**, pp.937-939, DOI: 10.1136/neurintsurg-2018-014421.

[14] D.L. Barrow, R.H. Spector, I.F. Braun, et al. (1985), "Classification and treatment of spontaneous carotid-cavernous sinus fistulas", *J. Neurosurg.*, **62(2)**, pp.248-256, DOI: 10.3171/jns.1985.62.2.0248.

[15] H. Baharvahdat, Y.C. Ooi, W.J. Kim, et al. (2020), "Updates in the management of cranial dural arteriovenous fistula", *Stroke Vasc. Neurol.*, **5(1)**, pp.50-58, DOI: 10.1136/svn-2019-000269.

[16] C.J. Chen, C.C. Lee, D. Ding, et al. (2015), "Stereotactic radiosurgery for intracranial dural arteriovenous fistulas: A systematic review", *J. Neurosur.*, **122(2)**, pp.353-362, DOI: 10.3171/2014.10.JNS14871.

[17] T.R. Miller, D. Gandhi (2015), "Intracranial dural arteriovenous fistulae: Clinical presentation and management strategies", *Stroke*, **46(7)**, pp.2017-2025, DOI: 10.1161/STROKEAHA.115.008228.

[18] K.H. Jung, B.J. Kwon, K. Chu, et al. (2011), "Clinical and angiographic factors related to the prognosis of cavernous sinus dural arteriovenous fistula", *Neuroradiology*, **53(12)**, pp.983-992, DOI: 10.1007/s00234-010-0805-3.

DOI: 10.31276/VJSTE.64(4).55-59

Lateral transperitoneal laparoscopic management for paragangliomas in para-aortic position: Surgical experience from case reports of three patients

Ngoc Hung Nguyen[1], Huy Du Nguyen[2], Hai Dang Do[2], Tuan Hiep Luong[2*], Thanh Khiem Nguyen[1]

[1]Gastrointestinal and Hepato-Biliary-Pancreatic Surgery Department, Bach Mai Hospital, 78 Giai Phong Street, Phuong Mai Ward, Dong Da District, Hanoi, Vietnam
[2]Department of Surgery, Hanoi Medical University, 1 Ton That Tung Street, Trung Tu Ward, Dong Da District, Hanoi, Vietnam

Received 14 January 2022; revised 21 March 2022; accepted 4 April 2022

Abstract:

Paragangliomas are a type of neuroendocrine tumour with catecholamine secretion outside of the adrenal medulla. These tumours are frequently found in the para-aortic region, which makes them difficult to manage laparoscopically. We herein present a case report of three patients with paragangliomas in a difficult superior para-aortic position. All cases had symptoms of catecholamine release with uncontrolled episodic hypertension and had their tumours successfully removed by transperitoneal laparoscopic management. We have found this method to be a suitable approach because it reveals a wide surgical field and a view over all organs, especially blood vessels around the tumour. Although blood and urine dopamine and noradrenaline levels were not measured postoperatively, all three patients were discharged without complication and they did not display any more clinical symptoms. Lateral transperitoneal laparoscopic surgery provided a wide field of view, which clearly revealed the tumour and surrounding components indicating it is one of the safest and most effective approaches, especially for paragangliomas adjacent to large blood vessels.

Keywords: **case report, laparoscopic, lateral transperitoneal, paragangliomas.**

Classification number: **3.2**

1. Introduction

Paraganglioma is a rare type of neuroendocrine tumour with an incidence of 0.0005-0.1% of the population and is most commonly found during the third through fifth decades of life in women [1, 2]. Tumours located outside the adrenal glands mainly secrete catecholamines and may have an origin of adrenal medulla (pheochromocytoma) or not (paraganglioma). Genetic factors account for about 25% of cases and most commonly in bilateral tumours [3]. The specific clinical symptoms, which are called fight-or-flight sympathetic symptoms, occur due to increased release of catecholamines including episodic hypertension, heart palpitations, headache, fatigue, and profuse sweating. Of these symptoms, the triad of profuse sweating, heart palpitations, and headache has a sensitive rate of 89% and a specific rate of 67%, which is close to 95% in patients with hypertension [4]. Most of the time, however, this tumour is often discovered incidentally by diagnostic imaging. These tumours are mostly benign, with only a small percentage of malignancies and metastases, and often insensitive to chemotherapy or radiotherapy. Surgical resection is thus the optimal treatment option [3]. Recently, with the development of minimally invasive techniques, laparoscopic surgery has become a new choice due to the advantages of decreased surgical time and hospital stay as well as a reduced postoperative complication rate. However, because the tumour is often located between the aorta and vena cava, laparoscopic surgery remains a challenge for surgeons [3]. Herein, we report successful laparoscopic surgeries used to treat three patients with paraganglioma located in relatively difficult locations. Our research objectives are aimed at sharing our approach and experiences with paragangliomas in the para-aortic region as well as reviewing the literature on the outcomes of laparoscopic surgery for paraganglionic tumours.

2. Case presentation

2.1. Surgical procedure

In this study, we placed the patient in the lateral position using the intra-abdominal laparoscopic technique.

For tumours located adjacent to or in between the abdominal aorta (AA) and inferior vena cava (IVC), the patient is placed in the left lateral position and the 4 trocars are placed, respectively, as follows: two 10-cm trocars above the umbilical and suprapubic and two 5-cm trocars in the right lumbar region and right iliac region. The patient underwent surgery according to the following procedure: fasciectomy, ascending colon and

*Corresponding author: Email: hiep1995hsgs@gmail.com

mesenteric mobility, descending duodenal dissection, inferior duodenal flexure, and horizontal duodenal dissection (if necessary). Then, reveal and resect the paraganglioma.

For tumours located between the left renal artery and the left renal vein, we had chosen the placement of 4 trocars as follows: two 10-mm trocars were placed just above the umbilical and below the left sternum, and two 5-mm trocars were placed in the right hypochondrium and the left anterior axillary line, respectively. Our surgical procedure was as follows: mobilise the descending colon, splenic flexure, distal pancreas, and spleen; then expose the left renal peduncle. The tumour was found between the renal artery and the renal vein, and a part of the tumour was attached to the left renal vein. Then, dissection was conducted, the tumour dissected, and the infiltrated left renal vein was sutured. In these cases, we chose to place the trocar and the approach was similar to the left adrenalectomy. We successfully performed this procedure in three cases.

2.2. Case 1

A male patient, 17 years old, with a history of glomerulonephritis and an appendectomy at 12 years old. The patient was admitted to the hospital because of paroxysmal hypertensive crises 3 months before admission with each episode lasting 5-6 minutes with sweating. The highest blood pressure recorded was 220 mmHg with generalised seizures one day before admission. Physical examination was unremarkable except for a few brief episodes of hypertension during the day. Computed tomography (CT) scans showed a retroperitoneal tumour in front of the AA, at the level of the inferior mesenteric artery, 48x28 mm in size and was clearly demarcated, hyper-enhanced, and possessed central necrosis (Fig. 1). Blood tests showed increased dopamine (328.5 pg/ml) and increased noradrenaline (1177.71 pg/ml) with slightly increased aldosterone, normal cortisol and adrenocorticotropic hormone (ACTH), and normal blood potassium concentration. The patient was then diagnosed with paraganglioma. Intraoperative exploration: The abdomen is dry, the peritoneum is smooth, and there is a tumour measuring 4x5 cm close to the IVC and the AA without invasion. Tumour resection was performed according to the following procedure: The patient lies on the left side with 4 trocars placed, respectively, as follows: two 10-cm trocars above the umbilical and suprapubic and two 5-cm trocars in right lumbar region and right iliac region. The patient underwent surgery according to the following procedure: fasciectomy, ascending colon and mesenteric mobility, extensive exposure of the retroperitoneal space including the aorta, and tumour, meticulous dissection of the tumour from the aorta and IVC. Surgery time was 80 minutes and blood loss was less than 30 ml. The pathological result was pheochromocytoma with pheochromocytoma of the adrenal gland scaled score (PASS) of 5 points. Postoperatively, the patient was stable and discharged without any complications.

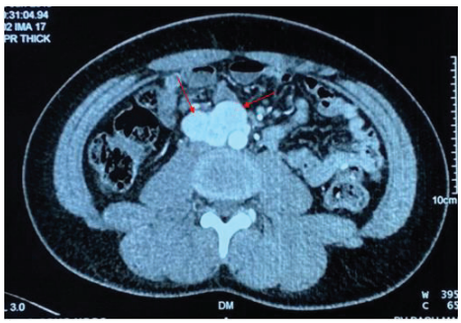


Fig. 1. CT scan showed a retroperitoneal tumour in front of the AA, at the level of the inferior mesenteric artery, with 48x28 mm in size, clearly demarcated, hyper-enhancing, with a central necrosis (red arrows).

2.3. Case 2

A male patient, 47 years old, with a history of episodes of paroxysmal hypertensive crisis peaking at 200/150 mmHg over five years but no intervention. In the recent year, the hypertensive crises began to appear more and more severe with an average of 10-15 attacks per day, each attack lasting about one minute, and the highest blood pressure was 275/135 mmHg. Magnetic resonance imaging (MRI) and Positron emission tomography and computed tomography (PET-CT) scans show a lymph-node-liked structure of the 3rd and 4th level of the lumbar spine with a size of 23x17 mm, clearly demarcated, and located in between the IVC and the AA (Fig. 2). Blood levels of dopamine, noradrenalin, aldosterone, and cortisol were 112.3 pg/ml, 686 pg/ml, 126 pg/ml (high), and 334.3 nmol/l (normal), respectively. Urine catecholamine test presented a urine dopamine level of 68.1 µg in 24 hours, and increased noradrenaline and adrenaline of 113.3 and 142.2 µg, respectively, in 24 hours. Location of trocar placement and the stages of surgery were similar to the first case. Operating time was 75 minutes and the blood loss was less than 30 ml. The pathological result was paraganglioma with raft-forming cells with large, irregular nuclei. The postoperative process was normal.

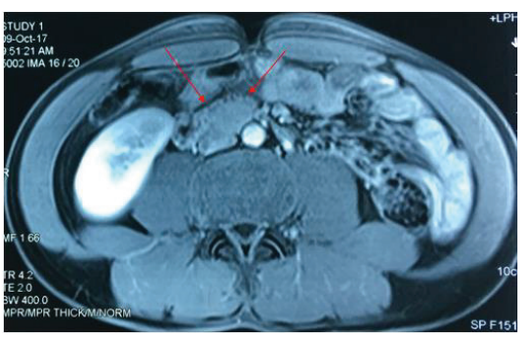


Fig. 2. The MRI scans show the lymph-node-liked structure of the lumbar spine 3rd and 4th level with the size of 23x17 mm, located between the inferior vena cava and the AA, clearly demarcated (red arrows).

2.4. Case 3

A female patient, 16 years old, with a history of stable polycythaemia vera was admitted to our hospital. The patient presented with intermittent dyspnoea, dizziness, and heart palpitations with the highest blood pressure of 220 mmHg. On the CT scan, two vascular-enhanced masses are seen at the level of the renal hilum and the right and left dimensions were 29 and 15 mm, respectively (Fig. 3). Blood tests showed a normal dopamine level of 84.83 pg/ml, a high noradrenaline level of 1153 pg/ml, a normal adrenaline level of 27.67 pg/ml, and pituitary hormones were within normal limits. After stabilizing medical treatment, the patient underwent surgery in two stages: the first stage was on April 13th, 2021, and the second was on April 19th, 2021. Unlike the above two cases, the tumour was located close to the renal artery and vein, and a part of the tumour was attached to the posterior surface of the renal vein. First, after releasing the splenic flexure and distal pancreas, the surgeon exposed the left renal peduncle, wiggled the renal vein and renal artery, and sutured the infiltrated vein (Fig. 4). Duration of the first surgery was 180 minutes. In the second surgery, a tumour in the renal hilum measuring 2x3 cm with solid density and firmly attached to the posterior genital vein was revealed. After dissecting and releasing the ascending colon, exposing the duodenum, and dissecting the tumour from the posterior genital vein, the tumour was resected. The operation time was 150 minutes and the patient was monitored at the Department of Endocrinology with stable condition after surgery. Postoperative pathological result was paraganglioma.

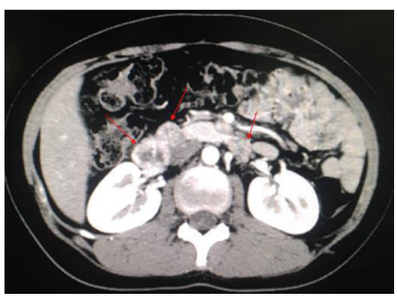


Fig. 3. On the CT scan, two vascular-enhanced masses are seen at the level of the renal hilum, the right and left dimensions are 29 and 15 mm, respectively (red arrows).

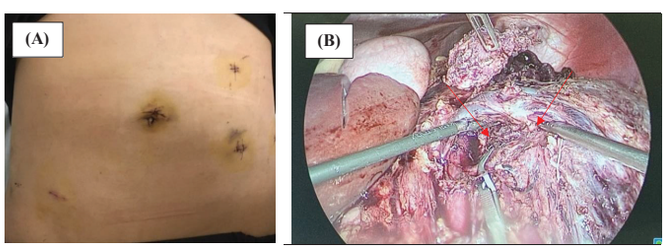


Fig. 4. (A) Trocars' placements; (B) Intraoperative injury: A tumour in the renal hilum measuring 2x3 cm, solid density, firmly attached to the posterior genital vein (black arrows).

Treatment of paraganglioma is a complicated process, and the surgeon also needs to consult and closely coordinate with many other specialties such as endocrinology, cardiology, and anaesthesiology to develop an appropriate strategy. All patients were consulted by an endocrinologist for a definitive diagnosis as well as a cardiologist and anaesthetist to maintain blood pressure pre-, intra-, and post-operatively.

3. Discussion

Paragangliomas were first described by Pick in 1912 and first operated on in 1937 by Charles Mayo [3]. Through the three reported cases, we found that the most specific symptom was an episodic hypertensive crisis due to increased secretion of catecholamine. About 50-60% of cases will present with headache, heart palpitations, and profuse sweating, and a paroxysmal hypertensive crisis. Depending on the type of tumour, the characteristics of hypertension are different. Of these characteristics, noradrenaline-secreting tumours are often associated with prolonged hypertension, while episodic hypertensive crises are common in both noradrenaline and adrenaline-secreting groups. The estimated malignancy rate is about 15% and the 5-year survival rate ranges from 40 to 85% [5, 6] with paragangliomas having a higher malignancy potential than pheochromocytomas [6]. Although the risk factors for malignancy have not been well identified, some factors predisposed to malignancy are tumour size >5 cm, genetic mutations, failed resection of primary tumour, and presence of synchronous metastases [7]. Therefore, it is necessary to follow-up with patients postoperatively to determine if there is recurrence [8].

Radical resection remains the optimal treatment for paragangliomas. There are different treatment methods such as open surgery, laparoscopic lateral transperitoneal adrenalectomy (LTA), and laparoscopic posterior retroperitoneal adrenalectomy (PRA) [9, 10]. Based on the criteria of surgical time, intraoperative blood loss, pain level, hospital stay, and time to return to normal life, as well as potential complications, most studies have shown that PRA has the advantage. However, recent studies show an equivalent role of these two methods in the treatment of adrenal tumours [11]. Currently, indications for laparoscopic surgery are applied in cases of hormone-secreting tumours such as paraganglioma and other equivalents with tumour size less than 7 cm. For tumour sizes over 7 cm, with highly qualified surgeons capable of endoscopic dissection of related components such as Kocher duodenum endoscopy, colonic motility, etc., laparoscopic management is still possible [12]. Z.A.R. Jawad, et al. (2017) [13] reported a successful laparoscopic surgery to remove an 8.2-cm paraganglioma.

In this study, we placed the patient in a supine position using the intra-abdominal laparoscopic technique because

of the short hospital stay and high efficiency compared to other methods [14]. Our trocar placement and approach were different from that of X. Ren, et al. (2020) [12]. These authors used four trocars with corresponding positions as follows: 10-mm trocar above the umbilicus, another 10-mm trocar in the epigastric region, one 5-mm trocar in the right hypochondrium, and one 5-mm trocar in the right anterior axillary line. The author's approach to the tumour consisted of the following steps: first, the right hepatic triangular ligament and the hepato-colic ligament were cut, the liver was lifted to expose the upper pole of the right kidney. Then, the mesentery of the ascending colon was dissected and extruded to expose the right kidney and right renal vein followed by dissecting sections D2 and D3 of the duodenum to reveal the tumour. Then, the vessels feeding the tumour were clamped and the tumour resected. Our tumour approach provided a wide field of view, which clearly revealed the tumour and surrounding components, especially the adjacent large blood vessels. In addition, it was possible to fully observe the abdominal viscera to investigate any other abnormalities. This approach is especially beneficial in cases of large tumours and those with vascular invasion. The operating time of our two cases were 80 and 75 minutes, respectively, with very little blood loss of just under 30 ml. This result was better than of [12], which took 120 minutes and 50 ml of blood in surgery. Our patients were in stable condition and did not have any postoperative complications. However, when performing this technique, we also found that there were some difficulties when we had to move many bowel segments such as the ascending colon, the mesentery, and a part of the duodenum to reach the tumour. The risk of injury to intra-abdominal organs is also higher than in the retroperitoneal approach.

For tumours located between the left renal artery and left renal vein, we chose a placement of 4 trocars as follows: two 10-mm trocars were placed just above the umbilical and below the left sternum, and two 5-mm trocars were placed in the right hypochondrium and the left anterior axillary line, respectively. We have found this to be a very suitable approach because it reveals a wide surgical field, which allowed an evaluation of all the organs and especially the blood vessels around the tumour. In our case, the tumour also invaded the left renal vein. Therefore, a wide and clear disclosure would make it easier and safer to remove the tumour and suture the invasive blood vessel. As a result, the intraoperative blood loss, in this case, was 140 ml and the surgery time was 180 minutes. In a study by S. Hattori, et al. (2014) [15] who performed surgery on 9 cases of paragangliomas, the operation time was 189.8±44.9 minutes and the intraoperative blood loss was about 404.9 to 1036.3 ml. It can be seen that our surgery time was similar, but we had a much lower amount of blood loss intraoperatively compared to this study. However, choosing this approach will make it difficult for the surgeon to reveal the tumour when moving the colon, spleen, and especially the distal pancreas. All manipulations have the risk of damaging organs and causing complications intra- and post-operatively.

Paraganglioma often occurs around the great vessels in the retroperitoneum and the location of the tumour significantly affects the difficulty of procedure. T. Hakariy, et al. (2019) [16] reported a case of paraganglioma located posterior to the IVC and bilateral renal veins. In that study, five trocars were placed in the abdomen as follows: one 10-mm trocar above the umbilicus and proximal to the right white line for camera placement, two 12-mm trocars placed below the right costal margin deviated to the sternal and in the longitudinal side of the right white line with the level of the umbilicus, and two 5-mm trocars placed in the right anterior axillary line, respectively. The process of approaching the tumour: dissection of the tumour from the underside of the liver, dissection of the tumour from the right renal vein, left renal vein, right kidney, and AA. It was identified that there were two veins going from the tumour that drained into the IVC, then resection was performed. The tumour was separated from the AA and the two arteries supplying the tumour were cut off; then, dissection of the tumour from the renal artery continued. Finally, the tumour was released from the IVC and completely removed from the surrounding surgery. In this case, the most difficult and dangerous step in the surgical procedure was dissecting the tumour from the IVC. Therefore, in order to safely release it, the author dissected tissue surrounding the tumour and achieved tumour mobility before separating the tumour from the IVC. In addition, the tumour was located between the right renal vein and the right renal artery. These vessels were all dissected from the tumour intact. The operation time was 231 minutes, and the blood loss was about 200 ml. The tumour was located behind many large blood vessels, and, although many reports suggest that the retroperitoneal approach is feasible, safer, and faster, the authors chose the transperitoneal approach instead of the retroperitoneal approach because it can provide a wide surgical field and make it easy to realise the anatomical relationship between the tumour and the surrounding components in the abdominal cavity. If necessary, this approach can be converted to open surgery.

It is important to control blood pressure pre-, peri- and post-operatively. Patients should be given anti-alpha- or beta-adrenergic drugs in preoperative preparation [17]. One note to the surgeons is to limit the impact on the tumour hemodynamically by ligating the adrenal vein early. According to N. Rao, et al. (2016) [18], the rate of using antihypertensive drugs after surgery depends on the number of episodes of hypertension and the degree of impact on the adrenal gland during surgery. Anaesthesiologists need to coordinate with surgeons to use invasive arterial blood pressure or antihypertensive drugs during the procedure. Intraoperative hemodynamic instabilities typically include hypertension before tumour removal and hypotension after tumour isolation. Because of this, the American Society of Anaesthesiologists recommends using central venous access and invasive blood pressure (IBP) monitoring. Hypertension and bradycardia/

tachycardia due to norepinephrine (NE) secretion should be managed with short-acting and potent vasodilators such as sodium nitroprusside and nitroglycerine or esmolol, a short-acting beta-receptor antagonist [17]. After ligation of the adrenal veins and tumour removal, many patients may experience a sudden drop in blood pressure (hypotension) and may require vasopressor support. Indeed, treatment of paraganglioma is a complicated process and it is imperative that the surgeon consult and closely coordinate with many specialties such as endocrinology, cardiology, and anaesthesiology to form an appropriate strategy [19].

4. Conclusions

Paraganglioma is a rare disease that is located frequently in para-aortic region, which makes it difficult to manage laparoscopically. However, lateral transperitoneal laparoscopic surgery is a safe and effective treatment, especially for tumours in difficult locations. Our tumour approach provides a wide field of view, which clearly reveals the tumour and surrounding components, especially adjacent large blood vessels. In addition, full observation of the abdominal viscera is another advantage of this method, which can be used to investigate any other abnormalities in the region. The treatment of paraganglioma requires the coordination of endocrinologists, cardiologists, surgeons, and anaesthesiologists for the most appropriate treatment strategy and follow-up.

CRediT author statement

Ngoc Hung Nguyen, Huy Du Nguyen, Hai Dang Do, Tuan Hiep Luong, Thanh Khiem Nguyen: Data curation, Software, Visualisation, Investigation, Writing original draft preparation, Conceptualization, Methodology.

AKNOWLEDGEMENTS

The authors would like to thank the Board of the Gastrointestinal and Hepato-Biliary-Pancreatic Surgery Department, Bach Mai Hospital for their assistance during the time in-hospital observation of this patient.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] E.A. Mittendorf, W.M Linehan, G. Eisenhofer, et al. (2007), "Pheochromocytoma: Advances in genetics, diagnosis, localization, and treatment", *Hematology/Oncology Clinics of North America*, **21(3)**, pp.509-525, DOI: 10.7326/0003-4819-134-4-200102200-00016.

[2] L.C. Pederson, J.E. Lee (2003), "Pheochromocytoma", *Current Treatment Options in Oncology*, **4(4)**, pp.329-337, DOI: 10.1007/s11864-003-0008-9.

[3] H.P.H. Neumann, W.F. Young, C. Eng (2019), "Pheochromocytoma and Paraganglioma", *The New England Journal of Medicine*, **381(6)**, pp.552-565, DOI: 10.1056/NEJMra1806651.

[4] F.A. Farrugia, G. Martikos, P. Tzanetis, et al. (2017), "Pheochromocytoma, diagnosis and treatment: Review of the literature", *Endocrine Regulations*, **51(3)**, pp.168-181, DOI: 10.1515/enr-2017-0018.

[5] S. Hescot, M.C. Freixes, T. Deutschbein, et al. (2019), "Prognosis of malignant pheochromocytoma and paraganglioma (MAPP-Prono Study): A European network for the study of adrenal tumours retrospective study", *The Journal of Clinical Endocrinology Metabolism*, **104(6)**, pp.2367-2374, DOI: 10.1210/jc.2018-01968.

[6] E.P.M. Corssmit, M. Snel, E. Kapiteijn (2020), "Malignant pheochromocytoma and paraganglioma: Management options", *Current Opinion in Oncology*, **32(1)**, pp.20-26, DOI: 10.1097/CCO.0000000000000589.

[7] P. Goffredo, J.A. Sosa, S.A. Roman (2013), "Malignant pheochromocytoma and paraganglioma: A population level analysis of long-term survival over two decades", *Journal of Surgical Oncology*, **107(6)**, pp.659-664, DOI: 10.1002/jso.23297.

[8] P.F. Plouin, L. Amar, O.M. Dekkers, et al. (2016), "European Society of Endocrinology Clinical Practice Guideline for long-term follow-up of patients operated on for a phaeochromocytoma or a paraganglioma", *European Journal of Endocrinology*, **174(5)**, pp.1-10, DOI: 10.1530/EJE-16-0033.

[9] Y.J. Chai, H. Kwon, H.W. Yu, et al. (2014), "Systematic review of surgical approaches for adrenal tumours: Lateral transperitoneal versus posterior retroperitoneal and laparoscopic versus robotic adrenalectomy", *International Journal of Endocrinology*, **2014**, DOI: 10.1155/2014/918346.

[10] J.W.M. Lenders, Q.Y. Duh, G. Eisenhofer, et al. (2014), "Pheochromocytoma and paraganglioma: An endocrine society clinical practice guideline", *The Journal of Clinical Endocrinology & Metabolism*, **99(6)**, pp.1915-1942, DOI: 10.1210/jc.2014-1498.

[11] A. Balla, M. Ortenzi, L. Palmieri, et al. (2019), "Laparoscopic bilateral anterior transperitoneal adrenalectomy: 24 years experience", *Surgical Endoscopy*, **33(11)**, pp.3718-3724, DOI: 10.1007/s00464-019-06665-6.

[12] X. Ren, J. Shang, R. Ren, et al. (2020), "Laparoscopic resection of a large clinically silent paraganglioma at the organ of Zuckerkandl: A rare case report and review of the literature", *BMC Urology*, **20(1)**, DOI: 10.1186/s12894-020-00732-0.

[13] Z.A.R. Jawad, A.B.F. Puerta, D. Lefroy, et al. (2017), "Complete laparoscopic excision of a giant retroperitoneal paraganglioma", *Annals of The Royal College of Surgeons of England*, **99(5)**, pp.148-150, DOI: 10.1308/rcsann.2017.0056.

[14] C. Aggeli, A.M. Nixon, C. Parianos, et al. (2017), "Surgery for pheochromocytoma: A 20-year experience of a single institution", *Hormones (Athens, Greece)*, **16(4)**, pp.388-395, DOI: 10.14310/horm.2002.1759.

[15] S. Hattori, A. Miyajima, Y. Hirasawa, et al. (2014), "Surgical outcome of laparoscopic surgery, including laparoendoscopic single-site surgery, for retroperitoneal paraganglioma compared with adrenal pheochromocytoma", *Journal of Endourology*, **28(6)**, pp.686-692, DOI: 10.1089/end.2013.0706.

[16] T. Hakariya, Y. Shida, H. Ito, et al. (2019), "Successful laparoscopic resection of a paraganglioma immediately behind the inferior vena cava and bilateral renal veins", *IJU Case Reports*, **2(5)**, pp.261-264, DOI: 10.1002/iju5.12097.

[17] R. Ramachandran, V. Rewari (2017), "Current perioperative management of pheochromocytomas", *Indian Journal of Urology*, **33(1)**, pp.19-25, DOI: 10.4103/0970-1591.194781.

[18] N. Rao, R. Ramachandran, N. Tandon, et al. (2016), "Surgical and hemodynamic outcomes in pheochromocytoma surgery: A prospective cohort study", *Urology*, **98**, pp.103-106, DOI: 10.1016/j.urology.2016.09.004.

[19] G. Janetschek, H.P. Neumann (2001), "Laparoscopic surgery for pheochromocytoma", *The Urologic Clinics of North America*, **28(1)**, pp.97-105, DOI: 10.1097/01.sla.0000164175.26785.06.

DOI: 10.31276/VJSTE.64(4).60-63

In-vitro synthesis of RNA fragments specifying envelope protein gene and *RdRp* gene of SARS-CoV-2 as positive standards for molecular diagnosis tests

Hang Thi Thu Dinh[1*], Mai Ngoc Nguyen[1, 2], Su Xuan Hoang[1]

[1]*Institute of Biomedicine and Pharmacy, Vietnam Military Medical University, 160 Phung Hung Street, Phuc La Ward, Ha Dong District, Hanoi, Vietnam*

[2]*University of Science, Vietnam National University - Hanoi, 334 Nguyen Trai Street, Thanh Xuan Trung Ward, Thanh Xuan District, Hanoi, Vietnam*

Received 22 July 2021; revised 7 October 2021; accepted 19 October 2021

Abstract:

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of Coronavirus Disease 2019 (COVID-19), has rapidly spread through the entire world and has become the worst pandemic from December 2019 until now. The establishment of positive standards for molecular diagnostic testing for SARS-CoV-2 plays a critical role in development and assessment of diagnostic assays associated with the shortage of positive specimens and viral culture fluids. This study aims to establish a novel assay for *in-vitro* RNA fragment synthesis based on amplicons of self-priming PCR targeting envelope protein gene and RNA-dependent RNA polymerase (*RdRp*) gene of SARS-CoV-2. The cDNA library of the targeted genes of SARS-CoV-2 was generated by using long-primers named sMn1 forward/reverse primer (*E* gene) and sMn2 forward/reverse primer (*RdRp* gene) for self-priming PCR assays. The synthesised amplicons that overlap the target sequence of the World Health Organization (WHO) assay were cloned into a pGEM-T easy vector, then transformed into *E. coli* competent cells by conventional methods. The recombinant plasmids were used as materials for *in-vitro* RNA transcription. Concentrations of the *in-vitro* transcribed RNA were 200-800 ng/µl with A260/A280 ratios of 2.0-2.2. Gel electrophoresis showed a single band of each RNA molecule with sizes of 216 and 214 bases for sMn1- E and sMn2- *RdRp* gene, respectively. Furthermore, we effectively evaluated the *in-vitro* transcribed RNA by a one-step, real-time RT-PCR assay according to the standard WHO protocol. The stability of in-vitro RNA over a 6-month storage period was then investigated. In conclusion, our assay for in-vitro synthesis of RNA fragments transcribed from self-priming amplicons were successfully established and thus these positive standards were useful for molecular diagnosing of SARS-CoV-2.

Keywords: ***in-vitro*** **transcribed RNA, one-step real-time RT-PCR, SARS-CoV-2, self-priming.**

Classification number: **3.2**

1. Introduction

COVID-19 is a severe acute respiratory syndrome (SARS) caused by a novel strain of coronavirus named SARS-CoV-2. This virus has been reported to be a highly transmissible and pathogenic coronavirus that arose in late 2019 from Wuhan, China [1] and has sparked a pandemic, posing a significant threat to human health, public safety, and many other aspects [2]. To date, more than 163 million cases have been confirmed with more than 3.37 million deaths related to COVID-19 worldwide as of May 16th, 2021 [3].

Similar to other coronaviruses, the genome of SARS-CoV-2 consists of a positive-sense single-stranded RNA molecule ranging from 26 to 32 kb in size, which was believed to have emerged from bats. However, there is no strong evidence to ensure its zoonotic origin [4]. Coronavirus particles are spherical in shape with diameters between 80 and 160 nm. The genetic structure of SARS-CoV-2 contains 4 structural proteins and 16 non-structural proteins with the arrangement of open reading frames (ORFs) as replicase and protease (1a-1b) and other major glycoproteins like Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N) proteins following a typical 5'-3' order [5]. Among these, the spike protein (S) coats the envelope surface and is responsible for virus penetration into host cells while membrane (M) and envelope (E) proteins cover the membrane. Also, in the SARS-CoV-2 genome, there is an essential gene for virus replication called the *RdRp* (RNA-dependent RNA polymerase) gene in the ORF1ab sequence, which is highly conserved and specific for this virus strain [6]. SARS-CoV-2 possesses many virulence factors that play an important role in the invasion, escape, and multiplication in host cells. This virus has the ability to stay in the human body for a very long time until it causes illness in the respiratory, digestive, and nervous systems of humans and other animals [7].

In order to manage COVID-19, the first step requires a rapid and accurate detection of SARS-CoV-2. Remarkably, the WHO has recommended reverse transcription-polymerase chain reactions (RT-PCR) targeting the *E* gene and *RdRp* gene for the confirmation of SARS-CoV-2 diagnosis. The establishment of positive standard

[*]*Corresponding author: E-mail: dinhhang@vmmu.edu.vn*

sources play a critical role in development and assessment of diagnostic kits as patients' specimens are not available for every laboratory. Besides, for virus culturing, these samples require handling in biosafety level 3 laboratories. Therefore, plasmids or RNA containing the target sequence of SARS-CoV-2 need to be prepared. Therefore, this study focused on the establishment of a novel assay for *in-vitro* transcribed RNA as positive standards based on amplicons of the self-priming PCR targeting envelope protein gene and *RdRp* gene of SARS-CoV-2.

2. Materials and methods

2.1. Envelope protein gene and RdRp gene cloning of SARS-CoV-2 in E. coli

E gene and RdRp gene construction and amplification: Using primer design software, the specific primers sMn1-Fwd/Rev (*E* gene) and sMn2-Fwd/Rev (*RdRp* gene) (Nucleotide position 26248 to 26366 and 15422 to 15538, GenBank accession no. NC_045512.2, respectively) were designed to amplify the target nucleotide sequence of the *E* and *RdRp* genes from SARS-CoV-2 (Table 1). Self-priming refers to the folding back of oligonucleotides. Given the advantage of self-priming that allows amplification initiation without the need for a DNA template, the forward and reverse primer were designed to interact with each other to form the double-stranded DNA fragments of interest. The self-priming PCR process consisted of denaturation (95°C/5 min), annealing and amplifying (94°C/30 s, 40°C/30 s, 72°C/40 s) for 35 cycles, followed by final extension step at 72°C/7 min. The PCR products (119 and 117 bp in length) were analysed by electrophoresis in a 1.2% agarose gel with ethidium bromide then purified by a PCR purification kit (Thermo Scientific, USA).

Table 1. Primers and probes used for this study.

No	Primer/ probe	Sequence (5'-3')	Target gene	Reference
1	sMn1-F	GAGACAGGTACGTTAATAGTTAATAGCGTACTTCTT TTTCTTGCTTTCGTGGTATTCTTGCTAGTTACAC	Envelope	This study
2	sMn1-R	ACAATATTGCAGCAGTACGCACACAATCGAAGCGC AGTAAGGATGGCTAGTGTAACTAGCAAGAATACCA	Envelope	This study
3	sMn2-F	AATATTGAGTGAAATGGTCATGTGTGGCGGTTCACT ATATGTTAAACCAGGTGGAACCTCATCAGGAGAT	*RdRp*	This study
4	sMn2-R	CAGCTGACAGCTTGACAAATGTTAAAAACACTATTAG CATAAGCAGTTGTGGCATCTCCTGATGAGGTTC	*RdRp*	This study
5	qPrimer1	ACAGGTACGTTAATAGTTAATAGCGT	Envelope	[8]
6	qPrimer2	ATATTGCAGCAGTACGCACACA	Envelope	[8]
7	nCoV-F	GTGAAATGGTCATGTGTGGCGG	*RdRp*	[9]
8	nCoV-R	CAAATGTTAAAAACACTATTAGCATA	*RdRp*	[9]
9	nCoVE_pr	/56-FAM/ACACTAGCCATCCTTACTGCGCTTCG /3BHQ_1/	Envelope	[8]
10	nCoV_pr	/56-FAM/ CAGGTGGAACCTCATCAGGAGATGC /3BHQ-1/	*RdRp*	[9]

Cloning of E gene and RdRp gene in E. coli DH5α: The purified PCR products were inserted directly into pGEM-T easy vector (Promega, USA) (Fig. 1) and transformed into *E. coli* DH5α competent cells by heat shock at 42oC/90 s. Then, recombinant plasmids were tested by a *SspI* restriction enzyme (Thermo Scientific, USA). The targeted sequences were then identified by Sanger method, and MegaX software was used to compare and analyse against other determined sequences on GenBank.

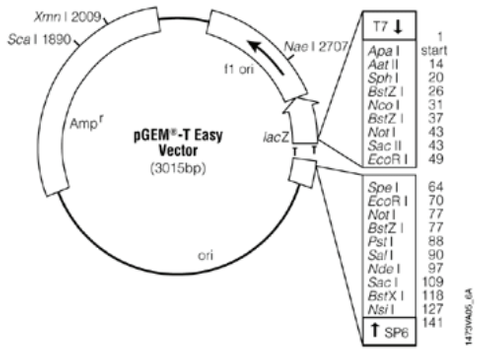


Fig. 1. The pGEM-T easy vector (pGEM-T easy vector systems, Promega, USA).

2.2. In-vitro synthesisation of E protein gene and RdRp gene of SARS-CoV-2

After analysis, the recombinant plasmid containing the sMn1/sMn2 sequence was cleaved at one point downstream of the T7 promoter and the segment to be transcribed using *NdeI* restriction enzyme creating a linear strand. The RNA product was then purified (PCR purification kit - Thermo Scientific, USA) for RNA *in-vitro* synthesis.

Synthesisation of the *E* and *RdRp* genes of SARS-CoV-2: Reaction components were 5X TranscriptAid Reaction buffer 4 µl; NTP mix (25 mM) 8 µl; TranscriptAid Enzyme Mix 2 µl (Thermo Scientific, USA); and DNA plasmid 6 µl; DEPC - treated water added up to the total volume of 20 µl. RNA *in-vitro* was synthesised at 37oC in 2 h. DNAseI was used to remove any remaining DNA at 37°C in 1.5 h. The RNA product was purified by QIAGEN RNeasy Mini kit (Qiagen, Germany) following the conventional protocol. Purified RNA were stored at -80°C for any further experiment (Fig. 2).

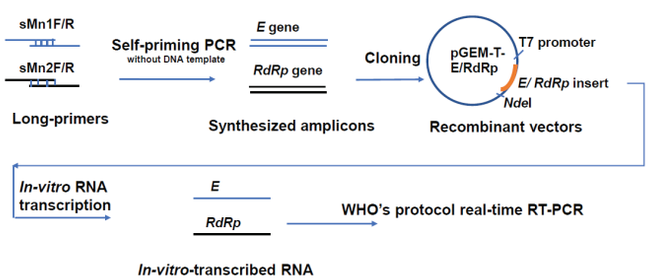


Fig. 2. Diagram of the experimental procedure for *in-vitro* transcribed RNA as positive standards based on amplicons of self-priming PCR targeting the *E* gene and *RdRp* gene of SARS-CoV-2.

A NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, USA) was used to quantify and assess the purity of the purified RNA. Also, their molecular size was visualised by conducting one-step RT-PCR (Qiagen, Germany) on 1.2% agarose gel electrophoresis. Following up with the WHO protocol

for one-step, real-time RT-PCR using Superscript III one-step RT-PCR system with Platinum *Taq* Polymerase (Thermo Scientific, USA), the primer qPrimer(1+2), probe nCoVE-pr and primer nCoV Fwd/Rev, probe nCoV_pr were used for the reaction of *in-vitro* synthesised RNAs sMn1 and sMn2, respectively (Table 1) [8]. After analysis by Sanger sequencing, the *in-vitro* transcribed RNAs were stored at -80oC to determine their stability after 1, 3, and 6 months using one-step real-time RT-PCR assays.

3. Results and discussion

3.1. Cloning envelope protein gene and RdRp gene of SARS-CoV-2 in E. coli

Observation from gel electrophoresis results of PCR products of primer sMn1 Fwd/Rev and sMn2 Fwd/Rev, the synthesising process succeeded with significant bands corresponding to sizes of 119 bp and 117 bp each (Fig. 3A).

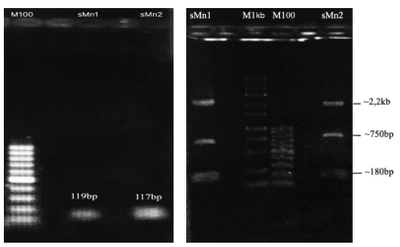


Fig. 3. (A) Self-priming PCR products of sMn1 (*E* gene) and sMn2 (*RdRp* gene) of SARS-CoV-2; (B) Recombinant plasmids of sMn1 and sMn2 after being cleaved by *Ssp* I restriction enzyme. Lane M100, 100 bp DNA ladder (Thermo Scientific, USA); Lane M1kb, 1kb DNA ladder (Thermo Scientific, USA).

As the SARS-CoV-2 genome contains positive-sense single-stranded RNA, forward and reverse primers were designed to create the gene fragment of interest for each envelope protein gene and *RdRp* gene. sMn1 (*E* gene) was synthesised using sMn1 Fwd/Rev primers and, similarly, sMn2 Fwd/Rev for sMn2 (*RdRp* gene).

The pGEM-T easy vector was chosen for its advantages in the cloning process such as its small size of 3015 bp and numerous restriction sites within the multiple cloning regions. Also, these are high-copy-number vectors in *E. coli* containing T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase. Insertional inactivation of the α-peptide allows the identification of recombinants by blue/white screening on indicator plates. Recombinant plasmids were checked by *SspI* restriction enzyme and sequencing. Their sizes were calculated to be 3134 and 3132 bp. Digestion by the *SspI* restriction enzyme resulted in three specific bands equivalent to approximately 0.18, 0.75, and 2.2 kb as measured. Recombinant plasmids were then confirmed by Sanger sequencing.

Sequencing results proved that the cloned genes are exactly sMn1/2 belonging to the *E* and *RdRp* genes of SARS-CoV-2. The comparison consequences indicated two point mutations on the sMn2 sequence at positions of $G_{02} \rightarrow A_{02}$ and $T_{32} \rightarrow C_{32}$, while the sMn1 sequence was completely identical to all other 4 reference sequences. The important takeaway from sequence analysis is to verify if the sequences of interest were cloned into the exact position downstream to the T7 promoter of pGEM-T easy vector, therefore confirming that the transcribed *in-vitro* RNA afterward will be specifically a positive-sense, single-stranded RNA of SARS-CoV-2. The plasmids containing the targeted genes are named as pGEMT_sMn1_E and pGEMT_sMn2_P (Fig. 4).

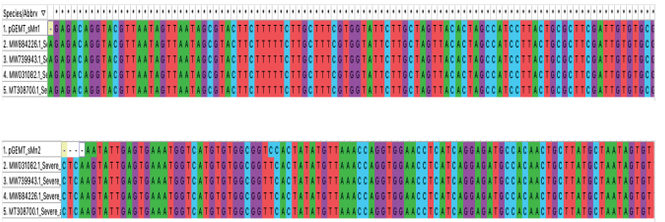


Fig. 4. Comparison between nucleotide sequence of sMn1/sMn2 of SARS-CoV-2 and other 4 reference Coronavirus SARS-CoV-2 strains on Genbank. Reference sequences are named in order of their code, name, and origin.

3.2. In-vitro synthesisation of envelope protein gene and RdRp gene of SARS-CoV-2

While the envelope protein gene (*E*) is a "universal" gene that codes for envelope proteins that are found in all coronaviruses, the RNA-dependent RNA polymerase gene (*RdRp*) of the ORF1ab sequence was found only in SARS-CoV-2. Given the importance of *RdRp* for viability and replication/transcription of RNA viruses, mutations in this gene are statistically less likely to occur [10]. Thus, numerous studies using these two genes have been provided to indicate SARS-CoV-2 based on real-time RT-PCR techniques. Our study focused on the highly conserved genomic region of the envelope protein and *RdRp* for cloning and transcription of *in-vitro* RNA, thus producing positive standards for further SARS-CoV-2 molecular diagnostic studies.

Transcription of *in-vitro* RNA was generated after the cleavage of the *NdeI* restriction enzyme to linearize the recombinant vectors. The *NdeI* enzyme was chosen as it has only one cleavage site on the recombinant plasmid at the position of 97 bp after the T7 promoter, which makes sure the process of transcription will result in a specific target RNA.

Theoretically, successfully transcribed *in-vitro* RNA sMn1 would have 216 bases and 214 bases for sMn2. Gel electrophoresis of the *in-vitro*-transcribed RNA showed clear bands and no by-products (Data not shown). The *in-vitro* transcribed RNA of the *E* and *RdRp* targets had concentrations of 200-800 ng/µl and A260/A280 ratios of 2.0-2.2. The number of copies was calculated by an online tool from http://www.scienceprimer.com [11] based on the concentration and the length of these single stranded *RNAs*, then RNA stocks were diluted into a series of 10-times concentration from 10^{-1} to 10^{6} copies/ml. All diluted RNA solutions were used as RNA templates for a real-time RT-PCR following WHO's protocol to confirm the *in-vitro*-transcribed RNA. Fig. 5 shows that the targeted RNAs were successfully created. Additionally, their

stability during 1, 3, and 6 months of conservation shows equivalent results on one specific concentration (Data not shown).

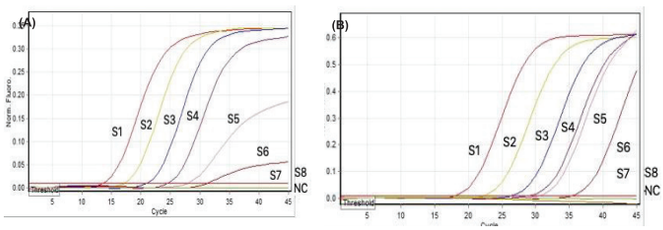


Fig. 5. Evaluation of quality of *in-vitro* transcribed *RNA* of sMn1 and sMn2 using the standard WHO protocol for SARS-CoV-2 detection by real-time RT-PCR. (A) sMn1 and **(B)** sMn2; S1-S8: in vitro RNA was diluted to the concentration of 10^{6}-10^{-1} copies/ul; NC: negative control.

In order to create a positive standard source enabling SARS-CoV-2 detection based on real-time RT-PCR, this study has proposed two cloned and transcribed *in-vitro* RNAs. Due to the properties of stability and ease of cloning into *E. coli* cells, plasmid-carrying target genes are normally used to establish positive standards. In spite of that, plasmids are only suitable for subjects with DNA as their genetic material. In this particular case, since SARS-CoV-2 uses RNA as its genetic material, the ideal positive standard must be RNA. Hence, *in-vitro*-transcribed RNA has been accessed by many studies [12-14].

Generally, two approaches are used to synthesise *in-vitro* RNA: direct transcription and indirect transcription through cloning. With direct transcription, RNA is produced using PCR products containing T7 promoter sequences at their 5'-end. This can be achieved by designing PCR primers having 5'-overhang ends including the T7 promoter [15, 16]. Meanwhile, for indirect transcription, no special design of the PCR primers is required, and plasmids are easily cloned with large amounts as the T7 promoter sequence was provided in the cloning vectors. On the other hand, through cloning techniques, target genes can be sustainably stored in plasmids as well as in *E. coli* cells. Our research had constructed an in vitro-transcribed RNA protocol of the *E* protein gene and *RdRp* gene of SAR-CoV-2 by indirect transcription resulting in RNA with relatively high purity when compared to direct assays [15, 16] and also higher than that of the in-vitro RNA synthesised according to Akyurek, et al. (2021) with about 10^{8} copies/µl (approx. 0.0536 ng/µl) [13] and P.T. Lan, et al (2020) [17] for 7.94 x10^{11} copies/µl.

4. Conclusions

This study successfully generated *in-vitro*-transcribed RNA coding for a specific fragment of the envelope protein gene (sMn1) and the *RdRp* gene (sMn2) of SARS-CoV-2 with concentrations of 200-800 ng/µl and A260/A280 ratios of 2.0-2.2 for sMn1 and sMn2. The *in-vitro* RNAs were tested for their content, purity, RNA electrophoresis, one-step real-time RT-PCR, and sequencing to confirm target fragments included in each RNA molecule. The stability of *in-vitro* RNA over a 6-month storage period was investigated. This experiment establishes a fundamental assay for the production of positive standards for RNA-based molecular techniques.

CRediT author statement

Hang Thi Thu Dinh, Mai Ngoc Nguyen, Su Xuan Hoang: Data curation, Software, Visualisation, Investigation, Writing original draft preparation, Conceptualisation, Methodology.

ACKNOWLEDGEMENTS

We would like to record our appreciation to everyone who participated in this work, especially the Institute of Biomedicine and Pharmacy, Vietnam Military Medical University for providing facilities and support toward our project. This work was supported by the Ministry of Science and Technology, Vietnam under grant number: DTDL.CN.29/20.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] J. Page, D. Hinshaw, B. McKay (2021), "In hunt for COVID-19 origin, patient zero points to second Wuhan market - The man with the first confirmed infection of the new coronavirus told the WHO team that his parents had shopped there", *The Wall Street Journal*, https://wsj.com/articles/in-hunt-for-covid-19-origin-patient-zero-points-to-second-wuhan-market-11614335404, accessed 26 February 2021.

[2] C. Zimmer (2021), "The secret life of a Coronavirus", *The New York Times*, https://nytimes.com/2021/02/26/opinion/sunday/coronavirus-alive-dead.html, accessed 26 February 2021.

[3] Google (2021), "New", https://news.google.com/covid19/map?hl=vi&mid=%2Fm%2F01crd5&gl = V&ceid=VN%3Avi, accessed 26 February 2021.

[4] Z.W. Ye, S. Yuan, K.S. Yuen, et al. (2020), "Zoonotic origins of human coronaviruses", *International Journal of Biological Sciences*, **16(10)**, pp.1686-1697, DOI: 10.7150/ijbs.45472.

[5] A.A.T. Naqvi, K. Fatima, T. Mohammad, et al. (2020), "Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach", *Biochim. Biophys. Acta Mol. Basis Dis.*, **1866(10)**, DOI: 10.1016/j.bbadis.2020.165878.

[6] R.H. Paul, I.I.G. Rather (2020), "The laboratory test for COVID-19, explained", *Science, The Wire*, https://science.thewire.in/the-sciences/rt-pcr-covid-19-rna-polymerase-thermal-cycler/, accessed 26 February 2021.

[7] Y. Yang, Z. Xiao, K. Ye, et al. (2020), "SARS-CoV-2: Characteristics and current advances in research", *Virology Journal*, **17(1)**, DOI: 10.1186/s12985-020-01369-z.

[8] WHO (2021), *Protocol: Real-Time RT-PCR Assays For The Detection of SARS-CoV-2*, https://who.int/docs/default-source/coronaviruse/real-time-rt-pcr-assays-for-the-detection-of-sars-cov-2-institut-pasteur-paris.pdf, accessed 26 February 2021.

[9] V.M. Corman, O. Landt, M. Kaiser, et al. (2020), "Diagnostic detection of 2019-nCoV by real-time RT-PCR", *Euro Surveill.*, **25(3)**, DOI: 10.2807/1560-7917.ES.2020.25.3.2000045.

[10] M. Pachetti, B. Marini, F. Benedetti, et al. (2020), "Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant", *Journal of Translational Medicine*, **18(1)**, DOI: 10.1186/s12967-020-02344-6.

[11] Science Primer (2021), *Copy Number Calculator for Realtime PCR*, http://scienceprimer.com/copy-number-calculator-for-realtime-pcr, accessed 28 February 2021.

[12] J. Bland, A. Kavanaugh, L.K. Hong, et al. (2021), "Development and validation of viral load assays to quantitate SARS-CoV-2", *Journal of Virological Methods*, **291**, DOI: 10.1016/j. jviromet.2021.114100.

[13] S. Akyurek, S.N.S. Demirci, Z. Bayrak, et al. (2021), "The production and characterization of SARS-CoV-2 RNA reference material", *Analytical and Bioanalytical Chemistry*, **413(13)**, pp.3411-3419, DOI: 10.1007/s00216-021-03284-w.

[14] T.G.W. Graham, C.D. Darzacq, G.M. Dailey, et al. (2021), "Open-source RNA extraction and RT-qPCR methods for SARS-CoV-2 detection", *PLOS ONE*, **16(2)**, DOI: 10.1371/journal.pone.0246647.

[15] E. Keyaerts, L. Vijgen, P. Maes, et al. (2006), "Viral load quantitation of SARS-coronavirus RNA using a one-step real-time RT-PCR", *International Journal of Infectious Diseases*, **10(1)**, pp.32-37, DOI: 10.1016/j.ijid.2005.02.003.

[16] M.L. Choudhary, V. Vipat, S. Jadhav, et al. (2020), "Development of *in vitro* transcribed RNA as positive control for laboratory diagnosis of SARS-CoV-2 in India", *Indian J. Med. Res.*, **151(2&3)**, pp.251-254, DOI: 10.4103/ijmr.IJMR_671_20.

[17] P.T. Lan, Q.C. Hoang, H.T. Linh, et al. (2020), "Development of standardised specimens with known concentrations for severe acute respiratory syndrome Coronavirus 2 Realtime-RT-PCR testing validation", *Bulletin of The World Health Organization*, DOI: 10.2471/BLT.20.259630.

DOI: 10.31276/VJSTE.64(4).64-69

Association of BRAF V600E immunoexpression with clinicopathological variant groups in papillary thyroid carcinoma

Hung Nguyen Van[1, 2*], Thanh Nguyen Tuan[3], Minh Tran Ngoc[2, 3], Tuyen Pham Van[1, 3], Luan Dao Thi[1, 2]

[1]*Department of Pathology, Hanoi Medical University, 1 Ton That Tung Street, Trung Liet Ward, Dong Da District, Hanoi, Vietnam*
[2]*Department of Pathology, Hanoi Medical University Hospital, 1 Ton That Tung Street, Trung Liet Ward, Dong Da District, Hanoi, Vietnam*
[3]*Centre of Pathology and Cytopathology, Bach Mai Hospital, 78 Giai Phong Street, Phuong Mai Ward, Dong Da District, Hanoi, Vietnam*

Received 24 December 2021; revised 11 February 2022; accepted 25 March 2022

Abstract:

The detection of BRAF V600E immunoexpression in papillary thyroid carcinoma (PTC) is extremely important because of its association with poor prognosis. Is poor prognosis really associated with all histological variants of thyroid papillary carcinoma? To contribute to this elucidation, using the BRAF V600E mutation-specific antibody, the present study was performed to determine the association between BRAF V600E immunoexpression and clinicopathological variant groups of PTC. Tumour tissue samples collected from 95 patients with papillary thyroid cancer and tumour size <3.5 cm in diameter were processed by routine microscopic technique and using Ventana Benchmark XT automated stainer. BRAF V600E immunoexpression in the group of the aggressive clinicopathological group is directly proportional to lymph node and distal metastasis (p=0.005) in contrast to the favourable clinicopathological group. Finally, the lymph nodal and distal metastasis (poor prognosis) of the aggressive clinicopathological group of PTC can be better predicted when combining BRAF V600E immunostaining with conventional histopathological technique, conversely, these results could not be achieved in the favourable clinicopathological group. These findings suggest that BRAF V600E immunoexpression is a potential candidate for the prognosis of the aggressive clinicopathological group.

Keywords: **BRAF V600E immunoexpression, favourable and aggressive clinicopathological groups, papillary thyroid carcinoma.**

Classification number: **3.2**

1. Introduction

PTC is the most common malignant thyroid disease, accounting for 45-80% of all thyroid cancers. The BRAF (B-isoform of Raf kinase) V600E mutation is common in PTC, with a reported frequency of 25-82.3% [1, 2]. The mutant BRAF protein activates serine/threonine BRAF kinase, which can activate mitosis. BRAF V600E also causes metalloprotease overexpression in the matrix, thereby increasing the mobility and spread of tumour cells. BRAF mutations have been detected in 20-80% of thyroid papillary carcinomas with favourable prognoses, and high frequencies of mutant BRAF are more common in papillary than in follicular variants [3-5].

The BRAF V600E mutation has been associated with age, lymph node metastasis, extra-thyroid expansion, tumour, node, metastasis (TNM) stage, lymph node metastasis, recurrence, and reduced survival but not distal metastasis [6-9]. However, several studies on the association between the BRAF V600E mutation and unfavourable clinical features in PTC showed controversial results and therefore, the association between BRAF V600E mutation and the clinical prognosis of PTC needs to be confirmed [10, 11]. Meanwhile, other studies have demonstrated that characteristics of the unfavourable/aggressive histological variant group (columnar cell, tall cell, hobnail, solid/trabecular) include large tumour size, extra-thyroid expansion, early lymph node metastasis, and high recurrence rate. Favourable histological variants (microcarcinoma, encapsulated, follicular, and cribriform) often have opposite characteristics [3, 4, 12].

Patients who have PTCs with the BRAF V600E mutation have a 1.5-2.1-times higher risk of disease recurrence, extrathyroidal expansion, lymph node metastasis, and TNM stage than patients carrying tumours with wild-type BRAF genotype [13]. According to the American Thyroid Association (ATA) guidelines, the BRAF V600E mutation is one of the risk stratification factors for PTC according to ATA guidelines [14]. The mutant-specific BRAF antibody is useful in detecting the BRAF V600 mutation with more than 95% sensitivity and specificity. Indeed, antibodies specific to BRAF mutations may be more sensitive than molecular tests in detecting BRAF V600E mutations [15-19]. Recently, targeted treatments based on the expression of mutant BRAF protein in thyroid cancer, especially the papillary variant, has increased with quite satisfactory results [20-22]. Thus, the precise identification of aggressive histological variants of the PTC and BRAF V600E mutation status is crucial in preventing sub-optimal or inappropriate patient treatment [12].

The presence of aggressive variants significantly impacts the prognosis and treatment of patients with PTC based on the ATA guidelines. The identification of aggressive variants can modify the therapeutic approach, especially in small thyroid tumours [14]. Therefore, this study aims to determine the relationship between BRAF V600E immunoexpression and clinicopathological variant groups (e.g., aggressive and favourable variants, tumour size, lymph node metastasis, and distal metastasis) of PTC.

*Corresponding author: Email: nguyenvanhung@hmu.edu.vn

2. Materials and methods

2.1. Tissue samples

Ninety-five patients with primary PTC who underwent an initial operation (total thyroid lobectomy with or without lymph node dissection) at the Hanoi Medical University Hospital, Vietnam, between October 2019 and June 2020 were enrolled in this retrospective study. Recurrent PTCs, secondary thyroid carcinomas, and thyroid tumours ≥3.5 cm in size were excluded. PTC tissue samples were obtained from patients residing in the northern regions of Vietnam.

Data were obtained from archived records including age, sex, lymph node metastasis, distal metastasis, histological type, primary tumour location, and tumour size. The corresponding slides and tissue samples embedded in paraffin blocks were also obtained. Formalin-fixed paraffin-embedded tissue blocks were used to prepare hematoxylin and eosin slides to determine areas of tumour cells. The identification of histological variants of PTC was based on the histological classification of thyroid cancer by the World Health Organization in 2017 [4].

Based on clinical behaviour, aggressive (high-risk) and favourable prognosis groups were identified among the PTC variants [12, 23]. The favourable prognosis group consisted of four variants: microcarcinoma, encapsulated, follicular, and cribriform. The clinically high-risk aggressive group (ability to penetrate and metastasize early and extra-thyroid expansion) also consisted of four variants: tall cell, columnar cell, hobnail, and solid/trabecular. The histological characteristics of the aggressive variant group were as follows: tall cell variant, ≥30% tall cells (height twice their width) with oncocytic cytoplasm; columnar cell variant, columnar cells with pseudostratified nuclei, no eosinophilic cytoplasm, and hypercellular neoplasm with thin papillary and glandular structures (the nuclear characteristics of this variant are not as clear as that for classic PTC); solid/trabecular, solid nested variant, lack of tumour necrosis, and absence of marked mitotic activity; and the hobnail variant, ≥30% hobnail cells, micropapillary growth patterns, and high nuclear/cytoplasmic ratio. The clinicopathological characteristics included large tumour size, diffuse tumours, extra-thyroid expansions, patient age (50-60 years) or infiltrative form with extra-thyroid expansion, rapid growth rate, early lymph node metastasis, high rate of recurrences, and higher risk of metastasis. Wild-type BRAF expression was more common than the BRAF V600E point mutation [3-5, 23, 24].

Distal metastases were evaluated via retrospective data analysis of imaging studies (scintigraphy) or histopathologic examination of resected tumours and/or metastases. In the present study, lymph node metastases were usually cervical ones. Distal metastases appeared in the lungs and/or mediastinal lymph nodes. This study has been reviewed and approved by the Ethics Committee of Hanoi Medical University and adheres to the ethical standards laid down in the Declaration of Helsinki amended by the World Medical Association (WMA) General Assembly, Seoul, Korea, October 2008.

2.2. Immunohistochemistry

Tissue slices (4-5 µm) were dried, dewaxed, and rehydrated. Immunostaining was performed using a Ventana Benchmark XT automatic stainer. A kit with BRAF V600E mutant-specific antibody (diluted 1:100, VE1, mouse monoclonal, LOT: E12804, RFF 790-4855, VENTANA) was used. An OptiView DAB immunohistochemistry (IHC) detection kit (Roche Diagnostics), which included pre-treatment with cellular conditioner 1 (pH 8.5) for 64 min, was used for visualization of antibody binding. Tissue slices were then incubated at 37°C with the VE1 antibody for 16 min. Incubation of the antibodies was followed by back-staining with hematoxylin and bluing reagent at room temperature for 4 min each. Subsequently, the immunostaining slides were removed, washed with a drop of dishwashing detergent, and mounted.

A Nikon optical microscope (Nikon Microscope ECLIPSE Ci-L, Epi-FL 4) was used to assess immunostaining. Immunostaining for the BRAF V600E antigen was considered positive when the cytoplasm of the tumour cells was reddish-brown. A BRAF V600E-positive colorectal adenocarcinoma was used for the positive control and a section with no primary antibody incubation was the negative control. Interpretation of immunostaining results was performed by two pathologists (HNV, TNT), who were blinded to the clinical evidence.

2.3. Scoring of BRAF V600E immunoreactivity

IHC results were reported as a function of tumour cells with cytoplasmic staining as follows: 0, no colouring or weak colouring intensity of 10% or less of tumour cells; 1+, weak colouring in more than 10% of tumour cells; 2+, moderate colouring in over 10% of tumour cells; and 3+, strong colouring in over 10% of tumour cells. The tumour was considered positive at 1+ or higher and negative at 0 [25].

2.4. Statistical analysis

Statistical analyses were accomplished using SPSS 17.0 for Windows (SPSS, Inc., Chicago, IL, U.S.A.). The link between the BRAF mutation and the clinicopathologic characteristics was analysed using an r-Pearson correlation, p-value. Differences were considered to be statistically significant at $p<0.05$.

3. Results

3.1. Patient characteristics

The clinicopathological features of PTC are shown in Table 1. Ninety-five patients with PTC were included in the study; 90.6% of

Table 1. Clinicopathological features of PTC.

Variables		n (%)
Gender		
	Male	9 (9.4)
	Female	86 (90.6)
Age (y)		
	10-19	1 (1.1)
	20-29	9 (9.5)
	30-39	23 (24.2)
	40-49	34 (35.8)
	50-59	17 (17.9)
	≥60	11 (11.6)
Location		
	Right lobe	47 (49.5)
	Left lobe	41 (43.1)
	Isthmus	7 (7.4)
Size		
	<1 cm	22 (23.2%)
	1-<2 cm	67 (70.5%)
	2-3.5 cm	6 (6.3)

patients were female (female:male ratio of 9.5:1). Patient ages ranged from 18-68 years (mean: 43.2±11.2 years). The highest number of cases were in the 40-49 years age group (35.8%, 34 cases), followed by 30-39 years (24.2%, 23 cases). Patients aged 10-19 years accounted for the lowest number of cases (1.1%, 1 case).

Tumours appeared in the right lobe in 47 patients (49.5%) and the left lobe in 41 patients. Seven patients had a tumour in the isthmus. No patients had tumours in both lobes. The minimum tumour size was 0.3 cm and the largest tumour size was 3.2 cm The average tumour size was 2±0.4 cm. Tumours were ≤1 cm in size in 22 cases (23.2%), 1-2 cm in 67 cases (70.5%), 2-3 cm in 5 cases, and 3.2 cm in 1 case.

3.2. Histopathological and immunohistochemistry characteristics

Table 2 shows that eleven variants of PTC were identified. The most common variants were conventional (26 tumours, 27.4%), microcarcinoma (22 tumours, 23.2%), and fasciitis-like stroma (16 tumours, 16.8%). The remaining 31 tumours consisted of eight PTC variants, including the clear cell variant. Four variants of PTC were not identified in the study population including diffuse sclerosing, oncocytic, spindle cell, and Warthin-like (Fig. 1).

Table 2. Correlation of histological variants with BRAF V600E in thyroid papillary carcinoma.

Histology		BRAF V600E	
Histological variants	*n (%)*	*BRAF+ (n, %)*	*BRAF- (n, %)*
Conventional	26 (27.4)	22 (84.6)	4 (15.4)
Follicular	6 (6.3)	2 (33.3)	4 (66.7)
Encapsulated	4 (4.2)	3 (75)	1 (25)
Microcarcinoma	22 (23.2)	17 (77.3)	5 (22.7)
Tall cells	8 (8.4)	5 (62.5)	3 (37.5)
Columnar cells	3 (3.2)	1 (33.3)	2 (66.7)
Cribriform	3 (3.2)	1 (33.3)	2 (66.7)
Hobnail	2 (2.1)	2 (100)	0
With fibromatosis/ fasciitis-like stroma	16 (16.8)	10 (62.5)	6 (37.5)
Solid/trabecular	4 (4.2)	2 (50)	2 (50)
Clear cells	1 (1.1)	1 (100)	0
Total	95 (100)	66 (69.5)	29 (30.5)

r-Pearson correlation =-0.325; p=0.01.

Positive BRAF V600E immunostaining was detected in 66 tumours (69.5%) while 28.4, 29.5, and 11.6% had immunostaining scores of 3+, 2+, and 1+, respectively. The remaining 29 tumours (30.5%) were negative for BRAF expression. BRAF V600E immunoexpression is associated with histological variant and was more common in the following variants: conventional (84.6%), encapsulated (75%), microcarcinoma (77.3%), tall cell (62.5%), hobnail (100%), fasciitis-like (62.5%), clear cell (100%), and less in variants: follicular, columnar, and cribriform (same at 33.3%) (r-Pearson correlation=-0.325; p=0.01) (Table 2). Thus, the high frequency of BRAF V600E immunoexpression is in both aggressive and non-aggressive variants.

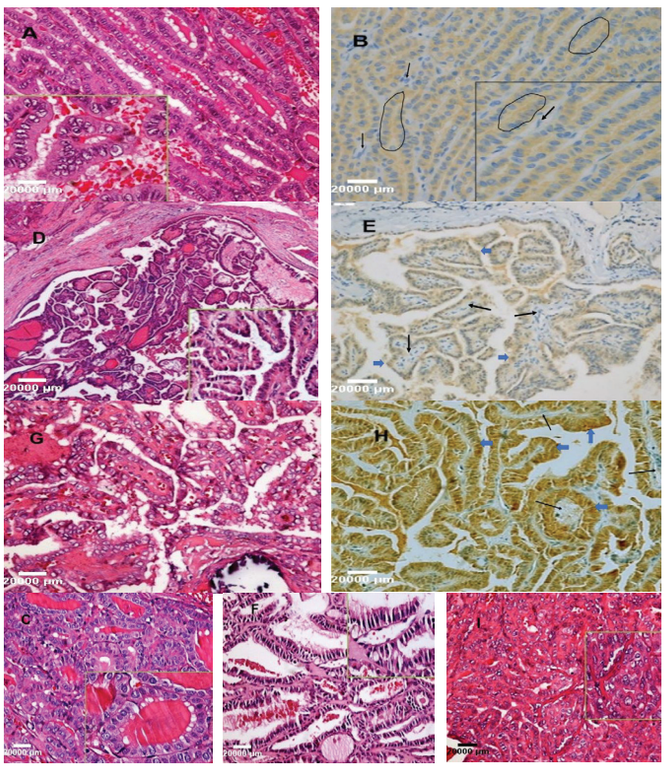


Fig. 1. Variants of PTC. (A) Columnar cell variant; **(B)** BRAF (++), columnar cell variant; **(C)** Follicular variant; **(D)** Encapsulated variant; **(E)** BRAF (+), encapsulated variant; **(F)** Tall cell variant; **(G)** Conventional variant; **(H)** BRAF (+++), conventional variant; **(I)** Solid variant. [H&E (A,C,D,F,G,I) and immunoperoxidase (B,E,H) stain; original magnification: 20x (large box) and 40x (inserts)].
(⇑): internal negative control; *circle and* (➡): positive for yellow brown.

The associations between positive BRAF expression and clinicopathologic factors are shown in Table 3. Tumours were positive for BRAF V600E in 66.7% of men and 69.8% of women (p>0.05). BRAF V600E-positive tumours tended to decrease with age from 88.9% (20-29 years) to 58.8% (50-59 years), then increase in the >60 years age group (90.9%). The differences according to age were significant (p<0.05). The highest incidence of BRAF V600E expression was found in tumours of the right lobe (74.5%), left lobe (65.9%), and isthmus (57.1%). Differences in positive BRAF immunostaining rate by tumour location were not statistically significant (p=0.332).

In the favourable histological group, 3/78 tumours (3.8%) with node metastasis were positive for BRAF V600E (p=0.636). In the aggressive histological group, 10/17 tumours (58.8%) with nodes and distant metastasis (3 lung metastases and 2 mediastinal metastasis) were positive for BRAF V600E (p=0.005). BRAF V600E immunoexpression was associated with patient age (p=0.039), histological variants (r-Pearson correlation=-0.325; p=0.01), and metastasis in the aggressive histological group (p=0.005), but not with gender, tumour location, or tumour size (Table 3).

Table 3. The association between positive BRAF expression and clinicopathologic factors.

Clinicopathologic factors	BRAF V600E (−) (n)	BRAF V600E (−) (n)	Total n (%)
Gende			
Male	3	6	9
Female	26	60	86
p	0.457		
Age			
10-19	1	0	1
20-29	1	8	9
30-39	6	17	23
40-49	13	21	34
50-59	7	10	17
≥60	1	10	11
p	0.039		
Location			
Right lobe	12	35	47
Left lobe	14	27	41
Isthmus	3	4	7
p	0.332		
Size			
<1 cm	18	47	65
1-<2 cm	9	14	23
2–3.5 cm	2	5	7
p	0.476		
Metastasis (lymph node, distal) of favourable group			
Yes	1	2	3
No	21	54	76
p	0.636		
Metastasis (lymph node, distal) of aggressive group			
Yes	0	10	10
No	7	0	7
p	0.005		

3.3. Association between BRAF V600E expression and histological variants according to clinical behaviour

As shown in Table 4, only 17/95 tumours could be classified into aggressive prognostic groups by histological type. BRAF V600E immunoexpression was associated with histological type by high expression rate in tall cell (62.5%), solid (50%), hobnail (100%), and lower in columnar variants (33.3%), (r-Pearson correlation=0.870; p=0.000). So, immunostaining with VE1 antibodies should be preferred for aggressive variants: tall cell, solid, and hobnail.

Table 4. Relationship between the BRAF V600E expression and clinicopathologic variant groups.

Histology			BRAF V600E		
Clinicopathologic variant groups		*n (%)*	*BRAF+ (n, %)*	*BRAF- (n, %)*	p
Favourable histological variant group	Conventional	26 (27.4)	22 (84.6)	4 (15.4)	r-Pearson correlation=-0.239; p=0.035
	Follicular	6 (6.3)	2 (33.3)	4 (66.7)	
	Encapsulated	4 (4.2)	3 (75)	1 (25)	
	Microcarcinoma	22 (23.2)	17 (77.3)	5 (22.7)	
	Clear cells	1 (1.1)	1 (100)	0	
	With fibromatosis/ fasciitis-like stroma	16 (16.8)	10 (62.5)	6 (37.5)	
	Cribriform	3 (3.2)	1 (33.3)	2 (66.7)	
Aggressive histological variant group	Tall cell	8 (8.4)	5 (62.5)	3 (37.5)	r-Pearson correlation=0.870; p=0.000
	Columnar cell	3 (3.2)	1 (33.3)	2 (66.7)	
	Solid/trabecular	4 (4.2)	2 (50)	2 (50)	
	Hobnail	2 (2.1)	2 (100)	0	

3.4. Association between histological variants according to clinical behaviour and clinicopathologic features

As shown in Table 5, only 52/95 tumours could be classified into prognostic groups by histological type. In the favourable variant group (35 tumours, 67.3%), 28 tumours were <1 cm (80.0%) and 7 were 2-3 cm (20.0%). Of the aggressive variants (17 tumours, 32.7%), 10 tumours (58.8%) were <1 cm. The differences in tumour size according to clinicopathological groups were statistically significant (p=0.04). Most tumours in the left lobe (90%) and the isthmus (83.3%) belonged to the group of favourable histological variants (p=0.004).

Table 5. Relationship between histological variant groups and other variables.

Variables	n	Group of favourable histological variants	Group of aggressive histological variants	p
Size				
<1 cm	38	28	10	0.04
1–<2 cm	11	7	4	
2–3.5 cm	3	0	3	
Location				
Right lobe	26	12	14	
Left lobe	20	18	2	0.004
Isthmus	6	5	1	
Lymph node metastasis				
Yes	7	1	6	0.003
No	45	34	11	
Distal metastasis				
Yes	4	0	4	0.009
No	48	35	13	

In the lymph node metastasis group, the majority of tumours (6/7, 85.7%) were aggressive variants (Table 5) and in the group without lymph node metastases, the majority of tumours (34/45, 75.6%) were favourable variants (p=0.003). In the group with distal metastases, all tumours (4/4, 100%) were aggressive variants and in the group with no distal metastases, the majority of tumours (35/48, 72.9%) were favourable variants (p=0.009). Therefore, BRAF V600E immunoexpression was related to patient age and histological variant and not to gender, tumour location, tumour size, lymph node metastasis, and distal metastasis. The aggressive variant group had a worse prognosis: larger tumour size (p=0.004), more lymph node metastasis (p=0.003), and more distal metastasis (p=0.009) than the favourable variant group; Therefore, it is necessary to closely manage the aggressive variant group with the BRAF V600E mutation to limit the poor prognosis as much as possible.

4. Discussion

The frequency at which the BRAF V600E mutation was detected in PTC varies worldwide, for example, in the Philippines (70.59%) [26], Korea (75.8-82.3%), China (62.7%), Hawaii (83.8%), Poland (72.4%), the United States (61.0-77.0%), and the present study, Vietnam (69.5%). The differences between these rates are not significant. These findings are also consistent with the increasing PTC mutation rates detected in recent years [27, 28]. Increasing mutation rates may be occurring because of improved detection methods with better sensitivities and specificities. High rates of BRAF V600E immunoexpression occur with the tall cell variant. Several studies showed that the tall cell variant has the highest BRAF V600E positive

rate (68-100%) [12, 24, 29]. H.Z. Kimbrell, et al. (2015) [1] proved that all cases of tall cell variants were positive. Therefore, in clinical settings, the determination of BRAF V600E mutations in tall cell variants in PTC should be prioritized. The frequency of BRAF V600E mutation in the follicular variant ranges from 21 to 54% [23, 30, 31]. This frequency is consistent with our results for the tall cell variant (62.5%).

Several other factors can influence the measurement of BRAF V600E expression including the research material (frozen or paraffin-embedded tissue), antibody purity, the epitope for the antibody, the amplification system for signal detection, incubation time, and fixation solution for tissue samples [30, 31]. Another possible cause for the varying expression is the gene pair mismatch self-repair that results in the BRAF V600E gene mutation. Excessive oxidation during thyroid hormone synthesis may cause genomic mismatches, which suggests a correlation between iodine concentration (a component in thyroid hormone synthesis) and BRAF mutations [32].

Molecular techniques have been used to detect the BRAF V600E point mutation. These methods are often very expensive, require a lot of time and effort, and are difficult to verify [33]. Immunohistochemical staining is a technique widely used in pathology laboratories. The IHC technique is faster and cheaper than molecular techniques. Importantly, IHC has the added advantage of recognizing individual antigen-carrying tumour cells and evaluating their uniformity [34]. Recently, two mutant-specific antibodies against the BRAF V600E protein have been developed and commercialized: the antibody clone, VE1, developed by Spring Bioscience (Pleasanton, California) and a monoclonal anti-mouse BRAF antibody developed by New East Bioscience (Malvern, Pennsylvania). The VE1 clone has been used in most studies [10, 16, 23]. Currently, available BRAF mutant-specific antibodies have sensitivities and specificities of more than 95% [10, 17-19], and they are more sensitive than molecular tests for detecting the BRAF V600E mutation [17-19, 34]. Molecular techniques are highly dependent on DNA quality and require a sufficient number of tumour cells in the sample [17, 34]. Unlike molecular testing, IHC is less dependent on DNA quality or the proportion of tumour cells present in the sample. Furthermore, the level of mutant protein expression in tumour cells at the individual cell level can be assessed. Thus, each tumour cell in the sample is evaluated for mutations more accurately than molecular analyses [19, 35].

M. Bullock, et al. (2012) [19] suggested that the mutant-specific VE1 antibody should be the gold standard for detecting the BRAF V600E point mutation in PTC. P. McKelvie, et al. (2013) [10] also showed that detection of the BRAFV600E mutation by IHC was more sensitive than standard real-time PCR using a commercial kit (Seegene) to detect the BRAFV600E mutation and recommended using the mutant-specific VE1 antibody (BRAF V600E) in pathology laboratories where conditions for routine sequencing of PTC specimens are not favourable. Several studies have shown that the number of positive cells detected using IHC was higher than the mutant allele frequency, which was perhaps due to the presence of heterogeneous clones for BRAF V600E within the same tumour. The BRAF V600E mutation was found only in PTC cells and not in the stroma. Given that the study samples contain numerous different stromal tissues without the BRAF V600E mutation (stroma and vascular), the allele frequency in PTC may be significantly reduced in molecular tests [17-19, 36].

The detection of the BRAF V600E mutation by IHC upon diagnosis can immediately provide individual prognostic information to guide treatment and monitoring in patients with PTC. The presence of the BRAF V600E mutation significantly increases the risk of lymph node metastasis in PTC. Considerably high frequencies of microscopic lymph node metastases were detected among lymph nodes evaluated as negative by preoperative ultrasound, which ranged from 42% for primary thyroid tumours <2 cm to 86% for tumours >3 cm [10, 17, 33, 34]. P. McKelvie, et al. (2013) [10] demonstrated a significant association between the BRAF V600E mutation and gender, but not with other factors such as age, primary tumour size, extra-thyroid expansion, lymph node metastasis, distal metastasis, or the AJCC stage at the time of diagnosis.

The follicular variant usually has a higher frequency of BRAF V600E mutations and has characteristics more similar to the conventional papillary variant than the encapsulated variant [16, 37, 38]. Synthetic analysis by H. Lui, et al. (2015) [33] of 20,764 patients showed that the BRAF V600E mutation was associated with several high-risk clinical variables used in the prognostic system, including extra-thyroid invasion, high TNM stage, lymph node metastasis, recurrence, and total survival but not with distal metastasis. The diagnosis of distal metastasis varies widely among countries and even among medical centres within the same country, which may cause discrepancies when comparing results across studies.

Although several authors have found an association between the BRAF mutation and tumour invasiveness, others have been unable to confirm this association. Accordingly, some studies have found that the presence of a BRAF V60E mutation was not correlated with clinical signs of aggression [10, 35] including age, extra-thyroid expansion, TNM stage, lymph node metastases, recurrence, reduced survival, or distal metastasis [39]. However, PTC histology has been significantly associated with the presence of the BRAF V600E mutation [40] whereas follicular variant histology was not associated with the BRAF V600E mutation.

Classifying the two clinicopathologic variants was useful in determining the association between BRAF V600 immunoexpression and patient age and histological variants in the favourable and aggressive histological groups. In the favourable variant group, the tumour size was often small <1 cm ($p=0.04$) and more preferred in the left lobe ($p=0.004$). The tumour was more prone to lymph node metastases ($p=0.003$) in the aggressive variant group and distant metastases ($p=0.009$).

5. Conclusions

Lymph nodal and distal metastases (poor prognosis) in the aggressive (or high-risk) clinicopathological group of PTC can be better predicted when the molecular feature of BRAF V600E is integrated with histopathological characteristics; conversely, these results could not be achieved in the favourable histological group. These features suggest that BRAF V600E immunoexpression may be considered as a potential candidate for the prognosis of the high-risk histological group. Therefore, the BRAF V600E test should be added to the diagnosis in the high-risk clinicopathological group.

CRediT author statement

Hung Nguyen Van, Thanh Nguyen Tuan, Minh Tran Ngoc, Tuyen Pham Van, Luan Dao Thi: Data curation, Software, Visualisation, Investigation, Writing original draft preparation, Conceptualization, Methodology.

ACKNOWLEDGEMENTS

The authors would like to thank the Ethics Committee of Hanoi Medical University for the study's approval and Enago (www.enago.com) for English language review.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] H.Z. Kimbrell, A.B. Sholl, S. Ratnayaka, et al. (2015), "BRAF testing in multifocal papillary thyroid carcinoma", *Biomed Res. Int.*, **2015**, pp.1-7, DOI: 10.1155/2015/486391.

[2] R. Ciampi, Y.E. Nikiforov (2007), "Minireview: RET/PTC rearrangements and braf mutations in thyroid tumorigenesis", *Endocrinology*, **148(3)**, pp.936-941, DOI: 10.1210/en.2006-0921.

[3] M. Nath, L.A. Erickson (2018), "Aggressive variants of papillary thyroid carcinoma: Hobnail, tall cell, columnar, and solid", *Adv. Anat. Pathol.*, **25(3)**, pp.172-179, DOI: 10.1097/PAP.0000000000000184.

[4] R.V. Lloyd, R.Y. Osamura, G. Klöppel, et al. (2017), *WHO Classification of Tumours of Endocrine Organs, 4th ed.*, International Agency for Research on Cancer (IARC) Publisher, **10**, 355pp.

[5] S. Asioli, L.A. Erickson, A. Righi, et al. (2013), "Papillary thyroid carcinoma with hobnail features: Histopathologic criteria to predict aggressive behaviour", *Hum. Pathol.*, **44(3)**, pp.320-328, DOI: 10.1016/j.humpath.2012.06.003.

[6] N. Fakhruddin, M. Jabbour, M. Novy, et al. (2017), "BRAF and NRAS mutations in papillary thyroid carcinoma and concordance in BRAF mutations between primary and corresponding lymph node metastases", *Sci. Rep.*, **7(1)**, DOI: 10.1038/s41598-017-04948-3.

[7] C. Liu, T. Chen, Z. Liu (2016), "Associations between BRAF V600E and prognostic factors and poor outcomes in papillary thyroid carcinoma: A meta-analysis", *World J. Surg. Oncol.*, **14(1)**, DOI: 10.1186/s12957-016-0979-1.

[8] G. He, B. Zhao, X.Zhang, et al. (2014), "Prognostic value of the BRAF V600E mutation in papillary thyroid carcinoma", *Oncol. Lett.*, **7(2)**, pp.439-443, DOI: 10.3892/ol.2013.1713.

[9] S. Choden, S. Keelawat, C.K. Jung, et al. (2020), "VE1 immunohistochemistry improves the limit of genotyping for detecting BRAFV600E mutation in papillary thyroid cancer", *Cancers (Basel).*, **12(3)**, DOI: 10.3390/cancers12030596.

[10] P. McKelvie, F. Chan, Y. Yu, et al. (2013), "The prognostic significance of the BRAFV600E mutation in papillary thyroid carcinoma detected by mutation-specific immunohistochemistry", *Pathology*, **45(7)**, pp.637-644, DOI: 10.1097/PAT.0000000000000008.

[11] D. Chen, W. Qi, P. Zhang, et al. (2017), "Investigation of BRAF V600E detection approaches in papillary thyroid carcinoma", *Pathol. Res. Pract.*, **214(2)**, pp.303-307, DOI: 10.1016/j.prp.2017.09.001.

[12] A.C. Pelaz, J.P. Shah, J.C.H. Prera, et al. (2020), "Papillary thyroid cancer-aggressive variants and impact on management: A narrative review", *Adv. Ther.*, **37(7)**, pp.3112-3128, DOI: 10.1007/s12325-020-01391-1.

[13] T.H. Kim, Y.J. Park, J.A. Lim, et al. (2012), "The association of the BRAFV600E mutation with prognostic factors and poor clinical outcome in papillary thyroid cancer: A meta-analysis", *Cancer*, **118(7)**, pp.1764-1773, DOI: 10.1002/cncr.26500.

[14] B.R. Haugen, E.K. Alexander, K.C. Bible, et al. (2016), "2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer", *Thyroid*, **26(1)**, pp.1-133, DOI: 10.1089/thy.2015.0020.

[15] K. Dorak, B. Aggeler, J. Palting, et al. (2014), "Immunohistochemistry with the anti-BRAF V600E (VE1) antibody: Impact of pre-analytical conditions and concordance with DNA sequencing in colorectal and papillary thyroid carcinoma", *Pathology*, **46**, pp.509-517, DOI: 10.1097/PAT.0000000000000119.

[16] J. Zagzag, A. Pollack, L. Dultz, et al. (2013), "Clinical utility of immunohistochemistry for the detection of the BRAF V600E mutation in papillary thyroid carcinoma", *Surg. (United States)*, **154(6)**, pp.1199-1205, DOI: 10.1016/j.surg.2013.06.020.

[17] D. Capper, M. Preusser, A. Habel, et al. (2011), "Assessment of BRAF V600E mutation status by immunohistochemistry with a mutation-specific monoclonal antibody", *Acta Neuropathol.*, **122(1)**, pp.11-19, DOI: 10.1007/s00401-011-0841-z.

[18] O. Koperek, C. Kornauth, D. Capper, et al. (2012), "Immunohistochemical detection of the BRAF V600E-mutated protein in papillary thyroid carcinoma", *Am. J. Surg. Pathol.*, **36(6)**, pp.844-850, DOI: 10.1097/PAS.0b013e318246b527.

[19] M. Bullock, C. O'Neill, A. Chou, et al. (2012), "Utilization of a MAB for BRAF V600E detection in papillary thyroid carcinoma", *Endocr. Relat. Cancer*, **19(6)**, pp.779-784, DOI: 10.1530/ERC-12-0239.

[20] M. Al-Jundi, S. Thakur, S. Gubbi, et al. (2020), "Novel targeted therapies for metastatic thyroid cancer - A comprehensive review", *Cancers (Basel).*, **12(8)**, pp.1-37, DOI: 10.3390/cancers12082104.

[21] Y.L. Han, X.Q. Yu, Y.Z. Yin, et al. (2021), "Identification of potential BRAF inhibitor joint therapy targets in PTC based on WGCAN and DCGA", *J. Cancer*, **12(6)**, pp.1779-1791, DOI: 10.7150/jca.51551.

[22] M. Cabanillas, M. Ryder, C. Jimenez (2019), "Targeted therapy for advanced thyroid cancer: Kinase inhibitors and beyond", *Endocr. Rev.*, **40(6)**, pp.1573-1604, DOI: 10.1210/er.2019-00007.

[23] H.S. Min, C. Lee, K.C. Jung (2013), "Correlation of immunohistochemical markers and BRAF mutation status with histological variants of papillary thyroid carcinoma in the korean population", *J. Korean Med. Sci.*, **28(4)**, pp.534-541, DOI: 10.3346/jkms.2013.28.4.534.

[24] X. Wang, W. Cheng, C. Liu, et al. (2016), "Tall cell variant of papillary thyroid carcinoma: Current evidence on clinicopathologic features and molecular biology", *Oncotarget*, **7(26)**, pp.40792-40799, DOI: 10.18632/oncotarget.8215.

[25] X. Zhu, Y. Luo, Q. Bai, et al. (2016), "Specific immunohistochemical detection of the BRAF V600E mutation in primary and metastatic papillary thyroid carcinoma", *Exp. Mol. Pathol.*, **100(1)**, pp.236-241, DOI: 10.1016/j.yexmp.2016.01.004.

[26] G. Espiritu, J.T Malana, A.J.G.V. Dumasis, et al. (2019), "High preponderance of BRAF V600E mutation in papillary thyroid carcinoma among filipinos: A clinicopathologic study abstract", *J. Glob. Oncol.*, **2019(5)**, pp.1-6, DOI: 10.1200/JGO.18.00085.

[27] F.A. Rashid, J. Munkhdelger, J. Fukuoka, et al. (2020), "Prevalence of BRAF V600E mutation in Asian series of papillary thyroid carcinoma - A contemporary systematic review", *Gland Surg.*, **9(5)**, pp.1878-1900, DOI: 10.21037/gs-20-430.

[28] H.G. Vuong, A.M.A. Altibi, A.H. Abdelhamid, et al. (2017), "The changing characteristics and molecular profiles of papillary thyroid carcinoma over time: A systematic review", *Oncotarget*, **8(6)**, pp.10637-10649, DOI: 10.18632/oncotarget.12885.

[29] R.V. Taibo, D.P. González, J.M.C. Agrícola, et al. (2017), "Aggressiveness of the tall cell variant of papillary thyroid carcinoma is independent of the tumour size and patient age", *Oncol. Lett.*, **13(5)**, pp.3501-3507, DOI: 10.3892/ol.2017.5948.

[30] A.E. Walts, J.M. Mirocha, S. Bose (2015), "Follicular variant of papillary thyroid carcinoma (FVPTC): Histological features, BRAF V600E mutation, and lymph node status", *J. Cancer Res. Clin. Oncol.*, **141(10)**, pp.1749-1756, DOI: 10.1007/s00432-015-1939-9.

[31] Y.Y. Jung, J.H. Yoo, E.S. Park, et al. (2015), "Clinicopathologic correlations of the BRAFV600E mutation, BRAF V600E immunohistochemistry, and BRAF RNA *in situ* hybridization in papillary thyroid carcinoma", *Pathol. Res. Pract.*, **211(2)**, pp.162-170, DOI: 10.1016/j.prp.2014.10.005.

[32] H.R. Chung (2014), "Iodine and thyroid function", *Ann. Pediatr. Endocrinol. Metab.*, **19(1)**, pp.8-12, DOI: 10.6065/apem.2014.19.1.8.

[33] H. Lui, F. Lin (2015), "Application of immunohistochemistry in thyroid pathology", *Arch. Pathol. Lab. Med.*, **139(1)**, pp.67-82, DOI: 10.5858/arpa.2014-0056-RA.

[34] G.V. Long, J.S. Wilmott, D. Capper, et al. (2013), "Immunohistochemistry is highly sensitive and specific for the detection of V600E BRAF mutation in melanoma", *Am. J. Surg. Pathol.*, **37(1)**, pp.61-65, DOI: 10.1097/PAS.0b013e31826485c0.

[35] A. Mathur, W. Moses, R. Rahbari, (2011), "Higher rate of BRAF mutation in papillary thyroid cancer over time: A single-institution study", *Cancer*, **117(19)**, pp.4390-4395, DOI: 10.1002/cncr.26072.

[36] C.A. Routhier, M.C. Mochel, K. Lynch, et al. (2013), "Comparison of 2 monoclonal antibodies for immunohistochemical detection of BRAF V600E mutation in malignant melanoma, pulmonary carcinoma, gastrointestinal carcinoma, thyroid carcinoma, and gliomas", *Hum. Pathol.*, **44(11)**, pp.2563-2570, DOI: 10.1016/j.humpath.2013.06.018.

[37] W. Smith, A. Alajegheh, M. Nassiri (2011), "Correlation between BRAF mutation and the clinicopathological parameters in papillary thyroid carcinoma with particular reference to a follicular variant", *Hum. Pathol.*, **42(4)**, pp.500-506, DOI: 10.1016/j.humpath.2009.09.023.

[38] M. Rivera, R.M. Tuttle, S. Patel, et al. (2009), "Encapsulated papillary thyroid carcinoma: A clinicopathologic study of 106 cases with emphasis on its morphologic subtypes (histologic growth pattern)", *Thyroid*, **19(2)**, pp.119-127.

[39] S.Y. Dong, R.C. Zeng, L.P. Jin, et al. (2017), "BRAFV600E mutation is not associated with central lymph node metastasis in all patients with papillary thyroid cancer: Different histological subtypes and preoperative lymph node status should be taken into account", *Oncol. Lett.*, **14(4)**, pp.4122-4134, DOI: 10.3892/ol.2017.6694.

[40] D.J. Givens, L.O. Buchmann, A.M. Agarwal, et al. (2014), "BRAF V600E does not predict aggressive features of pediatric papillary thyroid carcinoma", *Laryngoscope*, **124(9)**, pp.389-393, DOI: 10.1002/lary.24668.

DOI: 10.31276/VJSTE.64(4).70-78

Review Cannabis: A new strategy against methicillin-resistant *Staphylococcus aureus*

Emerson Joseph Addison*

Central Michigan University, 1200 S. Franklin Street, Mount Pleasant, Michigan, United States of America

Received 15 October 2021; 21 December 2021; accepted 7 January 2022

Abstract:

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a global health concern. Many antibiotics are no longer effective at treating MRSA, which causes an increase in adverse patient outcomes. This has led to calls for new antibiotics and treatment strategies to combat the spread of MRSA and multidrug resistance (MDR). The antimicrobial secondary metabolites found in plants are a promising source for new antibiotics and treatment strategies. *Cannabis sativa* L. is especially promising, as it produces dozens of antimicrobial secondary metabolites that are active against *Staphylococcus aureus* (*S. aureus*) and MRSA strains. In addition to its antimicrobial properties against *S. aureus* and MRSA, cannabis has many other desirable properties for potential antibiotics. Cannabis secondary metabolites are active against a wide range of microorganisms, are generally safe, target multiple bacterial processes and structures, have antimicrobial synergies, have a low potential for resistance development, can be produced inexpensively and combined with existing antibiotics to further reduce costs, and contain secondary metabolites capable of penetrating a variety of in vivo environments. These characteristics make cannabis a potential resource against MRSA and MDR bacteria.

Keywords: **antimicrobial secondary metabolites, cannabinoids, cannabis, *cannabis sativa*, combination therapy, methicillin-resistant, MRSA, *S. aureus*.**

Classification numbers: **3.2, 3.3**

1. Introduction

S. aureus is a commensal bacterium that colonizes approximately 30% of the human population and acts as an opportunistic pathogen [1]. Typically, hosts are asymptomatic; however, infections are common and can range from mild skin infections and abscesses to invasive and life-threatening infections including bacteraemia, endocarditis, and pneumonia [1, 2]. In 2017, *S. aureus* bacteraemia was responsible for approximately 20,000 deaths and 120,000 infections in the United States [2].

S. aureus develops antibiotic resistance (AR) quickly and AR in *S. aureus* is widespread. MRSA is of particular concern as MRSA rates in World Health Organization (WHO) regions typically exceed 20%, which increases risks for patients and necessitates the use of second line, more toxic drugs [3]. In the past, vancomycin was considered the antibiotic of last resort for MRSA infections; however, due to the risk for adverse reactions and increasing rates of vancomycin resistance, newer antibiotics, such as linezolid, daptomycin, quinupristin/dalfopristin, and tigecycline are often used [4]. Unfortunately, these newer antibiotics can be expensive and have risks for adverse reactions [5]. Additionally, resistances, though rare, have already developed for linezolid, daptomycin, quinupristin/dalfopristin and tigecycline [4]. Thus, the search for new antibiotics that are effective against MRSA and other MDR bacteria continues.

This review will examine *Cannabis sativa* as a potential source for new antimicrobial compounds for the treatment of MRSA and other MDR bacteria. Cannabis is promising in this regard because it produces an abundance of antimicrobial secondary metabolites and has many other qualities that are desirable in antibiotic therapy and antibiotic development.

1.1. Cannabis sativa

Cannabis has been cultivated and used as medicine for thousands of years [6-11]. After millions of years of evolution, thousands of years of traditional cultivation and generations of modern selective breeding, there is considerable diversity among varieties of cannabis, with different strains having different medicinal properties [12].

As different cannabis cultivars can be easily cross-bred, and typical plant characteristics like height and leaflet width are insufficient distinctions between varieties, cannabis is often classified by "chemovar" according to biochemical characteristics [13-15]. Each chemovar boasts unique genetics and combinations of the various secondary metabolites found in cannabis [10-16]. As of the writing of this review, there are as many as 700 different chemovars of cannabis [9, 10, 15]. In addition to the influence of genetics on the chemical profile of each chemovar, the chemical profile of cannabis is further influenced by environmental and external factors including nutrition, humidity, temperature, age of plant, harvest time, plant stress, and plant organ and storage conditions [17-20].

Cannabis chemovars produce more than 500 natural secondary metabolites from 18 different chemical classes, including more than 100 cannabinoids and more than 200 terpenes [11, 12, 14, 19, 21-25].

*Email: emersonaddison@gmail.com

Cannabis is noteworthy for its production of cannabinoids, which are lipophilic molecules with low water solubility. Cannabinoids have been only rarely detected in non-cannabis plants [24]. The best-known cannabinoids are tetrahydrocannabinol (THC), the primary psychoactive compound, and cannabidiol (CBD), which is known for having a variety of medicinal properties and for being an antipsychotic. THC is a "weak partial agonist on CB1 and CB2 receptors," while CBD is a "negative allosteric modulator of CB1" [26].

In addition to terpenes and cannabinoids, hundreds of other compounds have been identified in cannabis, including 27 nitrogenous compounds, 18 amino acids, 3 proteins, 6 enzymes, 2 glycoproteins, 34 sugars and related compounds, 50 hydrocarbons, 7 simple alcohols, 12 simple aldehydes, 13 simple ketones, 20 simple acids, 23 fatty acids, 12 simple esters, 1 lactone, 11 steroids, 25 non-cannabinoid phenols, 23 flavonoids, 1 vitamin, 2 pigments, and 9 elements [27]. Cannabis resin, which is naturally produced in the trichomes, is rich in both cannabinoids and terpenes and is "valued for its psychoactive and medicinal properties" [11].

1.2. Antimicrobial activity vs S. aureus and MRSA

The essential oils and extracts from cannabis, as well as many of the individual cannabinoids, have antimicrobial properties and are active against various strains of *S. aureus* and MRSA [6-8, 17, 21, 28-50]. Antimicrobial activity against *S. aureus* and MRSA strains has been demonstrated by many *cannabinoids*, including cannabichromenic acid (CBCA) [32, 39, 40], cannabichromene (CBC) [6, 30, 40], cannabichromene-C0 (CBC homolog) [6], cannabichromene-C1 (CBC homolog) [6], isocannabichromene-C_0 [6], (±)-3″-hydroxy-Δ($^{4''5''}$)-cannabichromene [33], cannabidiolic acid (CBDA) [8, 30, 40], cannabidiol (CBD) [8, 30, 38-40, 44, 48-50], cannabidivarin methyl ester (CBDVM) [32], cannabigerol acid (CBGA) [30, 40], cannabigerol (CBG) [30, 39, 40], 4-acetoxy-2-geranyl-5-hydroxy-3-n-pentylphenol (CGB derivative) [33], 5-acetoxy-6-geranyl-3-n-pentyl-1,4-benzoquinone [46], 5-acetyl-4-hydroxycannabigerol [33], methylated cannabigerol [30], cannabinol (CBN) [30, 39, 40], 8-hydroxycannabinolic acid A [33], 1′ S-hydroxycannabinol [21], carmagerol [30], pre-Δ9-tetrahydrocannabinol (Δ9-THCA) [30, 40], Δ^9-tetrahydrocannabinol (Δ^9-THC) [30, 39, 40, 45, 50], cannabidivarin (CBDV) [40], cannabidivarinic acid (CBDVA) [40], Δ8-tetrahydrocannabinol (Δ^8-THC) [40], tetrahydrocannabivarinic acid (THCVA) [40], Δ9-tetrahydrocannabivarin (THCV) [40], exo-olefin THC [40], and +/-11-OH Δ9-THC [40]. Additionally, THC has been demonstrated to protect mice from acute respiratory distress syndrome (ARDS) and toxicity caused by the cytokine storm triggered by Staphylococcal enterotoxin B (SEB), which is a toxin produced by *S. aureus* [45].

Many of the non-cannabinoid phytocompounds in cannabis are also active against *S. aureus* and MRSA strains [30, 38, 41, 46, 51-68]. Antimicrobial activity against *S. aureus* and MRSA strains has been observed in α-bisabolol (levomenol) [51], carvacrol [65-68], eugenol [52], nerolidol [51], limonene [53], para-cymene (p-cymene) [53], myrcene (β-myrcene) [38, 53], olivetol [30], 1,8-cineole [54, 57, 58, 64], α-pinene [38, 52, 59, 60, 64], β-pinene [38, 52], α-terpineol [58, 60], α-terpinolene [38], terpinen-4-ol [58, 60], thymol [53, 65, 66], β-caryophyllene [38, 59], humulene (α-caryophyllene) [59], β-amyrin [61], cannflavin A [46], naringenin [41, 62, 63], caffeic acid [55], and linoleic acid [56].

Other antimicrobial activity

Antibiotic combinations that are effective against a variety of microorganisms are useful as empirical therapy for the treatment of unidentified pathogens [69]. In addition to being effective against MRSA, the antimicrobial properties of cannabis have been tested against other pathogenic microorganisms, including many species of gram-positive and gram-negative bacteria, a variety of clinically significant fungi, and Leishmania protozoa.

Cannabis essential oil and extracts are active against many species of gram-positive bacteria, including *Bacillus* cereus [38]; *Bacillus pumilus* [29]; *Bacillus subtilis* [7, 17, 29, 38]; *Brevibacterium linens* and *Brochothrix thermosphacta* [17]; *Clostridium tyrobutyricum*, *Clostridium bifermentans*, *Clostridium butyricum* and *Clostridium sporogenes* [70]; *Enterococcus faecalis* [35, 38, 47]; *Enterococcus faecium* and *Enterococcus hirae* [38, 70]; *Micrococcus flavus* [29]; *Listeria monocytogenes* and *Staphylococcus epidermidis* [38]; *Streptococcus salivarius* [70]; and the gram-positive to gram-variable *Micrococcus luteus* [17].

Cannabis essential oil and extracts are active against a variety of gram-negative bacteria, including *Acinetobacter calcoaceticus*, *Aeromonas hydrophyla* and *Beneckea natriegens* [17]; *Bordetella bronchiseptica* [29]; *Escherichia coli* [7, 17, 35, 47]; *Enterobacter aerogenes* [47]; *Flavobacterium suaveolens* [17]; *Helicobacter pylori* [41]; *Pectobacterium carotovorum* [70]; *Pseudomonas aeruginosa* [7, 35, 43]; *Pseudomonas campestris*, *Pseudomonas corrugata*, *Pseudomonas fluorescens*, *Pseudomonas savastanoi*, *Pseudomonas syringae*, and *Pseudomonas viridiflava* [70]; *Proteus vulgaris* [29]; *Salmonella typhimurium* [47]; and *Yersinia enterocolitica* [17].

Cannabis essential oil and extracts are active against a variety of fungi, including *Aspergillus niger* [29]; *Candida albicans* [7, 29, 43]; and *Candida sake*, *Kluyveromyces marxianus*, *Pichia membranaefaciens*, *Schizosaccharomyces pombe*, *Schizosaccharomyces japonicus*, *Torulaspora delbrueckii* and *Zygosaccharomyces bailii* [70]. Fractional cannabis distillations showed activity against *Candida glabrata*, *Candida krusei*, and *Cryptococcus neoformans*; cannabis extracts have also demonstrated activity against the protozoa *Leishmania donovani* [36].

In addition to being active against MRSA, many of the secondary metabolites found in cannabis, including many of the cannabinoids, have also been individually tested against other microorganisms. CBD is of particular interest as a potential antimicrobial, and in one study showed a consistent MIC (minimum inhibitory concentration) of 1-4 μg/ml against more than 20 types of gram-positive bacteria, including multiple strains of MRSA, MDR *Streptococcus pneumoniae*, *E. faecalis*, and the anaerobic bacteria *Clostridioides difficile* and *Cutibacterium* acnes [49]. CBD is also active against *L. mono-cytogenes*, *E. faecalis*, and methicillin-resistant *S. epidermidis* (MRSE) [44]. THC and CBD are active against *Streptococcus pyogenes*, *Streptococcus milleri*, and *Streptococcus faecalis* [50]. CBD and CBDA are active against *S. epidermis* [8]; carvacrol is also active against *S. epidermis* [65]. CBD, α-pinene, β-pinene, β-myrcene, α-terpinolene, and β-caryophyllene are active against *S. epidermidis*, *L. monocytogenes*, *E. faecalis*, *E. faecium*, *E. hirae*, *B.*

subtilis, and *B. cereus* [38]. Naringenin is active against *H. pylori* [41]. CBC and its homologs, analogues and isomers are active against several bacteria, including *B. subtilis*, and *M. smegmatis*; CBC is also active against *S. cerevisiae* and *Trichophyton mentagrophytes*; many CBC homologs and isomers are also active against *T. mentagrophytes*, *C. albicans*, and *S. cerevisiae* [6]. α-Humulene is active against *C. Neoformans*, *C. Glabrata*, *C. Krusei* and *L. Donovani* [36], and against *C. bifermentans*, *E. hirae*, *E. faecium* and *S. salivarius*, *P. viridiflava*, *P. membranaefaciens*, *S. cerevisiae*, *S. japonicus*, and *Z. bailii* [70]. α-Pinene is active against *S. salivarius*, *C. tyrobutyricum*, *C. bifermentans*, *C. butyricum*, *C. sporogenes*, *E. hirae*, *E. faecium*, *P. savastanoi*, *P. carotovorum*, *P. corrugata*, *P. fluorescens*, *P. syringae*, *P. viridiflava*, *P. campestris*, *C. sake*, *K. marxianus*, *P. membranaefaciens*, *S. cerevisiae*, *Schizosaccharomyces pombe*, *S. japonicus*, *T. delbrueckii*, and *Z. bailii* [70]. β-Pinene, myrcene and carmagnola are active against *C. bifermentans*, *E. hirae*, *E. faecium*, *P. corrugata*, *P. fluorescens*, *P. viridiflava*, *C. sake*, *K. marxianus*, *P. membranaefaciens*, *S. pombe*, and *S. japonicus* [70]. β-Caryophyllene shows weak activity against *C. neoformans* [36]. The compound 1′ S-hydroxycannabinol is active against *L. donovani* and *P. falciparum* [21]. Other compounds isolated from cannabis with antimicrobial properties include 5-acetoxy-6-geranyl-3-n-pentyl-1, which is active against *L. donovani* and *Plasmodium falciparum*; cannflavin A, cannflavin C, and β-acetyl cannabispiranol, which are active against *L. donovani*; and 6-prenylapigenin, which is active against *P. falciparum*, *L. donovani*, and *C. albicans* [46]. Other anti-microbial cannabinoids that were recently discovered include (±)-3″-hydroxy-Δ($^{4″,5″}$)-cannabichromene, which is active against *C. albicans* and *C. krusei*; 4-acetoxy-2-geranyl-5-hydroxy-3-n-pentylphenol, active against C. krusei; 8-hydroxycannabinol, which is active against *C. albicans* and *M. intracellulare*; 8-hydroxycannabinolic acid A, which is active against *C. krusei* and *E. coli*; (±)-4-acetoxycannabichromene and 5-acetyl-4-hydroxycannabigerol, which are both active against *L. donovani* and *P. falciparum*; and (±)-3″-hydroxy-Δ($^{4″5″}$)-cannabichromene and 4-acetoxy-2-geranyl-5-hydroxy-3-n-pentylphenol, which are both active against *L. donovani* [33].

The ability to target a wide range of pathogenic microorganisms must be contrasted against the potential to damage the microbiome, as "selective inhibition is of the utmost importance for the maintenance of healthy gut microbiota" [47]. Cannabis has promise in this regard, too. Cannabis extract displayed "no inhibitory effects on the growth of probiotic strains" *Lactobacillus paracasei*, *Lactobacillus reuteri*, *Lactobacillus brevis*, *Lactobacillus plantarum*, *Bifidobacterium bifidum*, *Bifidobacterium longum* and *Bifidobacterium breve* [47].

1.4. Mechanism of action and synergistic interactions

Cannabis is also desirable as a potential antibiotic source for its ability to engage multiple bacterial targets and synergistic interactions. Synergistic interactions that potentiate antimicrobial effects are an evolutionary strategy against microorganisms [69, 71]. Many successful antibiotics engage multiple bacterial targets, structures, or processes, and typically resistance takes longer to emerge when multiple targets are engaged [69]. Combination antibiotic therapies often utilize antibiotic synergies.

Plants produce an abundance of secondary metabolites that often rely on synergistic combinations [72]. Multi-target engagement is also common among plants, and the essential oils and secondary metabolites produced by plants target microorganisms in multiple ways that affect their pathological processes [73-76]. Essentially, plant secondary metabolites often work together synergistically and engage multiple targets using different mechanisms of action [74, 75].

Common mechanisms of action include the disruption of cytoplasmic membrane function and structure (including the efflux system), interaction with the membrane proteins (ATPases and others), interruption of DNA/RNA synthesis and function, destabilization of the proton motive force with leakage of ions, prevention of enzyme synthesis, induction of coagulation of cytoplasmic constituents, and interruption of normal cell communications (quorum sensing) [74, 75].

The high diversity of cannabis chemovars and antimicrobial metabolites increases the likelihood of antimicrobial synergies [38]. Additionally, cannabis produces many unique secondary metabolites that are known to attack multiple targets in *S. aureus* and MRSA. CBD is particularly notable for multiple target engagement in *S. aureus*. CBD sharply inhibits protein, DNA, RNA, peptidoglycan and lipid synthesis [49]. CBD is active against MRSA biofilms [40, 49] and causes depolarization of the cytoplasmic membrane [44, 49]. Additionally, CBD shows low resistance frequency and has a low propensity to induce resistance [49]. At high concentrations, many other cannabinoids are active against biofilms, including CBG, CBN, CBC, CBCA, Δ9-THC, Δ8-THC, exo-olefin THC, Δ^9-THCA, THCV, CBGA, CBDV, CBDA, and +/- 11-OH Δ^9-THC [40]. CBCA induces rapid degradation of the bacterial lipid membrane and bacterial nucleoid [32]. CBG targets the cytoplasmic membrane; represses biofilm formation and eradicates preformed biofilms; kills persisters by rapidly eradicating them to below detection thresholds within 30 minutes of treatment; and shows no resistance development after being challenged for spontaneous resistance mutations [39].

Many of the non-cannabinoid secondary metabolites common in cannabis also engage multiple targets in *S. aureus* and MRSA. Carvacrol affects the lipid bilayer of bacterial cytoplasmic membranes causing loss of integrity and collapse of proton motive force, which results in a leakage of cellular material [67], reduces [66, 67] and eradicates biofilms [66], and is active against dual-species biofilms [68]. Myrcene acts synergistically with many essential oil components against *S. aureus* [54]. 1,8-cineole, α-terpineol, and terpinen-4-ol are active on the cytoplasmic membrane causing predisposition to lysis, loss of 260 nm absorbing material, altered morphology, and loss of tolerance to NaCl [58]. Linoleic acid inhibits the efflux pump and is synergistic with erythromycin [56]. Naringenin inhibits the growth of *S. aureus*, disrupts the cytoplasmic membrane, and affects the expression of fatty-acid synthesizing genes [63]. At high levels, naringenin damages the cytoplasmic membrane and interacts with DNA by changing conformations and molecular morphology [62]. Levomenol and nerolidol enhance membrane permeability, thereby increasing susceptibility of *S. aureus* to many common antibiotics [51]. Thymol inhibits biofilm formation and eradicates biofilms [66]. Quercetin can decrease the

proton-motive force [76]. Caffeic acid is active on efflux pumps by inhibiting the MrsA pumps of the *S. aureus* strain RN-4220 and the NorA pump of *S. aureus* strain 1199B [55].

1.5. Antibiotic resistance

Resistance development is another consequence of conventional antibiotic therapies. The multitude of antimicrobial secondary metabolites in cannabis may help prevent or delay resistance. One of the benefits that is often argued of combination therapy is that the simultaneous use of multiple antibiotics can delay resistance development [77-80]. Increasing the number of drugs used in combinations could be an effective strategy as high-order combinations can potentially slow resistance development [81].

With dozens of antimicrobial secondary metabolites that are active against MRSA, cannabis is a potential source for high-order antimicrobial combinations. Moreover, the bioactivity of the antimicrobial secondary metabolites found in plants typically does not confer resistance, and the use of antimicrobial plant extracts is relatively effective at preventing and reducing resistance [72].

Recent studies suggest that resistance is unlikely to development to either CBD [49] or CBG [39]. Although some species of bacteria are capable of developing resistance to some essential oils, resistance development to essential oils is generally rare and the development of resistance may be dependent upon oil composition and species of bacteria [82].

To further prevent or counter resistance development, different chemovars, with different compositions of antimicrobial secondary metabolites, could be easily substituted in a manner similar to cycling and mixing strategies, which rely on switching antibiotic regimens on time intervals or a per-patient basis in order to reduce selective pressure for resistance development.

1.6. The entourage effect

The term "entourage effect" is sometimes used to describe the complex interactions and variety of effects that "inactive" compounds found in cannabis are thought to have on active compounds [9, 12-14, 18, 19]. Many studies suggest therapeutic synergies in cannabis, and it is often observed that the effects of the entire plant are greater than the effects of individual components [9, 12-14, 18, 19, 83, 84]. In addition to these studies, many consumers of cannabis attribute different physiological effects to different chemovars [9, 11, 12, 19].

Although some chemovars might be inappropriate for some patients, with more than 700 chemovars available, the diversity of chemovars makes it likely that a chemovar with an appropriate balance of secondary metabolites could be found or bred for specific conditions and most patients [12, 16, 34].

Cannabis could potentially be used to treat multiple conditions simultaneously. Although in vivo studies and clinical trials are needed to determine if cannabis or cannabis secondary metabolites are suitable for treating *S. aureus* and MRSA infections, cannabis is already recognized for its analgesic properties [9, 12, 31]. Should cannabis prove suitable for *S. aureus* and MRSA treatment, it could potentially be used to treat both the infection and associated pain, possibly eliminating the need for a separate analgesic. Clinical trials could also be conducted to determine if cannabis could replace multiple drugs when treating *S. aureus* infection presenting with SEB-induced ARDS, as THC is a potent anti-inflammatory that halts the cytokine storm caused by the overactive immune response to SEB [45].

Additionally, medicinal plants and plant-based antimicrobials are generally less expensive and easier to obtain than synthetic drugs, and can be combined with other antibiotics to reduce testing and development costs [72].

1.7. Pharmacokinetic profile

Pharmacokinetic considerations are an important factor of antimicrobial therapy. Penetration differences between antibiotics administered simultaneously can accelerate the development of MDR by allowing for the stepwise accumulation of mutations, which can lead to an increase in both the rate of mutation acquisition and the rate of selection for pre-existing mutations [85-87].

Cannabis has desirable pharmacokinetic properties. Many of the antimicrobial secondary metabolites of cannabis share similar penetration profiles, which could allow multiple antimicrobial compounds to penetrate many in vivo environments. Cannabinoids cross the blood-brain barrier and the placental barrier, are present in breast milk, reduce inflammation and can penetrate *S. aureus* biofilms [6, 12, 18, 23, 45, 83]. Terpenoids, flavonoids, anthocyanins, and many other secondary metabolites present in cannabis are known to reduce inflammation, cross the blood-brain barrier, and destroy *S. aureus* biofilms [12, 18, 19, 22, 66-68, 88].

1.8. Toxicity, drug-drug interactions

Cannabis generally poses a low risk for toxicity, which is another important consideration with antibiotic therapy. On a population scale, cannabis has as estimated margin of exposure, which is the ratio of toxicological threshold to estimated human intake, in excess of 10,000 [89] and many studies have concluded that it is nearly impossible to consume lethal quantities of either cannabis or THC [89-92]. Further evidence of the low toxicity of cannabis can be found in the fact that there are no reported deaths from cannabis overdose [9, 26, 93] despite the fact that cannabis is used globally by more than 100 million people [93]. The absence of overdose deaths from cannabis is likely due to the lack of CB1 receptors in the brainstem cardiorespiratory centres, causing minimal interaction in areas of the brain involved in respiration [9, 26, 31, 94]. CBD has low toxicity when tested against red blood cells and keratinocytes [8].

In addition to low toxicity, many of the adverse effects associated with cannabis tend to decrease with tolerance or can be mitigated by other constituents of the plant. Cardiovascular effects of cannabis, such as tachycardia and increased blood pressure, are minimal or transient, and subside with tolerance [92]. Tolerance to the psychoactive effects of cannabis typically develops over several days; however, tolerance generally does not develop to the medical benefits, allowing patients to maintain dose consistency for many years [26]. The side effects of THC are mitigated by other secondary metabolites present in cannabis, and natural cannabis causes fewer psychological side effects than synthetic THC [95]. Numerous studies demonstrate the antipsychotic properties of

CBD and suggest that CBD can counter or mitigate many of the adverse psychoactive effects of THC [16, 18, 25, 95-98]. CBD has been observed to reduce anxiety, tachycardia, hunger, and sedation [25]. CBD is also being studied as a potential treatment for psychosis and schizophrenia [18, 98-101]. Moreover, not all chemovars of cannabis contain THC. For example, hemp, which is grown primarily for CBD and industrial applications, has very low quantities of THC, usually less than 0.3% [17].

Cannabis generally does not decrease effectiveness of concomitant medications and significant drug interactions, though rare, are typically associated with concurrent use of depressants [26]. However, many cannabinoids are known to interact with enzymes [20, 23, 25] and drug-drug interactions due to cytochrome P450 (CYP450) inhibition could occur [20, 23, 25, 102]. CBD inhibition of CYP450 has been associated with adverse drug events and has the potential to cause pharmacokinetic and pharmacodynamic drug-drug interactions [102]. There is no evidence that cannabis increases overdose lethality from other drugs [92].

2. Conclusions

In vitro studies and animal model studies suggest cannabis and its secondary metabolites as a potential source for new antimicrobial compounds against *S. aureus* and MRSA strains. However, more research is needed to understand the complex pharmacokinetics and pharmacodynamics of cannabis and its secondary metabolites before effective antimicrobial therapies can be developed.

Specifically, *in vivo* studies are needed to determine penetration profiles, screen for drug-drug interactions, and to test suitability for treatment of systemic infection. Given the high number of cannabis secondary metabolites active against *S. aureus* and MRSA strains, the multiplicity of antimicrobial mechanisms of action, and the potential for synergistic interactions, cannabis secondary metabolites should also be studied as antibiotic combinations and as possible adjuvants to be administered alongside conventional antibiotics. In addition to possible use in antibiotic combination therapy, the low potential for overdose and the generally safe profile of cannabis could allow cannabis-based therapies to be administered at high doses. Furthermore, the diversity among cannabis chemovars could allow for other antibiotic strategies such as cycling and mixing. Further research is necessary.

Due to the number of antimicrobial secondary metabolites produced by cannabis and the diversity of cannabis chemovars, it could also be possible to cross-breed a chemovar that produces specific antimicrobial secondary metabolites that are active against a target pathogen. If the antibiotic potential of cannabis is confirmed upon further testing, cannabis could provide an inexpensive and abundant source of new antibiotics for the developing world.

Appendix

The following appendix contains a list of approximately 50 cannabis secondary metabolites and derivatives that demonstrate antimicrobial activity against *S. aureus* and MRSA strains (Table 1). This list could be useful in cross breeding a chemovar that produces dozens of antimicrobial secondary metabolites active against *S. aureus*. This chemovar could serve as the starting point for the production and extraction of a full-spectrum oil that contains specific antimicrobial secondary metabolites, in consistent proportions, that target *S. aureus* and MRSA strains. Such a combination of secondary metabolites could effectively emulate the high-order combination strategy that evolved in plants. Ultimately, the secondary metabolites of cannabis, used wisely and in the correct proportions, could provide a new treatment strategy for MRSA that improves patient outcomes and minimizes

Table 1. Antimicrobial activity of cannabis secondary metabolites against different strains of *S. aureus*.

Cannabis secondary metabolite	Description of activity
Cannabinoids	
Cannabichromenic acid (CBCA)	MIC of 2 µg/ml against MRSA USA 300 [39, 40]. Bactericidal against MRSA at 3.9 µM and MSSA 34397 at 7.8 µM [32].
Cannabichromene (CBC):	Anti-inflammatory; 24 and 48 h MIC of 1.56 µg/ml against *S. aureus* ATCC 6538 [6]. Represses biofilm formation of MRSA USA 300; MIC of 8 µg/ml against MRSA USA 300 [39, 40]. Inhibits SA-1199B, RN-4220, ATCC 25923, EMRSA-15, and EMRSA-16 at 2 µg/ml; inhibits XU-212 at 1 µg/ml [30].
Cannabichromene-C0 (CBC homolog)	Anti-inflammatory; 24 and 48 h MIC of 12.5 µg/ml against *S. aureus* ATCC 6538 [6].
Cannabichromene-C1 (CBC homolog)	Anti-inflammatory; 24 and 48 h MIC of 3.12 µg/ml against *S. aureus* ATCC 6538 [6].
Isocannabichromene-C0	Anti-inflammatory; 24 and 48 h MIC of 12.5 µg/ml against *S. aureus* ATCC 6538 [6].
(±)-3″-hydroxy-$\Delta^{(4'',5'')}$-cannabichromene	IC_{50} against MRSA ATCC 35591 at 24.4 µM; IC_{50} against *S. aureus* ATCC 29213 29.6 µM [33].
Cannabidiolic acid (CBDA)	Inhibits SA-1199B, RN-4220, XU-212, ATCC 25923, EMRSA-15, and EMRSA-16 at 2 µg/ml [30]. MIC of 2 µg/ml against *S. aureus* ATCC 25923; MIC of 4 µg/ml against MRSA USA 300 [8]. Against MRSA USA 300, MIC of 16 µg/ml [39, 40]; inhibits biofilm formation [40].
Cannabidiol (CBD)	Engages multiple targets against *S. aureus*. Inhibits protein, DNA, RNA and peptidoglycan synthesis against *S. aureus* RN42200 at 2-3 µg/ml, rapidly shutting down synthesis pathways; reduces lipid synthesis at concentrations below MIC; membrane depolarization; MIC of 1-4 µg/ml against multiple strains MRSA, with similar MIC against VRSA; MIC90 against 132 MRSA and MSSA ATCC strains and Australian clinical isolates at 4 µg/ml; MIC50 and MIC_{90} of 1 µg/ml against an additional 50 MSSA and 50 MRSA USA-derived isolates; rapid bactericidal activity (<3 h) with minimum bactericidal concentration (MBC) of 2 µg/ml against MRSA ATCC 43300; able to penetrate and kill biofilms; minimum biofilm eradication concentration (MBEC) of 1-2 µg/ml against MSSA biofilms; MBEC of 2-4 µg/ml against MRSA biofilms; low innate resistance frequency value ($<3.78\times10^{-10}$) at 2x MIC against MRSA ATCC 43300); unlikely to induce resistance against MRSA ATCC 43300 (after 20 days of daily passage with 8 replications, 1.5-fold increase in MIC against CBD vs. 26-fold increase against daptomycin; non-toxic to human red blood cells, no signs of haemolysis up to 256 µg/ml; modest cytotoxicity against HEK-293 cells (human embryonic kidney), with CC_{50} around 200 µg/ml [49]. Causes depolarization of the cytoplasmic membrane against MRSA USA 300 at concentrations of 0.1 and 0.2 µg/ml; when combined with bacitracin, reduces MIC of BAC by 64-fold, and causes morphological changes including septa formations and membrane irregularities [44]. Bacteriostatic and bactericidal against *S. aureus* ATCC 6538 in nutrient broth agar between 1-5 µg/ml and in horse blood agar between 20-50 µg/ml [50]. Represses biofilm formation; MIC of 2 µg/ml against MRSA USA 300 [39, 40]. Inhibits SA-1199B, RN-4220, XU-212, EMRSA-15, and EMRSA-16 at 1 µg/ml; inhibits ATCC 25923 at 0.5 µg/ml [30]. MIC of 8 µg/ml against *S. aureus* ATCC 6538; MIC of 32 µg/ml against *S. aureus* 18As; MIC of 32 µg/ml against *S. aureus* 386 [38]. MIC against *S. aureus* ATCC 25923 and MRSA USA 300 at 1 µg/ml [8]. Against MRSA USA 300, a Canadian study published in 2021 reported CBD had an MIC value of 2.5 µg/ml and an MBC of 10 µg/ml; CBD powder had an inhibition zone of 11 mm and CBD oil had an inhibition zone of 9 mm [48].
Cannabidivarinic acid (CBDVA)	MIC of 32 µg/ml against MRSA USA 300 [40].
Cannabidivarin (CBDV)	MIC of 8 µg/ml against MRSA USA 300; inhibits biofilm formation [40].
Cannabidivarin methyl ester (CBDVM)	Bactericidal against MRSA at 15.6 µM [32].

Pre-Cannabigerol (Cannabigerolic-acid / CBGA)	Inhibits SA-1199B, XU-212, ATCC-25923, and EMRSA-16 at 4 µg/ml; inhibits RN-4220 and EMRSA-15 at 2 µg/ml [30]. MIC of 4 µg/ml against MRSA USA 300 [40].
Cannabigerol (CBG)	Active on the cytoplasmic membrane of gram-positive bacteria; represses biofilm formation of MRSA USA 300 by 50% at 0.5 µg/ml; MIC of 2 µg/ml against MRSA USA 300; eradicate preformed biofilms of MRSA USA 300 at 4 µg/ml; killed persisters in a concentration-dependent manner starting at 5 µg/ml; eradicated a population of ~108 CFU/ml MRSA persisters to below detection threshold within 30 minutes; MIC_{90} against 96 clinical isolates of MRSA ranged from 2-8 µg/ml, with one outlier isolate MIC_{90} of 0.0625; frequency of resistance less than 10^{-10} for MRSA; in vivo efficacy of CBG in systemic MRSA USA 300 mouse infection was comparable to vancomycin administered at a similar dose [39, 40]. Inhibits SA-1199B, RN-4220, XU-212, ATCC 25923, and EMRSA-16 at 1 µg/ml; inhibits EMRSA-15 at 2 µg/ml [30].
Cannabicyclol (CBL)	Represses biofilm formation against MRSA USA 300 [40].
4-Acetoxy-2-geranyl-5-hydroxy-3-n-pentylphenol (CGB - derivative)	IC_{50} against MRSA ATCC 35591 at 6.7 µM; IC50 against *S. aureus* ATCC 29213 12.2 µM [33].
5-Acetoxy-6-geranyl-3-n-pentyl-1,4-benzoquinone	IC_{50} against MRSA ATCC 43300 at 15 µg/ml [46].
5-Acetyl-4-hydroxycannabigerol	IC50 against MRSA ATCC 35591 at 53.4 µM [33].
+/- 11-OH Δ^9-THC	Represses biofilm formation of MRSA USA 300 [40].
Methylated Cannabigerol	Inhibits SA-1199B and XU-212 at 64 µg/ml [30].
Cannabinol (CBN)	Represses biofilm formation; MIC of 2 µg/ml against MRSA USA 300 [39, 40]. Inhibits SA-1199B, RN-4220, XU-212, ATCC-25923, and EMRSA-15 at 1 µg/ml [30].
8-Hydroxycannabinolic acid A	IC_{50} against *S. aureus* ATCC 29213 3.5 µM [33].
1′S-hydroxycannabinol	Active against MRSA ATCC 43300 at IC_{50} 10.0 µg/ml [21].
Carmagerol	Inhibits SA-1199B, RN-4220, and EMRSA-16 at 32 µg/ml; inhibits XU-212, ATCC 25923, and EMRSA-15 at 16 µg/ml [30].
(-)Δ8-tetrahydrocannabinol (-Δ8THC)	Against MRSA USA 300, MIC of 2 µg/ml; inhibits biofilm formation [39, 40].
Pre-Δ9 -Tetrahydrocannabinol (Δ^9-THCA)	Inhibits SA-1199B, XU-212, and EMRSA-15 at 8 µg/ml; inhibits RN-4220, ATCC 25923, and EMRSA-16 at 4 µg/ml [30]. Inhibits biofilm formation against MRSA USA 300 [40].
Δ9-tetrahydrocannabinolic acid A (THCAA)	MIC of 4 µg/ml against MRSA USA 300 [40].
Δ□-tetrahydrocannabinol (Δ9-THC)	Bacteriostatic and bactericidal against S. aureus ATCC 6538 in nutrient broth agar between 2-5 µg/ml and in horse blood agar between 20-50 µg/ml [50]. Represses biofilm formation; MIC of 2 µg/ml against MRSA USA 300 [39, 40]. Inhibits EMRSA-16 at 0.5 µg/ml; inhibits RN-4220, XU-212, and ATCC 25923 at 1 µg/ml; inhibits SA-1199B and EMRSA-15 at 2 µg/ml [30]. Protects mice from ARDS and toxicity post-SEB exposure by suppression of inflammatory cytokines and cessation of cytokine storm, attenuating SEB-mediated lung injury [45].
Tetrahydrocannabivarinic acid (THCVA)	MIC of 16 µg/ml against MRSA USA 300 [40].
Δ9-tetrahydrocannabivarin (THCV)	MIC of 4 µg/ml against MRSA USA 300 [40].
Exo-olefin THC	Represses biofilm formation; MIC of 2 µg/ml against MRSA USA 300 [39, 40].
Terpenes and Terpenoids	
Alpha-bisabolol (α-bisabolol; levomenol)	Sesquiterpenoid. Disrupts bacterial cell membranes; increases susceptibility of *S. aureus* ATCC 6538 to many common antibiotics [51].
Carvacrol	A secondary terpene found in some cultivars. Targets the lipid bilayer of bacterial cytoplasmic membranes. MIC values against 25 strains of *S. aureus* range from 0.015-0.03% (v/v) [65]. Effective against *S. aureus* 6-ME, 810-CT, 815-CT, 808-CT, 5-ME, and 74-CCH: MIC of 0.015-0.031% (v/v); MBC of 0.062-0.125% (v/v); BIC (biofilm inhibitory concentration) of 0.031-0.125% (v/v); BEC (biofilm eradication concentration) of 0.125-0.5% (v/v) [66]. In both liquid and vapour forms, causes significant reduction in biofilm biomass and cultivable cell numbers of *S. aureus* 815 [67]. Interferes with formation of dual-species biofilms consisting of *S. aureus* NCTC 10788/ Salmonella enterica serovar Typhimurium NCTC 74; total inhibition of dual-species biofilm at high doses [68].
Eugenol	Inhibits growth and cell viability of a variety of *S. aureus* strains. MIC of 10 µg/ml against *S. aureus* ATCC 13150, *S. aureus* ATCC 6538, *S. aureus* ATCC 25923, and *S. aureus* ATCC LB 126 [52].
Nerolidol	Disrupts bacterial cell membranes; increases susceptibility of *S. aureus* ATCC 6538 to many common antibiotics [51].
Limonene	A cyclic monoterpene. MIC of about 80 µg/ml and MBC of about 110 µg/ml against MRSA ATCC 43300: [53].
Para-Cymene (p-cymene)	MIC of about 50 µg/ml and MBC of about 100 µg/ml against MRSA ATCC 43300 [53].
Myrcene (β-myrcene)	Synergizes the antibiotic potency of other essential oil components against *S. aureus* and a number of other bacteria [54]. MIC of 8 µg/ml against *S. aureus* ATCC 6538 and *S. aureus* 18As; MIC of 32 µg/ml against *S. aureus* 386 [38].
Olivetol	Inhibits SA-1199B, RN-4220, XU-212, EMRSA-15, and EMRSA-16 at 64 µg/ml; inhibits ATCC 25923 at 128 µg/ml [30].
1,8-Cineole	Bacteriostatic and bactericidal against *S. aureus* [54, 57]. Against *S. aureus* NCTC 6571: MIC of 0.5% (v/v); MBC of 1 % (v/v) [57]. Causes predisposition to lysis, loss of NaCl tolerance, loss of 260-nm-absorbing material on *S. aureus* ATCC 9144 [58]. MIC of 250 µg/ml against MRSA samples obtained from Eskisehir Osmangazi University [64].
α-Pinene	Inhibits growth and cell viability of a variety of *S. aureus* strains. Against *S. aureus* ATCC 25923, MIC_{20} of 13.6 µg/ml [59]. MIC of 1.25-2.5% (v/v) against *S. aureus* NCTC 9518 [60]. MIC of 20 µg/ml against *S. aureus* ATCC 13150, *S. aureus* ATCC 6538, and *S. aureus* ATCC 25923; MIC of 10 µg/ml against *S. aureus* ATCC LB 126 [52]. Against *S. aureus* ATCC 6538, MIC of 4 µg/ml; against *S. aureus* 18As and *S. aureus* 386, MIC of 16 µg/ml [38]. MIC of 1000 µg/ml against MRSA samples obtained from Eskisehir Osmangazi University [64].
β-Pinene	Inhibits growth and cell viability of a variety of *S. aureus* strains. MIC of 20 µg/ml against *S. aureus* ATCC 13150, *S. aureus* ATCC 6538, *S. aureus* ATCC 25923, and *S. aureus* ATCC LB 126 [52]. MIC against *S. aureus* ATCC 6538 at 4 µg/ml; MIC *S. aureus* 18As at 32 µg/ml; MIC *S. aureus* 386 at 8 µg/ml [38].
α-Terpineol	MIC between 0.16-0.31% (v/v) against *S. aureus* NCTC 9518 [60]. Causes predisposition to lysis, loss of NaCl tolerance, loss of 260-nm-absorbing material on *S. aureus* ATCC 9144 [58].
α-Terpinolene	MIC against *S. aureus* ATCC 6538 at 8 µg/ml; MIC *S. aureus* 18As and *S. aureus* 386 at 32 µg/ml [38].
Terpinen-4-ol	MIC between 0.31-0.63% (v/v) against *S. aureus* NCTC 9518 [60]. Causes predisposition to lysis, loss of NaCl tolerance, loss of 260-nm-absorbing material; electron microscopy showed formation of mesosomes and loss of cytoplasmic contents on *S. aureus* ATCC 9144 [58].
Thymol	A monoterpene. MIC values against 25 strains of *S. aureus* range from 0.03-0.06% (v/v) [65]. Effective against *S. aureus* 6-ME, 810-CT, 815-CT, 808-CT, 5-ME, and 74-CCH: MIC of 0.031-0.062% (v/v); MBC of 0.062-0.125% (v/v); BIC (biofilm inhibitory concentration) of 0.062-0.125% (v/v); BEC (biofilm eradication concentration) of 0.125-0.250% (v/v) [66]. MIC of about 80 µg/ml and MBC of about 110 µg/ml against MRSA ATCC 43300 [53].
β-Caryophyllene	Against *S. aureus* ATCC 25923 MIC20 of 5.1 µg/ml [59]. Against *S. aureus* ATCC 6538, MIC of 16 µg/ml; *S. aureus* 18As and *S. aureus* 386, MIC of 32 µg/ml [38].
Humulene (α-Caryophyllene)	Against *S. aureus* ATCC 25923, MIC_{20} of 2.6 µg/ml [59].
β-amyrin	MIC of 2.5 mg/ml against *S. aureus* NCTC 7447 [61].
Flavonoids	
Cannflavin A	IC_{50} against MRSA ATCC 43300 at 15 µg/ml [46].
Naringenin	Against *S. aureus* ATCC 6538, disrupts the cytoplasmic membrane at low levels; at high levels, damages cytoplasmic membrane, causing leakage of intracellular substances; DNA targeting effects; MIC of 1.84 mM (0.50 g l^{-1}) [62]. Significantly reduces growth rate of S. aureus cells in the concentration range of 0 to 2.20 mM, with no growth detected within 14 h when concentration was 2.20 mM; disrupts the cytoplasmic membrane, affects the expression of fatty-acid synthesizing genes [63]. MIC of 512 µg/ml and an MBEC corresponding to 2048 µg/ml against *S. aureus* 105 [41].
Acids	
Caffeic Acid	Caffeic acid is active on efflux pumps, inhibiting the MrsA pumps of the *S. aureus* strain RN-4220 and the NorA pump of the *S. aureus* strain 1199B [55].
Linoleic Acid	Efflux pump inhibition against MRSA RN-4220/pUL5054. At 16 µg/ml, linoleic acid displayed synergistic effects with erythromycin, reducing MIC value of erythromycin from 256 to 16 µg/m [56].

the development of new resistances.

Other studies

A 1987 murine study showed that aqueous marijuana extract and marijuana smoke inhibited *S. aureus* NCTC 9789 [28].

A 1995 study at the University of Punjab found that cannabis extracts had strong inhibitory effect on *S. aureus* [29].

A 2001 study of 5 EOs from different cultivars of low-THC cannabis (SwissMix, Felina 34, Fedrina 74, Kompolti, and Secuemi)

found inhibitory zones against *S. aureus* to be 7.1, 14.4, 10.0, 5.2 and 9.6 mm, respectively [17].

A 2008 study of native and naturalized plants in Minnesota and Wisconsin found that cannabis extracts had an inhibition zone of 25 mm against *S. aureus* ATCC 12600 [42].

A 2011 study of cannabis extracts found strong antimicrobial activity against *S. aureus*. Inhibition zone diameter was positively correlated with extraction time (at 2 h, 8 h, and 18 h) and extraction method (acetone extraction vs. methanol extraction). Acetone extraction inhibition zones at 2, 8 and 18 h of extraction time were 12, 16 and 20 mm, respectively; methanol extraction inhibition zones were 10, 14 and 20 mm, respectively [43].

A 2012 cannabis study from Chinese Medicine tested cannabis seed oil, and cannabis petroleum ether and methanol extracts of the whole plant against a number of microorganisms, including *S. aureus* 25923. Seed oil had an inhibition zone of 28 mm, while petroleum ether extract had an inhibition zone of 23 mm and methanol extracts had an inhibition zone of 12 mm. MIC value of methanol extract of seed oil was 25 µg/ml; methanol extract of whole plant was 50 µg/ml [7].

A 2014 study from Hazara University found in vitro activity of C. sativa leaf extracts against *S. aureus* ATCC 6538. The average inhibition zone of cannabis extract was 10.3 mm [35].

A study published in 2016 in the Records of Natural Products demonstrated antibacterial activity of a number of volatile fractions isolated from high potency *C. sativa* oil. IC50 values for *S. aureus* ATCC 29213 and MRSA ATCC 33591 were obtained. The volatile oil was active against *S. aureus* at MIC50 of 44.71 µg/ml, and against MRSA at MIC50 of 98.79 µg/ml. Six subfractions demonstrated potential antibacterial activity against both *S. aureus* and MRSA, with IC50 values between 0.93 µg/ml and 19.9 µg/ml against *S. aureus* and between 0.82 µg/ml and 17.34 µg/ml against MRSA [36].

A study published in 2018 in the Journal of Integrative Medicine evaluated the efficiency of ethanolic extracts of *C. sativa*, *T. orientalis*, and *P. guajava* against 20 MRSA strains. Cannabis extracts were effective individually at inhibiting MRSA strains; however, profound synergism was observed when cannabis extract was combined with *T. orientalis* extract [37].

A study published in 2018 in Molecules demonstrated antibacterial potential of cannabis essential oil and naringenin against several strains of *S. aureus* (*S. aureus* ATCC 29213, *S. aureus* 101 TV, *S. aureus* 104, and *S. aureus* 105). Essential oil was tested on all strains for MIC, MBC, and MBEC. Against *S. aureus* ATCC 29213, *S. aureus* 101 TV, and *S. aureus* 104, MIC, MBC, and MBEC values were identical, with MIC of 8 mg/ml, MBC of 16 mg/ml and MBEC of 24 mg/ml. Against *S. aureus* 105 TV, MIC and MBC values were also reported at 8 and 16 mg/ml, respectively; MBEC was 16 mg/ml [41].

A study of EOs from different strains of fibre-type cannabis, published in Molecules in 2019, revealed 4 strains that inhibited *S. aureus* ATCC 6538 at MIC ranging from 2 to 16 µg/ml; 3 strains inhibited *S. aureus* 18As at MIC ranging from 16 to 32 µg/ml; and 3 strains that inhibited *S. aureus* 386 at MIC ranging from 16 to 32 µg/ml [38].

In 2020 a study published in LWT - Food Science and Technology, it was found that hemp seed extract had an MIC of 1 mg/ml against *S. aureus* ATCC 35556 and *S. aureus* ATCC 25923. Complete biofilm inhibition of *S. aureus* ATCC 35556 occurred at concentrations of 0.5 mg/ml and 1 mg/ml [47].

COMPETING INTERESTS

The author declares that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] S. Tong, J.S. Davis, E. Eichenberger, et al. (2015), "Staphylococcus aureus infections: Epidemiology, pathophysiology, clinical manifestations, and management", *Clinical Microbiology Reviews*, **28(3)**, pp.603-661, DOI: 10.1128/CMR.00134-14.

[2] A.P. Kourtis, K. Hatfield, J. Baggs, et al. (2019), "Vital signs: Epidemiology and recent trends in methicillin-resistant and in methicillin-susceptible Staphylococcus aureus bloodstream infections-United States", *Morbidity and Mortality Weekly Report*, **68(9)**, pp.214-219, DOI: 10.15585/mmwr.mm6809e1.

[3] World Health Organization (2014), *Antimicrobial Resistance: Global Report on Surveillance*.

[4] A. Shariati, M. Dadashi, Z. Chegini, et al. (2020), "The global prevalence of Daptomycin, Tigecycline, Quinupristin/Dalfopristin, and Linezolid-resistant *Staphylococcus aureus* and coagulase-negative staphylococci strains: A systematic review and meta-analysis", *Antimicrobial Resistance & Infection Control*, **9(56)**, pp.1-20, DOI: 10.1186/s13756-020-00714-9.

[5] M.S. Morehead, C. Scarbrough (2018), "Emergence of global antibiotic resistance", *Prim. Care.*, **45(3)**, pp.467-484, DOI: 10.1016/j.pop.2018.05.006.

[6] C.E. Turner, M.A. Elsohly (1981), "Biological activity of cannabichromene, its homologs and isomers", *J. Clin. Pharmacol.*, **21(S1)**, pp.283S-291S, DOI: 10.1002/j.1552-4604.1981.tb02606.x.

[7] E.M.M. Ali, A.Z.I. Almagboul, S.M.E. Khogali, et al. (2012), "Antimicrobial activity of *Cannabis sativa* L.", *Chin. Med.*, **3(1)**, pp.61-64, DOI: 10.4236/cm.2012.31010.

[8] L.D. Martinenghi, R. Jønsson, T. Lund, et al. (2020), "Isolation, purification, and antimicrobial characterization of cannabidiolic acid and cannabidiol from *Cannabis sativa* L.", *Biomolecules*, **10(6)**, DOI: 10.3390/biom10060900.

[9] L.E. Klumpers, D.L. Thacker (2019), "A brief background on cannabis: From plant to medical indications", *Journal of AOAC International*, **102(2)**, pp.412-420, DOI: 10.5740/jaoacint.18-0208.

[10] J. Sherma, F. Rabel (2019), "Thin layer chromatography in the analysis of Cannabis and its components and synthetic cannabinoids", *Journal of Liquid Chromatography & Related Technologies*, **42(19-20)**, pp.613-628, DOI: 10.1080/10826076.2019.1663529.

[11] J.K. Booth, J. Bohlmann (2019), "Terpenes in *Cannabis sativa* - From plant genome to humans", *Plant Science*, **284**, pp.67-72, DOI: 10.1016/j.plantsci.2019.03.022.

[12] E.P. Baron (2018), "Medicinal properties of cannabinoids, terpenes, and flavonoids in cannabis, and benefits in migraine, headache, and pain: An update on current evidence and cannabis science", *Headache: The Journal of Head and Face Pain*, **58(7)**, pp.1139-1186, DOI: 10.1111/head.13345.

[13] E.B. Russo (2019), "The case for the entourage effect and conventional breeding of clinical cannabis: No 'strain', no gain", *Frontiers in Plant Science*, **9**, DOI: 10.3389/fpls.2018.01969.

[14] M.A. Lewis, E.B. Russo, K.M. Smith (2018), "Pharmacological foundations of cannabis chemovars", *Planta Medica*, **84(4)**, pp.225-233, DOI: 10.1055/s-0043-122240.

[15] A. Hazekamp, J.T. Fischedick (2012), "Cannabis - from cultivar to chemovar", *Drug Testing And Analysis*, **4(7-8)**, pp.660-667, DOI: 10.1002/dta.407.

[16] B.A. Whittle, G.W. Guy, P. Robson (2001), "Prospects for new cannabis-based prescription medicines", *Journal of Cannabis Therapeutics*, **1(3-4)**, pp.183-205, DOI: 10.1300/J175v01n03_12.

[17] J. Novak, K.Z. Eglseer, S.G. Deans, et al. (2001), "Essential oils of different cultivars of *Cannabis sativa* L. and their antimicrobial activity", *Flavour and Fragrance Journal*, **16(4)**, pp.259-262, DOI: 10.1002/ffj.993.

[18] K.W. Green (2018), "The united chemicals of cannabis: Beneficial effects of cannabis phytochemicals on the brain and cognition", *Recent Advances in Cannabinoid Research*, DOI: 10.5772//intechopen.79266.

[19] G. Nahler, T.M. Jones, E.B. Russo (2019), "Cannabidiol and contributions of major hemp phytocompounds to the entourage effect, possible mechanisms", *Journal of Alternative Complementary & Integrative Medicine*, **5(2)**, DOI: 10.24966/ACIM-7562/100070.

[20] C.M. Andre, J.F. Hausman, G. Guerriero (2016), "Cannabis sativa: The plant of the thousand and one molecules", *Frontiers in Plant Science*, **7(19)**, DOI: 10.3389/fpls.2016.00019.

[21] S.A. Ahmed, S.A. Ross, D. Slade, et al. (2015), "Minor oxygenated cannabinoids from high potency *Cannabis sativa* L.", *Phytochemistry*, **117**, pp.194-199, DOI: 10.1016/j.phytochem.2015.04.007.

[22] R. Gallily, Z. Yekhtin, L.O. Hanuš (2018), "The anti-inflammatory properties of terpenoids from cannabis", *Cannabis and Cannabinoid Research*, **3(1)**, pp.282-290, DOI: 10.1089/can.2018.0014.

[23] B.C. Foster, H. Abramovici, C.S. Harris (2019), "Cannabis and cannabinoids: Kinetics and interactions", *The American Journal of Medicine*, **132(11)**, pp.1266-1270, DOI: 10.1016/j.amjmed.2019.05.017.

[24] L.O. Hanuš, S.M. Meyer, E. Muñoz, et al. (2016), "Phytocannabinoids: A unified critical inventory", *Natural Product Reports*, **33(12)**, pp.1357-1392, DOI: 10.1039/c6np00074f.

[25] E.B. Russo, J. Marcu (2017), "Cannabis pharmacology: The usual suspects and a few promising leads", *Advances in Pharmacology*, **80**, pp.67-134, DOI: 10.1016/bs.apha.2017.03.004.

[26] C.A. MacCallum, E.B. Russo (2018), "Practical considerations in medical cannabis administration and dosing", *European Journal of Internal Medicine*, **49**, pp.12-19, DOI: 10.1016/j.ejim.2018.01.004.

[27] M.A. ElSohly, D. Slade (2005), "Chemical constituents of marijuana: The complex mixture of natural cannabinoids", *Life Sciences*, **78(5)**, pp.539-548, DOI: 10.1016/j.lfs.2005.09.011.

[28] M.K. Ashfaq, E.S. Watson, H.N. ElSohly (1987), "The effect of subacute marijuana smoke inhalation on experimentally induced dermonecrosis by *S. aureus* infection", *Immunopharmacology and Immunotoxicology*, **9(2-3)**, pp.319-331, DOI: 10.3109/08923978709035217.

[29] K. Wasim, I. Haq, M. Ashraf (1995), "Antimicrobial studies of the leaf of *cannabis sativa* L.", *Pakistan Journal of Pharmaceutical Sciences*, **8(1)**, pp.29-38.

[30] G. Appendino, S. Gibbons, A. Giana, et al. (2008), "Antibacterial cannabinoids from Cannabis sativa: A structure-activity study", *Journal of Natural Products*, **71(8)**, pp.1427-1430.

[31] G. Appendino, G. Chianese, O. Taglialatela-Scafati (2011), "Cannabinoids: Occurrence and medicinal chemistry", *Current Medicinal Chemistry*, **18(7)**, pp.1085-1099, DOI: 10.2174/092986711794940888.

[32] M. Galletta, T.A. Reekie, G. Nagalingam, et al. (2020), "Rapid antibacterial activity of cannabichromenic acid against methicillin-resistant *Staphylococcus aureus*", *Antibiotics*, **9(8)**, DOI: 10.3390/antibiotics9080523.

[33] M.M. Radwan, M.A. Elsohly, D. Slade, et al. (2009), "Biologically active cannabinoids from high-potency *Cannabis sativa*", *Journal of Natural Products*, **72(5)**, pp.906-911, DOI: 10.1021/np900067k.

[34] E.B. Russo (2011), "Taming THC: Potential cannabis synergy and phytocannabinoid terpenoid entourage effects", *British Journal of Pharmacology*, **163(7)**, pp.1344-1364, DOI: 10.1111/j.1476-5381.2011.01238.x.

[35] M. Naveed, T. Khan, I. Ali, et al. (2014), "*In vitro* antibacterial activity of *Cannabis sativa* leaf extracts to some selective pathogenicbacterial strains", *Int. J. Biosci.*, **4(4)**, pp.65-70.

[36] A.S. Wanas, M.M. Radwan, Z. Mehmedic, et al. (2016), "Antifungal activity of the volatiles of high potency *Cannabis sativa* L. against *Cryptococcus neoformans*", *Records of Natural Products*, **10(2)**, pp.214-220.

[37] S. Chakraborty, N. Afaq, N. Singh, et al. (2018), "Antimicrobial activity of *Cannabis sativa*, *Thuja orientalis* and *Psidium guajava* leaf extracts against methicillin-resistant *Staphylococcus aureus*", *Journal of Integrative Medicine*, **16(5)**, pp.350-357, DOI: 10.1016/j.joim.2018.07.005.

[38] R. Iseppi, V. Brighenti, M. Licata, et al. (2019), "Chemical characterization and evaluation of the antibacterial activity of essential oils from fibre-type *Cannabis sativa* L. (Hemp)", *Molecules*, **24(12)**, DOI: 10.3390/molecules24122302.

[39] M.A. Farha, O.M.E. Halfawy, R.T. Gale, et al. (2020a), "Uncovering the hidden antibiotic potential of Cannabis", *ACS Infect. Dis.,* **6(3)**, pp.338-346, DOI: 10.1021/acsinfecdis.9b00419.

[40] M.A. Farha, O.M.E. Halfawy, R.T. Gale, et al. (2020), "Uncovering the hidden antibiotic potential of Cannabis", *ACS Infect. Dis.,* **6(3)**, pp.338-346, DOI: 10.1021/acsinfecdis.9b00419.

[41] G. Zengin, L. Menghini, A.D. Sotto, et al. (2018), "Chromatographic analyses, *in vitro* biological activities, and cytotoxicity of *Cannabis sativa* L. Essential oil: A multidisciplinary study", *Molecules*, **23(12)**, DOI: 10.3390/molecules23123266.

[42] J.R. Borchardt, D.L. Wyse, C.C. Sheaffer, et al. (2008), "Antimicrobial activity of native and naturalized plants of Minnesota and Wisconsin", *J. Med. Plants Res.*, **2(5)**, pp.98-110.

[43] V.N. Mkpenie, E. Essien, I.I. Udoh (2012), "Effect of extraction conditions on total polyphenol contents, antioxidant and antimicrobial activities of *Cannabis sativa* L.", *Electron. J. Environ. Agric. Food Chem.*, **11(4)**, pp.300-307.

[44] C.S. Wassmann, P. Højrup, J.K. Klitgaard (2020), "Cannabidiol is an effective helper compound in combination with bacitracin to kill Gram-positive bacteria", *Scientific Reports*, **10(1)**, DOI: 10.1038/s41598-020-60952-0.

[45] A. Mohammed, H. Alghetaa, M. Sultan, et al. (2020), "Administration of Δ9 tetrahydrocannabinol (THC) post staphylococcal enterotoxin B exposure protects mice from acute respiratory distress syndrome and toxicity", *Frontiers in Pharmacology*, **11(893)**, DOI: 10.3389/fphar.2020.00893.

[46] M.M. Radwan, M.A. Elsohly, D. Slade, et al. (2008), "Non-cannabinoid constituents from a high potency *Cannabis sativa* variety", *Phytochemistry*, **69(14)**, pp.2627-2633, DOI: 10.1016/j.phytochem.2008.07.010.

[47] S. Frassinetti, M. Gabriele, E. Moccia, et al. (2020), "Antimicrobial and antibiofilm activity of *Cannabis sativa* L. seeds extract against *Staphylococcus aureus* and growth effects on probiotic *Lactobacillus* spp.", *LWT*, **124**, DOI: 10.1016/j.lwt.2020.109149109149.

[48] K.N. Tahsin, D. Watson, A. Rizkalla, et al. (2021), "Antimicrobial studies of Cannabidiol as biomaterials against superbug MRSA", *CMBES Proceedings*, **44**, pp.1-10

[49] M.A.T. Blaskovich, A.M. Kavanagh, A.G. Elliott, et al. (2021), "The antimicrobial potential of cannabidiol", *Communications Biology*, **4(1)**, DOI: 10.1038/s42003-020-01530-y.

[50] B.V. Klingeren, M.T. Ham (1976), "Antibacterial activity of Δ9-tetrahydrocannabinol and cannabidiol", *Antonie Van Leeuwenhoek*, **42(1)**, pp.9-12, DOI: 10.1007/BF00399444.

[51] B.F.B. Stecher, E.A. Johnson (2003), "Sensitization of *Staphylococcus aureus* and *Escherichia coli* to antibiotics by the sesquiterpenoids nerolidol, farnesol, bisabolol, and apritone", *Antimicrobial Agents and Chemotherapy*, **47(10)**, pp.3357-3360, DOI: 10.1128/AAC.47.10.3357-3360.2003.

[52] A.M. Leite, E.O. Lima, E.L. Souza, et al. (2007), "Inhibitory effect of beta-pinene, alpha-pinene and eugenol on the growth of potential infectious endocarditis causing Gram-positive bacteria", *Revista Brasileira de Ciências Farmacêuticas*, **43(1)**, pp.121-126, DOI: 10.1590/S1516-93322007000100015.

[53] H. Li, T. Yang, F.Y. Li, et al. (2014), "Antibacterial activity and mechanism of action of *Monarda punctata* essential oil and its main components against common bacterial pathogens in respiratory tract", *International Journal of Clinical and Experimental Pathology*, **7(11)**, pp.7389-7398.

[54] F. Grotenhermen, E.B. Russo (2006), *Handbook of Cannabis Therapeutics: From Bench to Bedside*, Haworth Integrative Healing Press, 490pp.

[55] J.F.S.D. Santos, S.R. Tintino, T.S. Freitas, et al. (2018), "*In vitro* e *in silico* evaluation of the inhibition of *Staphylococcus aureus* efflux pumps by caffeic and gallic acid", *Comparative Immunology, Microbiology and Infectious Diseases*, **57**, pp.22-28, DOI: 10.1016/j.cimid.2018.03.001.

[56] B.C.L. Chan, X.Q. Han, S.L. Lui, et al. (2015), "Combating against methicillin-resistant *Staphylococcus aureus* - two fatty acids from *Purslane* (*Portulaca oleracea* L.) exhibit synergistic effects with erythromycin", *Journal of Pharmacy and Pharmacology*, **67(1)**, pp.107-116, DOI: 10.1111/jphp.12315.

[57] C.F. Carson, T.V. Riley (1995), "Antimicrobial activity of the major components of the essential oil of *Melaleuca alternifolia*", *Journal of Applied Bacteriology*, **78(3)**, pp.264-269, DOI: 10.1111/j.1365-2672.1995.tb05025.x.

[58] C.F. Carson, B.J. Mee, T.V. Riley (2002), "Mechanism of action of *Melaleuca alternifolia* (tea tree) oil on *Staphylococcus aureus* determined by time-kill, lysis, leakage, and salt tolerance assays and electron microscopy", *Antimicrobial Agents and Chemotherapy*, **46(6)**, pp.1914-1920, DOI: 10.1128/AAC.46.6.1914-1920.2002.

[59] A. Pichette, P.L. Larouche, M. Lebrun, et al. (2006), "Composition and antibacterial activity of *Abies balsamea* essential oil", *Phytotherapy Research: An International Journal Devoted to Pharmacological and Toxicological Evaluation of Natural Product Derivatives*, **20(5)**, pp.371-373, DOI: 10.1002/ptr.1863.

[60] A. Raman, U. Weir, S.F. Bloomfield (1995), "Antimicrobial effects of tea tree oil and its major components on *Staphylococcus aureus*, *Staph. epidermidis* and *Propionibacterium acnes*", *Letters in Applied Microbiology*, **21(4)**, pp.242-245.

[61] N.A. Raouf, N.M. Al-Enazi, A.A. Al-Homaidan, et al. (2015), "Antibacterial β-amyrin isolated from *Laurencia microcladia*", *Arabian Journal of Chemistry*, **8(1)**, pp.32-37, DOI: 10.1016/j.arabjc.2013.09.033.

[62] L.H. Wang, M.S. Wang, X.A. Zeng, et al. (2017), "Membrane and genomic DNA dual-targeting of citrus flavonoid naringenin against *Staphylococcus aureus*", *Integrative Biology*, **9(10)**, pp.820-829, DOI: 10.1039/C7IB00095B.

[63] L.H. Wang, X.A. Zeng, M.S. Wang, et al. (2018), "Modification of membrane properties and fatty acids biosynthesis-related genes in *Escherichia coli* and *Staphylococcus aureus*: Implications for the antibacterial mechanism of naringenin", *Biochimica et Biophysica Acta (BBA)-Biomembranes*, **1860(2)**, pp.481-490, DOI: 10.1016/j.bbamem.2017.11.007.

[64] G. Özek, F. Demirci, T. Ozek, et al. (2010), "Gas chromatographic-mass spectrometric analysis of volatiles obtained by four different techniques from *Salvia rosifolia* Sm., and evaluation for biological activity", *Journal of Chromatography A*, **1217(5)**, pp.741-748, DOI: 10.1016/j.chroma.2009.11.086.

[65] A. Nostro, A.R. Blanco, M.A. Cannatelli, et al. (2004), "Susceptibility of methicillin-resistant staphylococci to oregano essential oil, carvacrol and thymol", *FEMS Microbiology Letters*, **230(2)**, pp.191-195, DOI: 10.1016/S0378-1097(03)00890-5.

[66] A. Nostro, A.S. Roccaro, G. Bisignano, et al. (2007), "Effects of oregano, carvacrol and thymol on *Staphylococcus aureus* and *Staphylococcus epidermidis* biofilms", *Journal of Medical Microbiology*, **56(4)**, pp.519-523, DOI: 10.1099/jmm.0.46804-0.

[67] A. Nostro, A. Marino, A.R. Blanco, et al. (2009), "*In vitro* activity of carvacrol against staphylococcal preformed biofilm by liquid and vapour contact", *Journal of Medical Microbiology*, **58(6)**, pp.791-797, DOI: 10.1099/jmm.0.009274-0.

[68] J.R. Knowles, S. Roller, D.B. Murray, et al. (2005), "Antimicrobial action of carvacrol at different stages of dual-species biofilm development by *Staphylococcus aureus* and *Salmonella enterica* serovar Typhimurium", *Applied and Environmental Microbiology*, **71(2)**, pp.797-803, DOI: 10.1128/AEM.71.2.797-803.2005

[69] M. Tyers, G.D. Wright (2019), "Drug combinations: A strategy to extend the life of antibiotics in the 21st century", *Nature Reviews Microbiology*, **17(3)**, pp.141-155, DOI: 10.1038/s41579-018-0141-x.

[70] L. Nissen, A. Zatta, I. Stefanini, et al. (2010), "Characterization and antimicrobial activity of essential oils of industrial hemp varieties (*Cannabis sativa* L.)", *Fitoterapia*, **81(5)**, pp.413-419, DOI: 10.1016/j.fitote.2009.11.010.

[71] G.L. Challis, D.A. Hopwood (2003), "Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces species*", *Proceedings of the National Academy of Sciences*, **100(2)**, pp.14555-14561, DOI: 10.1073/pnas.1934677100.

[72] M.J. Cheesman, A. Ilanko, B. Blonk, et al. (2017), "Developing new antimicrobial therapies: Are synergistic combinations of plant extracts/compounds with conventional antibiotics the solution?", *Pharmacognosy Reviews*, **11(22)**, pp.57-72, DOI: 10.4103/phrev.phrev_21_17.

[73] P.D. Gupta, T.J. Birdi (2017), "Development of botanicals to combat antibiotic resistance", *Journal of Ayurveda and Integrative Medicine*, **8(4)**, pp.266-275, DOI: 10.1016/j.jaim.2017.05.004.

[74] F. Nazzaro, F. Fratianni, L.D. Martino, et al. (2013), "Effect of essential oils on pathogenic bacteria", *Pharmaceuticals*, **6(12)**, pp.1451-1474, DOI: 10.3390/ph6121451.

[75] C.L. Gorlenko, H.Y. Kiselev, E.V. Budanova, et al. (2020), "Plant secondary metabolites in the battle of drugs and drug-resistant bacteria: New heroes or worse clones of antibiotics?", *Antibiotics*, **9(4)**, DOI: 10.3390/antibiotics9040170.

[76] I. Górniak, R. Bartoszewski, J. Króliczewski (2019), "Comprehensive review of antimicrobial activities of plant flavonoids", *Phytochemistry Reviews*, **18(1)**, pp.241-272, DOI: 10.1007/s11101-018-9591-z.

[77] G.J. Sullivan, N.N. Delgado, R. Maharjan, et al. (2020), "How antibiotics work together: Molecular mechanisms behind combination therapy", *Current Opinion in Microbiology*, **57**, pp.31-40, DOI: 10.1016/j.mib.2020.05.012

[78] P.R. Gonzales, M.W. Pesesky, R. Bouley, et al. (2015), "Synergistic, collaterally sensitive β-lactam combinations suppress resistance in MRSA", *Nature Chemical Biology*, **11(11)**, pp.855-861, DOI: 10.1038/nchembio.1911.

[79] X. Zheng, A.D. Berti, S. McCrone, et al. (2018), "Combination antibiotic exposure selectively alters the development of vancomycin intermediate resistance in *Staphylococcus aureus*", *Antimicrobial Agents and Chemotherapy*, **62(2)**, DOI: 10.1128/aac.02100-17.

[80] B. Raymond (2019), "Five rules for resistance management in the antibiotic apocalypse, a road map for integrated microbial management", *Evolutionary Applications*, **12**, pp.1079-1091, DOI: 10.1111/eva.12808.

[81] K. Yilancioglu, M. Cokol (2019), "Design of high-order antibiotic combinations against *M. tuberculosis* by ranking and exclusion", *Scientific Reports*, **9**, DOI: 10.1038/s41598-019-48410-y.

[82] R. Becerril, C. Nerín, R. Gómez-Lus (2012), "Evaluation of bacterial resistance to essential oils and antibiotics after exposure to oregano and cinnamon essential oils", *Foodborne Pathogens and Disease*, **9(8)**, pp.699-705, DOI: 10.1089/fpd.2011.1097.

[83] J.M. McPartland, P.L. Pruitt (1999), "Side effects of pharmaceuticals not elicited by comparable herbal medicines: The case of tetrahydrocannabinol and marijuana", *Alternative Therapies in Health and Medicine*, **5(4)**, pp.57-62.

[84] J.W. Fairbairn, J.T. Pickens (1981), "Activity of cannabis in relation to its delta'-trans-tetrahydro-cannabinol content", *British Journal of Pharmacology*, **72(3)**, pp.401-409, DOI: 10.1111/j.1476-5381.1981.tb10990.x.

[85] S.M. Gamez, A.L. Hill, D.I.S. Rosenbloom, et al. (2015), "Imperfect drug penetration leads to spatial monotherapy and rapid evolution of multidrug resistance", *Proceedings of the National Academy of Sciences*, **112(22)**, pp.E2874-E2883, DOI: 10.1073/pnas.142418411.

[86] Q. Zhang, G. Lambert, D. Liao, et al. (2011), "Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments", *Science*, **333(6050)**, pp.1764-1767, DOI: 10.1126/science.1208747.

[87] R. Hermsen, J.B. Deris, T. Hwa, et al. (2012), "On the rapidity of antibiotic resistance evolution facilitated by a concentration gradient", *Proceedings of The National Academy of Sciences*, **109(27)**, pp.10775-10780, DOI: 10.1073/pnas.1117716109.

[88] K.A. Youdim, M.S. Dobbie, G. Kuhnle, et al. (2003), "Interaction between flavonoids and the blood-brain barrier: *In vitro* studies", *Journal of Neurochemistry*, **85(1)**, pp.180-192, DOI: 10.1046/j.1471-4159.2003.01652.x.

[89] D.W. Lachenmeier, et al. (2015), "Comparative risk assessment of alcohol, tobacco, cannabis and other illicit drugs using the margin of exposure approach", *Scientific Reports*, **5**, DOI: 10.1038/srep08126.

[90] P. Beaulieu (2005), "Toxic effects of cannabis and cannabinoids: Animal data", *Pain Research and Management*, **10(A)**, pp.23A-26A, DOI: 10.1155/2005/763623.

[91] R.S. Gable (2004), "Comparison of acute lethal toxicity of commonly abused psychoactive substances", *Addiction*, **99(6)**, pp.686-696, DOI: 10.1111/j.1360-0443.2004.00744.x.

[92] World Health Organization (2020), *WHO Expert Committee on Drug Dependence: Forty-first report*, 68pp.

[93] E. Zahra, S. Darke, L. Degenhardt, et al. (2020), "Rates, characteristics and manner of cannabis-related deaths in Australia 2000-2018", *Drug and Alcohol Dependence*, **212**, DOI: 10.1016/j.drugalcdep.2020.108028.

[94] M. Herkenham, A.B. Lynn, M.D. Little, et al. (1990), "Cannabinoid receptor localisation in brain", *Proceedings of The National Academy of Sciences*, **87(5)**, pp.1932-1936, DOI: 10.1073/pnas.87.5.1932.

[95] J.M. McPartland, E.B. Russo (2001), "Cannabis and cannabis extracts: Greater than the sum of their parts?", *Journal of Cannabis Therapeutics*, **1(3)**, pp.103-132, DOI: 10.1300/J175v01n03_08.

[96] A. Englund, P.D. Morrison, J. Nottage, et al. (2013), "Cannabidiol inhibits THC-elicited paranoid symptoms and hippocampal-dependent memory impairment", *Journal of Psychopharmacology*, **27(1)**, pp.19-27, DOI: 10.1177/0269881112460109.

[97] A.W. Zuardi, I. Shirakawa, E. Finkelfarb, et al. (1982), "Action of cannabidiol on the anxiety and other effects produced by Δ9-THC in normal subjects", *Psychopharmacology*, **76**, pp.245-250, DOI: 10.1007/BF00432554.

[98] A.W. Zuardi, J.A.S. Crippa, J.E.C. Hallak, et al. (2012), "A critical review of the antipsychotic effects of cannabidiol: 30 years of a translational investigation", *Curr. Pharm. Des.*, **18(32)**, pp.5131-5140, DOI: 10.2174/138161212802884681.

[99] A. Batalla, H. Janssen, S.S. Gangadin, et al. (2019), "The potential of cannabidiol as a treatment for psychosis and addiction: Who benefits most? A systematic review", *Journal of Clinical Medicine*, **8(7)**, DOI: 10.3390/jcm8071058.

[100] C. Davies, S. Bhattacharyya (2019), "Cannabidiol as a potential treatment for psychosis", *Therapeutic Advances in Psychopharmacology*, **9**, DOI: 10.1177/2045125319881916.

[101] P. McGuire, P. Robson, W.J. Cubala, et al. (2018), "Cannabidiol (CBD) as an adjunctive therapy in schizophrenia: A multicenter randomized controlled trial", *American Journal of Psychiatry*, **175(3)**, pp.225-231, DOI: 10.1176/appi.ajp.2017.17030325.

[102] J.D. Brown, A.G. Winterstein (2019), "Potential adverse drug events and drug-drug interactions with medical and consumer cannabidiol (CBD) use", *Journal of Clinical Medicine*, **8(7)**, DOI: 10.3390/jcm8070989.

DOI: 10.31276/VJSTE.64(4).79-83

Triterpenoid saponins from the root of *Weigela florida*

Duc Hung Nguyen*, Quang Tan Tu, Hoang Mau Chu

University of Education, Thai Nguyen University, 20 Luong Ngoc Quyen Street, Quang Trung Ward, Thai Nguyen City, Thai Nguyen Province, Vietnam

Received 3 December 2021; revised 10 January 2022; accepted 21 January 2022

Abstract:

Plants belonging to the *Weigela* genus are traditionally used to flourish gardens, homes, and other living areas. This genus contains ten species that are mainly found in Asia, but many cultivars have been cultivated around the world. Previous studies on some species of this genus revealed the occurrence of triterpenoid saponins, which possess interesting biological activities such as cytotoxicity, anti-complementary activity, and anti-inflammatory activity. Based on this suggestion, a phytochemical study on the roots of *W. florida* "Jean's Gold" was carried out using different chromatographic methods, spectroscopic analysis in 1D- and 2D-NMR, and mass spectroscopic method (HRESI-MS in negative mode). The isolation of three known oleanane-type triterpenoid saponins was revealed as olean-12-en-28-oic acid, 3-[(O-β-D-xylopyranosyl-(1→4)-O-β-D-glucopyranosyl-(1→4)-O-β-D-xylopyranosyl-(1→3)-O-6-deoxy-α-L-mannopyranosyl-(1→2)-β-D-xylopyranosyl)oxy]-, (3β)- (compound 1), olean-12-en-28-oic acid, 3-[(O-β-D-xylopyranosyl-(1→4)-O-β-D-glucopyranosyl-(1→4)-O-β-D-xylopyranosyl-(1→3)-O-6-deoxy-α-L-rhamnopyranosyl-(1→2)-α-L-arabinopyranosyl)oxy]-, (3β)- (compound 2), and olean-12-en-28-oic acid, 3-[(O-α-L-arabinopyranosyl-(1→3)-O-[α-L-xylopyranosyl-(1→4)]-O-β-D-glucopyranosyl-(1→4)-O-β-D-xylopyranosyl-(1→3)-O-6-deoxy-α-L-rhamnopyranosyl-(1→2)-α-L-arabinopyranosyl)oxy]-, (3β)- (compound 3). Full characterisation of the three saponins presented in this study gives a better understanding of the phytochemistry of ornamental plants, which can be used further for medicinal purposes.

Keywords: **Caprifoliaceae, NMR, ornamental shrub, triterpenoid saponin, *Weigela florida* "Jean's Gold".**

Classification number: **3.3**

1. Introduction

Ornamental plants can have many useful purposes. Although they possess beautiful flowering and a lovely attraction, they are being grown not only for environment beauty in workplaces, homes, universities, etc., but also for spreading freshness and eliminating toxic volatile substances inside homes, hospitals, and workplaces [1, 2]. Another interest in ornamental plants comes from their medicinal use as they possess many bioactive compounds like flavonoids, phenolic acids, and vitamins [3, 4]. Plants in the *Weigela* genus are traditionally used to flourish gardens, homes, and other living areas. This genus contains ten species that are mainly found in Asia, but many cultivars have been cultivated [5]. Among those, *W. florida* "Jean's Gold" is a medium-sized ornamental shrub with green leaves and red flowers. Previous studies on some species of this genus led to a revelation of the presence of triterpenoid saponins, which possess interesting biological activities such as cytotoxicity, anti-complementary activity, and anti-inflammatory activity [5-8]. Based on this suggestion, a phytochemical study on the roots of *W. florida* "Jean's Gold" was carried out using different chromatographic methods, analysis of 1D- and 2D-NMR, and mass spectroscopy using negative mode HRESI-MS. The isolation of three known oleanane-type triterpenoid saponins was further revealed. A suggestion for future phytochemical research on the Weigela genus were carried out together with the biological activities of components and specifically on oleanane-type triterpenoid saponins. This research will be useful to establish a relationship between the structure and activity of those saponins.

2. Materials and methods

2.1. Plant material

The ornamental shrub *W. florida* "Jean's Gold" (Fig. 1) was collected in 2016 from the flower boutique Botanic, Quetigny, France at 47°18'45.6"N, 5°05'41.4"E. A voucher specimen was stored in Laboratoire de Pharmacognosie EA 4267, Université de Bourgogne, Dijon, France.

Fig. 1. The ornamental shrub *W. florida* "Jean's Gold".

2.2. General experimental procedure

An automatic polarimeter AA-10R (Optical Activity®, England) was used to record optical rotation values. A Varian VNMR-S 600 MHz (Agilent Technologies®, USA) was used to record the 1D and 2D spectra including 1H, ^{13}C, HSQC, ROESY, HMBC, TOCSY, COSY NMR, and further recorded at the temperature of 35°C in pyridine-*d5* (C_5D_5N). HRESI-MS in the negative mode was done by a micrOTOF II mass spectrometer (Bruker®, Germany). A microwave apparatus

*Corresponding author: Email: hungnd@true.edu.vn

(MARS 6, CEM®, USA) was used to perform the extractions. The performance of VLC (Vacuum liquid chromatography) was carried out on normal phased silica gel 60 (60-200 µm, Merck®, Germany). MPLC (Medium-pressure liquid chromatography) was carried out on normal phased silica gel 60 (Merck®, Germany) with an M305 pump (Gilson®, USA). HPTLC (High-performance thin-layer chromatography, Merck®, Germany) and TLC (Thin layer chromatography, Silicycle®, Canada) were precoated silica gel plates $60F_{254}$. A vanillin reagent prepared by dissolving 1% vanillin in a solvent of EtOH:H_2SO_4 (50:1, v/v) was used to reveal the saponins in the samples. The phytochemical study was carried out at the Laboratory of Pharmacognosy, University of Burgundy, Dijon, France. The spectroscopic data were recorded at the Faculty of Science and Technology, University of Siegen, Siegen, Germany.

2.3. Extraction and isolation of compounds

After drying the whole plant at room temperature, each part of *W. florida* "Jean's Gold" was separated individually for different purposes. The dried root (58 g) was ground and further presented to an extraction using the microwave-assisted method with a solvent of EtOH: H_2O (75:35) in the following conditions: 60°C, 200 W, 45 min, 400 ml each x 3 times. After removing solvent using a rotary evaporator at 53°C, 5.2 g of crude extract was collected and further submitted to a VLC reversed-phased silica gel RP-18 with each 500 ml of solvents including 100% H_2O, 50% EtOH/50% H_2O, and 100% EtOH yielding three fractions A-C. Fraction B (295.4 mg) was separated by an MPLC on normal-phased silica gel 60 using solvent $CHCl_3$/MeOH/H_2O 75/25/3, 70/30/5, 60/32/7 (v/v/v) to give 4 subfractions B1-B4. Subfraction B2 was fractionated again by a successive MPLC on normal-phased silica gel 60 ($CHCl_3$/MeOH/H_2O 70/30/5, v/v/v) resulting in compound **1** (wbr2314) (3.5 mg). Subfractions B3 was presented to a normal phased MPLC on silica gel 60 using solvent $CHCl_3$/MeOH/H_2O 70/30/5, 60/32/7 (v/v/v) affording compound **2** (wbr2311) (4.2 mg) and compound **3** (wbr2321) (3.5 mg) (Figs. 2-6).

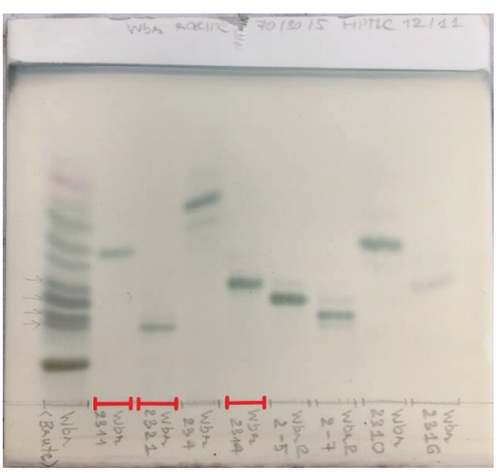


Fig. 2. HPTLC of isolated compounds (in red).

Compound **1**: Amorphous powder. ^{1}H-NMR (pyridine-d_5, 600 MHz) and ^{13}C-NMR (pyridine-d_5, 150 MHz), see Tables 1 and 2. HRESI-MS (negative mode) *m/z* 1160.4 $[M-H]^-$, $C_{57}H_{92}O_{24}$.

Compound **2**: Amorphous powder. ^{1}H-NMR (pyridine-d_5, 600 MHz) and ^{13}C-NMR (pyridine-d_5, 150 MHz), see Tables 1 and 2. HRESI-MS (negative mode) *m/z* 1160.1 $[M-H]^-$, $C_{57}H_{92}O_{24}$.

Compound **3**: Amorphous powder. ^{1}H-NMR (pyridine-d_5, 600 MHz) and ^{13}C-NMR (pyridine-d_5, 150 MHz), see Tables 1 and 2. HRESI-MS (negative mode) *m/z* 1322.1 $[M-H]^-$, $C_{63}H_{102}O_{29}$.

3. Results and discussion

The aqueous ethanolic crude extract of the roots of *W. florida* "Jean's Gold" was isolated and purified by various chromatographic methods running on normal phased silica gel affording compounds

Table 1. The spectroscopic data of the aglycone moiety of 1-3 (C_5D_5N, δ in ppm, *J* in Hz).

	1		2		3	
	δ_C	δ_H	δ_C	δ_H	δ_C	δ_H
1	39.0	0.94, 1.55	38.7	0.94, 1.52 m	39.1	0.96, 1.53
2	27.0	1.84, 2.12 m	26.5	1.83, 2.13	26.8	1.87, 2.12 *m*
3	88.6	3.33 *dd* (11.7, 4.0)	88.8	3.31 *dd* (12.1, 3.6)	89.0	3.31 *dd* (12.0, 3.9)
4	39.5	–	39.6	–	39.6	–
5	56.2	0.85 *d* (12.0)	56.0	0.84 *d* (12.0)	56.3	0.83
6	18.5	1.30, 1.55 *m*	18.4	1.27, 1.49 *m*	19.0	1.33, 1.50 *m*
7	33.3	1.24, 1.50 *m*	32.9	1.30, 1.46	33.5	1.26, 1.48 *m*
8	39.7	–	39.9	–	39.8	–
9	48.0	1.64 *dd* (15.2, 8.7)	48.1	1.65 *dd* (15.2, 8.8)	48.1	1.66
10	37.0	–	37.2	–	37.2	–
11	23.8	1.88, 1.94	23.6	1.89, 1.96	23.8	1.90, 1.95
12	122.5	5.45 *t-like* (3.7)	123.0	5.46 *t-like* (3.6)	123.1	5.45 *t-like* (3.8)
13	145.0	–	145.1	–	145.0	–
14	42.1	–	42.2	–	42.3	–
15	28.3	1.22 *m*, 2.12	27.9	1.22, 2.16	27.9	1.23, 2.16
16	24.0	1.96, 2.20	24.1	1.96, 2.15	24.0	1.97, 2.17
17	46.6	–	46.5	–	46.7	–
18	42.0	3.31 *dd* (11.8, 3.9)	42.0	3.33 *dd* (12.0, 4.0)	42.2	3.36 *dd* (11.8, 3.9)
19	46.5	1.33, 1.82	46.4	1.32, 1.82	46.6	1.30, 1.83
20	30.9	–	31.0	–	31.2	–
21	34.2	1.22 *m*, 1.50 *m*	34.4	1.24, 1.46	34.3	1.23, 1.46
22	33.4	1.82, 2.05 *m*	33.2	1.85, 2.10	34.0	1.85, 2.05
23	28.5	1.34 *s*	27.9	1.31 *s*	27.8	1.35 *s*
24	17.1	1.21 *s*	17.2	1.12 *s*	17.4	1.21 *s*
25	15.4	0.90 *s*	16.0	0.91 *s*	15.6	0.90 *s*
26	17.3	0.98 *s*	17.5	0.99 *s*	17.4	1.01 *s*
27	26.2	1.31 *s*	26.2	1.33 *s*	26.3	1.33 *s*
28	180.0	–	180.2	–	180.3	–
29	33.5	0.96 *s*	33.5	0.95 *s*	33.6	0.97 *s*
30	23.5	1.02 *s*	23.7	1.05 *s*	23.5	1.05 *s*

Table 2. The spectroscopic data of the sugar moieties of 1-3 (C5D5N, δ in ppm, J in Hz).

	1		2		3	
	δ_C	δ_H	δ_C	δ_H	δ_C	δ_H
Ara-1			105.1	4.89 *d* (6.0)		
2			75.5	4.56 *t* (6.0)		
3			74.4	4.24		
4			69.2	4.23 *m*		
5			65.3	3.81,4.33 *dd* (11.6, 3.6)		
Rha-1	101.5	6.49 *br s*	101.4	6.19 *br s*	101.5	6.51 *br s*
2	71.6	4.92 *br s*	71.8	4.86 *br s*	71.8	4.97 *br s*
3	83.3	4.71 *dd* (9.2, 2.8)	83.1	4.65 *dd* (9.6, 2.8)	83.4	4.75 *dd* (9.5, 3.5)
4	72.7	4.46 *dd* (9.6, 9.2)	72.8	4.47 *dd* (9.6, 9.2)	72.7	4.47 *dd* (9.5, 9.3)
5	69.5	4.77 *dq* (9.6, 6.4)	69.6	4.62 *dq* (9.2, 6.0)	69.5	4.77 *dq* (9.3, 5.9)
6	18.3	1.64 *d* (6.4)	18.4	1.55 *d* (6.0)	18.5	1.66 *d* (5.9)
Xyl-1	106.1	4.84 *d* (7.3)	106.9	5.27 *d* (7.5)	105.9	4.84 *d* (7.5)
2	77.7	4.24	75.4	4.05	77.7	4.19
3	76.3	4.15	76.1	4.17	79.8	4.15
4	71.6	4.17	77.7	4.28	71.7	4.17
5	67.2	3.71, 4.35	64.8	3.64, 4.39 *dd* (11.3, 5.1)	67.3	3.75 *t* (11.1), 4.32 *dd* (11.1, 4.9)
Xyl II-1	107.4	5.27 *d* (7.5)	105.5	5.07 *d* (7.5)	107.0	5.27 *d* (7.5)
2	75.4	4.05	74.5	4.01	75.5	4.06
3	76.3	4.14	78.4	4.13	76.3	4.12
4	77.8	4.26	70.9	4.16	77.7	4.21
5	64.6	3.67, 4.36	67.5	3.67 *t* (10.6), 4.26	64.8	3.64 *t* (11.1), 4.38 *dd* (11.1, 4.8)
Xyl III-1	105.6	5.05 *d* (7.6)			102.8	5.47 *d* (6.0)
2	74.7	3.99			73.8	4.16
3	78.4	4.09			76.3	4.18
4	70.6	4.14			70.5	4.24
5	67.5	3.66, 4.26			66.3	3.73 *t* (11.3), 4.63 *dd* (11.3, 4.1)
Glc I-1	103.3	4.99 *d* (8.0)	103.2	4.98 *d* (7.6)	102.7	4.92 *d* (7.6)
2	74.0	4.03	74.3	4.03	74.4	4.02
3	76.1	4.24	76.2	4.21	82.4	4.44
4	80.8	4.26	80.7	4.25	74.3	4.56 *dd* (9.4, 9.1)
5	76.5	3.90	76.7	3.93 *m*	77.6	3.83 *m*
6	61.4	4.47, 4.49	61.6	4.45, 4.49	61.4	4.34, 4.37
Glc II-1					104.6	5.51 *d* (8.0)
2					75.4	4.05
3					71.4	4.16
4					78.2	4.19
5					78.6	3.93 *m*
6					62.6	4.26 *dd* (11.9, 5.3), 4.42

1, **2,** and **3**. The structural determination of these compounds was carried out by spectroscopic methods (1D- and 2D-NMR) and mass spectroscopy (HRESI-MS in negative mode). The sugar moieties were determined by 2D-NMR analyses including HMBC, COSY, HSQC, ROESY, and TOCSY as three xylopyranosyl, one rhamnopyranosyl, one glucopyranosyl for compound **1**, two xylopyranosyl, one arabinopyranosyl, one rhamnopyranosyl, one glucopyranosyl for compound **2**, and three xylopyranosyl, one rhamnopyranosyl, two glucopyranosyl for compound **3**. Their absolute configurations were further determined to be D for xylose (Xyl) and glucose (Glc) and L for rhamnose (Rha) and arabinose (Ara) according to previous methods of acid hydrolysis and absolute configuration determination [9]. The relatively large values of Ara, Xyl, and Glc from 7.2 to 8.0 Hz revealed a β anomeric orientation for Xyl and Glc, and the α for Ara. The relatively large $J_{C\text{-}1,H\text{-}1}$ value of Rha from 165 to 167 Hz indicated an α-equatorial pyranoid form for this anomeric proton.

Compound **1** displayed a pseudo-molecular ion peak $[M\text{-}H]^-$ at 1160 compatible with a molecular formula of $C_{57}H_{92}O_{24}$ (see Fig. 6 in Supplementary data). Seven angular methyl groups were showed on the aglycone of the HSQC spectrum of compound **1** at δ_H/δ_C 1.34 (*s*)/28.5 (23-Me), 1.21 (*s*)/17.1 (24-Me), 0.90 (*s*)/15.4 (25-Me), 0.98 (*s*)/17.3 (26-Me), 1.31 (*s*)/26.2 (27-Me), 0.96 (*s*)/33.5 (29-Me) and 1.02 (*s*)/23.5 (30-Me) (Table 1). In addition, the observation of the characteristic signal was reached as one olefinic proton at δ_H/δ_C 5.45 (*br t*, *J*=3.6 Hz)/122.5 (C-12). The positions of those signals were assigned by HMBC cross-peaks (Fig. 3). The configurations of H-3, H-5, H-9, H-12 and H-18 were determined by analysing ROESY spectra at δ_H/δ_H 1.34 (23-Me α-equatorial)/3.33 (H-3 α-axial), 1.34 (23-Me α-equatorial)/ 0.85 (H-5 α-axial), 1.31 (27-Me α-axial)/1.64 (H-9 α-axial), 1.02 (30-Me β-axial)/3.31 (H-18 β-axial), and 3.31 (H-18 β-axial)/5.45 (H-12 β-equatorial). Those of cross-peaks were validated the configuration of 3-OH as β-equatorial (Fig. 4). All the evidence above led to an assignment to the aglycone of compound **1** as oleanolic acid which was in agreement with literature data [8].

The sugar region of compound **1** displayed five signal protons at δ_H 4.84 (*d*, *J*=7.3 Hz), 5.27 (*d*, *J*=7.5 Hz), 5.05 (*d*, *J*=7.6 Hz), 4.99 (*d*, *J*=8.0 Hz) and 6.49 (*br s*), further identified the correlations with five signal carbons through the HSQC spectrum at δ_C 106.1, 107.4, 105.6, 103.3 and 101.5 respectively (Table 2). Analysing the extensive NMR spectra including ^1H, ^{13}C, HSQC, HMBC, TOCSY, ROESY and TOCSY led to an identification of three Xyl, one Glc and one Rha moiety. The establishment of the oligosaccharidic sequence of compound **1** was carried out using ROESY and HMBC spectra: the HMBC correlation at δ_H/δ_C 4.84 (*d*, *J*=7.3 Hz)/88.6 (C-3) and the ROESY cross peak at δ_H/δ_H 4.84 (*d*, *J*=7.3 Hz, Xyl I-1)/3.33 (*dd*, *J*=11.7, 4.0 Hz, 3-H) displayed the linkage between the Xyl I and the C-3 position of the aglycone. The Rha was identified to be linked to the Xyl I by the HMBC correlation at δ_H/δ_C 6.49 (*br s*, Rha-1)/77.7 (Xyl I-2) together with the ROESY correlation at δ_H/δ_H 6.49 (*br s*, Rha-1)/4.24 (Xyl I-2). The (1→3) cross peak of Xyl II and Rha was determined by the

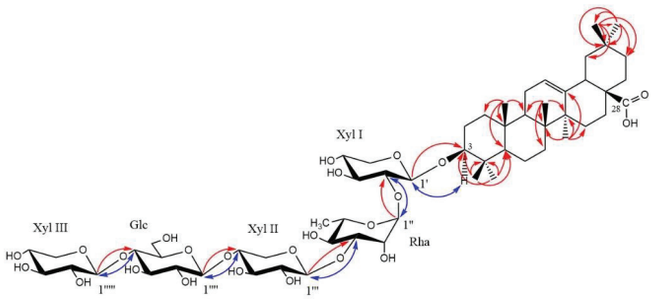


Fig. 3. Structure of compound 1 (HMBC correlations in red, ROESY correlations in blue).

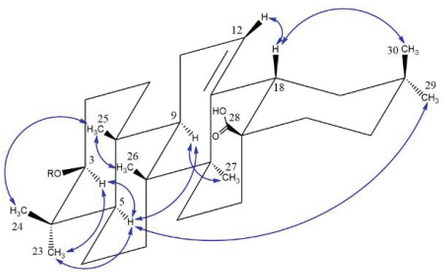


Fig. 4. ROESY correlations in the aglycone region of compound 1 (in blue).

HMBC correlation at δ_H/δ_C 5.27 (*d*, *J*=7.5 Hz, Xyl II-1)/83.3 (Rha-3) together with the ROESY cross peak at δ_H/δ_H 5.27 (*d*, *J*=7.5 Hz, Xyl II-1)/4.71 (*br s*, Rha-3). The linkage between the Glc and the Xyl II was identified by the HMBC cross peak at δ_H/δ_C 4.99 (*d*, *J*=8.0 Hz, Glc-1)/77.8 (Xyl II-4) and the ROESY cross peak at δ_H/δ_H 4.99 (*d*, *J*=8.0 Hz, Glc-1)/4.26 (Xyl II-4). Finally, the terminal sequence β-D-xylopyranosyl-(1→4)-β-D-glucopyranosyl was established by the HMBC cross peak at δ_H/δ_C 5.05 (*d*, *J*=7.6 Hz, Xyl III-1)/80.8 (Glc-4) and the ROESY correlation at δ_H/δ_H 5.05 (*d*, *J*=7.6 Hz, Xyl III-1)/4.26 (Glc-4).

Due to the results above, the structure as Olean-12-en-28-oic acid, 3-[(*O*-β-D-xylopyranosyl-(1→4)-*O*-β-D-glucopyranosyl-(1→4)-*O*-β-D-xylopyranosyl-(1→3)-*O*-6-deoxy-α-L-mannopyranosyl-(1→2)-β-D-xylopyranosyl)oxy]-, (3β)- was established for compound **1**. Comparison with literature data revealed that this compound was firstly isolated from *Pterocephalus hookeri* [10].

The similar HRESI-MS of compound **2** as compound **1** was shown with a pseudo-molecular ion peak $[M-H]^-$ at 1160 related to a molecular formula of $C_{57}H_{92}O_{24}$ (see Fig. 12 in Supplementary data). Analysing 1D and 2D NMR revealed that the aglycone of compound **2** was the same as compound **1** (Table 1). The difference between these compounds was only displayed in the sugar region which further identified that the Xyl I moiety in **1** was replaced by the Ara moiety in compound **2** (Table 2). This identification was confirmed by analysing mainly ROESY and HMBC spectra: the HMBC cross peak at δ_H/δ_C 4.89 (*d*, *J*=6.0 Hz, Ara-1)/88.8 (C-3), together with the ROESY cross peak at δ_H/δ_H 4.89 (*d*, *J*=6.0 Hz, Ara-1)/3.31 (*dd*, *J*=12.1, 3.6 Hz, 3-H) displayed a linkage of the Ara at the C-3 position of the aglycone. The Xyl I was determined to be linked to the Rha by the HMBC correlation at δ_H/δ_C 6.19 (*br s*, Rha-1)/75.5 (Ara-2) and the ROESY correlation at δ_H/δ_H 6.19 (*br s*, Rha-1)/4.56 (Ara-2). The oligosaccharidic sequence was further established as Xyl II-(1→4)-Glc-(1→4)-Xyl-(1→3)-Rha-(1→2)-Ara by comparison with those data of compound **1** (Fig. 5).

According to all the evidence above, the structure as Olean-12-en-28-oic acid, 3-[(*O*-β-D-xylopyranosyl-(1→4)-*O*-β-D-glucopyranosyl-(1→4)-*O*-β-D-xylopyranosyl-(1→3)-*O*-6-deoxy-α-L-rhamnopyranosyl-(1→2)-α-L-arabinopyranosyl)oxy]-, (3β)- was elucidated for compound **2**. This saponin has been already isolated from *W*. x "Bristol ruby" [7].

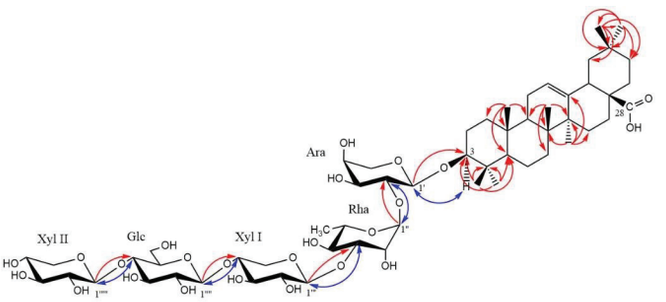


Fig. 5. Structure of compound 2 (HMBC correlations in red, ROESY correlations in blue).

The pseudo-molecular ion peak $[M-H]^-$ at *m/z* 1322 in the HRESI-MS led to an establishment of the molecular formula of compound **3** as $C_{63}H_{102}O_{29}$ (see Fig. 18 in Supplementary data). Comparison with signal data in the aglycone region of compound **1** gave an identification that the aglycone of compound **3** was the same as compounds **1** and **2** (Table 1). The molecular weight of compound **3** differs than compound **1** by 162 amu indicating the supplement of one additional sugar unit. Six anomeric cross-peaks in the HSQC spectrum at δ_H/δ_C 4.84 (*d*, *J*=7.5 Hz)/105.9, 5.27 (*d*, *J*=7.5 Hz)/107.0, 5.47 (*d*, *J*=6.0 Hz)/102.8, 4.92 (*d*, *J*=7.6 Hz)/102.7, 5.51 (*d*, *J*=8.0 Hz)/104.6 and 6.51 (*br s*)/101.5 were identified as three β-D-xylopyranosyl, one α-L-rhamnopyranosyl, and two β-D-glucopyranosyl.

The establishment of the oligosaccharidic chain and heterosidic linkage was further achieved by analysing the 2D NMR spectra (HSQC, COSY, ROESY, HMBC, TOCSY). The HMBC cross peak at δ_H/δ_C 4.84 (*d*, *J*=7.5 Hz)/89.0 together with the ROESY cross peak at δ_H/δ_H 4.84 (*d*, *J*=7.5 Hz)/3.31 (*dd*, *J*=12.0, 3.9) confirmed that the Xyl I was attached to the C-3 position of the aglycone (Table 2). The oligosaccharidic sequence attached to the C-2 position of the Xyl I was confirmed by the HMBC cross peaks at δ_H/δ_C 6.51 (Rha-1)/77.7 (Xyl I-2), 5.27 (Xyl II-1)/83.4 (Rha-3), 4.92 (Glc I-1)/77.7 (Xyl II-4), 5.51 (Glc II-1)/82.4 (Glc I-3) and 5.47 (Xyl III-1)/74.3 (Glc I-4), in combination with the ROESY cross peaks at δ_H/δ_H 6.51 (Rha-1)/4.19 (Xyl I-2), 5.27 (Xyl II-1)/4.75 (Rha-3), 4.92 (Glc I-1)/4.21 (Xyl II-4), 5.51 (Glc II-1)/4.44 (Glc I-3) and 5.47 (Xyl III-1)/4.56 (Glc I-4). The conclusion was

further reported as the (1→3) and (1→4) linkages between Glc II, Xyl III and Glc I, the (1→4) linkage between Glc I and Xyl II, the (1→3) linkage between Xyl II and Rha (Fig. 6). Based on above results, structural characterization of compound **3** was established as Olean-12-en-28-oic acid, 3-[(*O*- β-D-glucopyranosyl-(1→3) -*O*-[α-L-xylopyranosyl-(1→4)]-*O*-β-D-glucopyranosyl-(1→4) -*O*-β-D-xylopyranosyl-(1→3)-*O*-6-deoxy-α-L-rhamnopyranosyl-(1→2)-β-D-xylopyranosyl)oxy]-, (3β)-. Literature revealed that this compound was firstly isolated from *W*. x "Bristol ruby" [7].

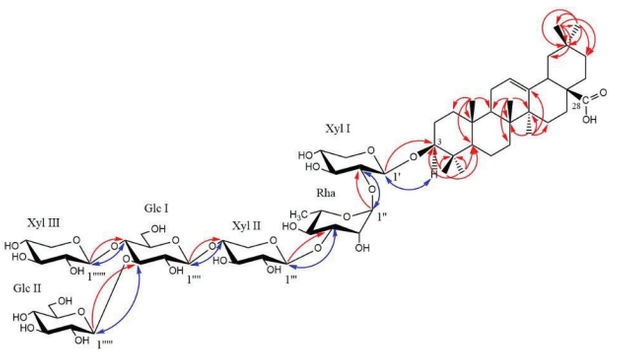


Fig. 6. Structure of compound 3 (HMBC correlations in red, ROESY correlations in blue).

Previous studies led to the conclusion that the oleanane-type triterpenoid saponins exhibit various biological activities including anti-tumour, anti-allergic, anti-inflammatory, immunological, anti-bacterial, and molluscicidal activities [11]. Specifically, various phytochemical and biological studies have been carried out on some species of the *Weigela* genus, which revealed that the oleanane-type triterpenoid saponins potentially possess biological activities. The investigations on cytotoxic activity against some cancer cell lines for the triterpenoid saponins isolated from *W. stelzneri* and *W*. x "Bristol ruby" were carried out against SW480, EMT6, and CT26 cancer cell lines that further exhibited a potent result for the oleanane-type triterpenoid saponins with high IC_{50} values [7, 8]. The immunological activity of oleanane-type triterpenoid saponins isolated from *W. florida* "rumba" (Bunge) A. DC was further evaluated by A.S.C. Tixier, et al. (2018) [12]. The first evaluation on toxicological properties in an *in vivo* zebrafish assay of those saponins revealed a significant result from the larvae in the range of 1 µM [5]. The characterisation of those saponins gives a better understanding of the phytochemistry and biological activities of the *Weigela* genus and suggests more evaluations of biological activity to be performed in order to establish structure activity related to these compounds.

4. Conclusions

Three oleanane-type triterpenoid saponins were characterised from the roots of *W. florida* "Jean's Gold" using various chromatographic methods. Their structures were further established by spectroscopic analysis of 1D- and 2D-NMR and mass spectroscopy (HRESI-MS in negative mode) in combination with data from the literature. Full structural characterisation of these three compounds together with structure activity discussion based on facts from the literature afford us a better understanding of the *Weigela* genus phytochemistry and putative biological activities.

CRediT author statements

Duc Hung Nguyen, Quang Tan Tu, Hoang Mau Chu: Data curation, Software, Visualisation, Investigation, Writing original draft preparation, Conceptualization, Methodology, Supervision.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] M.W. Iersel, M.R. Chappell, P.A. Thomas (2016), "Optimizing growth, qual[1] M.W. Iersel, M.R. Chappell, P.A. Thomas (2016), "Optimizing growth, quality, and profits through precision irrigation in ornamental plant production", *Acta Horticulturae*, **1131**, pp.57-64, DOI: 10.17660/ActaHortic.2016.1131.8.

[2] X. Wei, S. Lyu, Y. Yu, et al. (2017), "Phylloremediation of air pollutants: Exploiting the potential of plant leaves and leaf-associated microbes", *Frontiers in Plant Science*, **8**, DOI: 10.3389/fpls.2017.01318.

[3] M. Goleniowski, M. Bonfill, R. Cusido, et al. (2013), "Phenolic acids", *Natural Products*, **3**, pp.1951-1973, DOI: 10.1007/978-3-642-22144-6_64.

[4] K. Jakubczyk, K.J. Milczarek, K. Watychowicz, et al. (2018), "Garden nasturtium (*Tropaeolum majus* L.) - A source of mineral elements and bioactive compounds", *Rocz. Państwowego Zakładu Hig.*, **69(2)**, pp.119-126.

[5] N. Andriamisaina, et al. (2018), "Phytochemistry of *Weigela* x 'kosteriana variegata' (Caprifoliaceae)", *Nat. Prod. Commun.*, **13(4)**, pp.403-406, DOI: 10.1177/1934578X1801300406.

[6] P.T. Thuong, B.S. Min, W.Y. Jin, et al. (2006), "Anti-complementary activity of ursane-type triterpenoids from *Weigela subsessilis*", *Biol. Pharm. Bull.*, **29(4)**, pp.830-833, DOI: 10.1248/bpb.29.830.

[7] D.H. Nguyen, A.C.M. Offer, S. Maroso, et al. (2019), "Cytotoxic glycosides from the roots of *Weigela* x 'Bristol Ruby'", *Fitoterapia*, **137**, DOI: 10.1016/j.fitote.2019.104242.

[8] A. Rezgui, A.C.M. Offer, T. Miyamoto, et al. (2016), "Oleanolic acid and hederagenin glycosides from *Weigela stelzneri*", *Phytochemistry*, **123**, pp.40-47, DOI: 10.1016/j.phytochem.2015.12.016.

[9] D.H. Nguyen, A.C.M. Offer, T. Miyamoto, et al. (2021), "Steroidal glycosides from the Vietnamese cultivar *Cordyline fruticosa* 'Fairchild red'", *Phytochemistry*, **192**, DOI: 10.1016/j. phytochem.2021.112966.

[10] T. Jun, F.E. Wu, M.H. Qiu, et al. (1993), "Triterpenoid saponins from *Pterocephalus hookeri*", *Phytochemistry*, **32(6)**, pp.1535-1538, DOI: 10.1016/0031-9422(93)85174-p.

[11] Z. Fang, J. Li, R. Yang, et al. (2020), "A review: The triterpenoid saponins and biological activities of *Lonicera Linn*", *Molecules*, **25(17)**, DOI: 10.3390/molecules25173773.

[12] A.S.C. Tixier, A.C.M. Offer, F.R. Fernandez, et al. (2018), "Oleanane-type glycosides from the roots of *Weigela florida* 'rumba' and evaluation of their antibody recognition", *Fitoterapia*, **128**, pp.198-203, DOI: 10.1016/j.fitote.2018.04.017.

DOI: 10.31276/VJSTE.64(4).84-90

Bipolar disorder traits: An electroencephalogram systematic review

Tien Dat Nguyen[1, 2], Van Toi Vo[1, 2], Thi Thanh Huong Ha[1, 2*]

[1]*School of Biomedical Engineering, International University - Vietnam National University, Ho Chi Minh City, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam*
[2]*Vietnam National University - Ho Chi Minh City, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam*

Received 2 August 2022; revised 20 October 2022; accepted 31 October 2022

Abstract:

Bipolar disorder (BD) is a serious mental disorder that globally affected 40 million people in 2019. According to the National Alliance on Mental Illness (NAMI), the present state of scientific knowledge only permits psychiatrists to diagnose BD using subjective and imprecise questionnaires. Therefore, developing a diagnostic tool with objective and precise biomarkers should be a major focus of research in this field. Among the potential biomarkers for BD, electroencephalogram (EEG)-based signatures of BD are considered to be the most optimal marker due to their strong links with behavioural symptoms and also their non-invasiveness. The goal of this review is to give a detailed summary of current techniques for investigating the traces of BD through EEG abnormalities. In this review, 13 studies from databases such as ScienceDirect and PubMed seeking to utilize EEG characteristics to diagnose BD were selected. The search keywords were "EEG in BD diagnosis", "EEG microstates in BD", and "EEG features for BD patients". The publication date was set from 2007 to 2021. From these studies, we synthesize the effects of BD on each EEG feature, as well as detail the pros and cons when using each feature as a biomarker for BD. Results showed that EEG microstates demonstrate their potential among the seven EEG properties discussed in this article, as shown by several studies. By definition, EEG microstates are a dynamic representation of the spatial distribution of the scalp's electric potential as it varies over time. Specifically, four microstate classes recorded in different brain regions are classified into A (right-frontal left-posterior), B (left-frontal right-posterior), C (midline frontal-occipital), and D (midline frontal topographies). Greater presence of microstate class B in BD patients during task-free resting states are a distinctive characteristic of BD patients from which BD can be differentiated from other psychiatric illnesses. Besides microstates, EEG resting states are also considered to have a bright future in BD diagnosis. Specifically, by investigating brain frequency bands, researchers have discovered that BD patients exhibit abnormal delta and alpha signals as compared to healthy controls (HCs). The abnormalities of microstate B in EEG microstate characteristics would be the most promising biomarker for detecting BD. In addition, anomalies in delta and alpha signals during resting EEG states are possible BD diagnostic indicators.

Keywords: **bipolar disorder, EEG diagnosis, EEG features.**

Classification number: **3.6**

1. Introduction

BD is a severe mental disorder that afflicted 40 million individuals globally in 2019. BD is defined as a brain condition that causes a person to experience significant changes in emotions, mood, and energy levels ranging from extremely low to extremely high [1]. Life is split into two realities for the many millions of people who suffer from bipolar illness across the world: mania and depression. BD is divided into three typical types, although there are numerous variations.

Bipolar type I is defined as a person who has manic episodes or signs of manic episodes for at least seven days [1]. Bipolar type I individuas may necessitate emergency hospitalization in some situations. Depressive episodes have also been known to last for at least two weeks. In addition, bipolar type II is defined as a sequence of depressed and hypomanic episodes rather than the full-blown manic episodes as seen in BD type I. Finally, cyclothymia, another form of BD, is defined as a mood disorder

BD typical symptoms

As mentioned before, bipolar patients mainly suffer from two distinctive phases:

Phase 1: Depression period

During the depression period, patients likely experience:

+ Lack of motivation, energy
+ Irritability
+ Feeling negative about everything
+ Suicidal thoughts
+ Insomnia, lack of appetite
+ Delusion, hallucination, and illogical thinking are discovered in emergency cases

Phase 2: Maniac period

During the maniac period, patients likely suffer from:

+ Overly happy, elated
+ Feeling irritated and agitated
+ Spending significant sums of money on expensive and, at times, unaffordable objects is an example of conduct that frequently has negative outcomes.
+ Making decisions or speaking things that others perceive as hazardous or damaging and that are out of character
+ Insomnia, lack of appetite

*Corresponding author: Email: htthuong@hcmiu.edu.vn

in which sufferers experience hypomania and mild depression (at least for two years). People with cyclothymia find it difficult to find a good balance in their lifestyles [1].

According to NAMI, a physician may perform a physical examination, conduct an interview, and request laboratory testing to diagnose BD [2]. While a blood test or body scan cannot detect BD, they may help rule out other disorders that might mimic it, such as hyperthyroidism [3]. Therefore, developing a diagnostic tool with a greater scientific basis is the main target of this research with the understanding that EEG can be applied to diagnose BD and especially to distinguish BD from other serious psychiatric disorders such as schizophrenia (SCH). Realistically speaking, EEG is a diagnostic tool that is used to record electrical activity in the brain. By using small, metal electrodes attached to the scalp, EEG can record the electrical impulses that are communicated by our brain cells and display them as wavy lines on an EEG recording [4].

In neuro-engineering, EEG has been widely used to detect many serious brain disorders such as epilepsy, SCH, and so on [4]. To the best of the authors' knowledge, despite being applied in diagnosing many psychiatric disorders, there is not much research about using EEG as a diagnostic tool to detect BD. Traditionally, to diagnose BD, the results are mainly dependent on patient reports, which are subjective [5]. Therefore, this study aims to fulfil that missing piece by focusing on using EEG signals to detect and classify BD with other disorders. Therefore, in this review, the cognitive and affective changes in BD patients are meticulously discussed. Based on those abnormalities, studies about utilizing EEG features in diagnosing BD will be presented in a systematic way. Moreover, this review also suggests the most optimal solution for the diagnosis.

2. Paper selection criteria

Fifty papers were selected from academically trusted scientific library sources such as PubMed and ScienceDirect. After analysing all articles, only 13 papers fit all criteria (listed below) and were utilized for this review. In the scope of this review, the literature contains several articles with focus on specific biomarkers, literature reviews, or computational methods for BD diagnosis. This review is structured by the cognitive and affective changes in BD patients, followed by a brief explanation of EEG features to diagnose BD based on these changes. Finally, the research topics will be presented as well as some research constraints.

Included criteria: All research papers applied the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) as the standard subject recruitment criterion. Selected papers must be authored by credentialed professionals or reliable sources. In addition, researchers must provide precise results with a justifiable explanation.

Excluded criteria: After vetting the papers with multiple online sources, papers that only give hypotheses and do not have convincing arguments will be rejected. Besides that, papers that were published before 2,000 were also excluded to maintain the applicability of this review.

3. BD cognitive and affective changes

BD exemplifies the notion of mood and affect. Numerous studies suggest that bipolar illness may be associated with the disturbance of the brain or other bodily functions [3]. However, no distinct theoretical standpoint has been established. However, the viewpoint of applying EEG in psychiatry as an alternative to the DSM for understanding and treating people with mental disorders has been mentioned. This method is commonly used by healthcare professionals for the treatment and education of individuals with mental problems. BD patients often encounter physiologic alterations. Hence, in our review, all of the main findings about the cognitive and affective alterations in BD patients are summarized in Fig. 1 below:

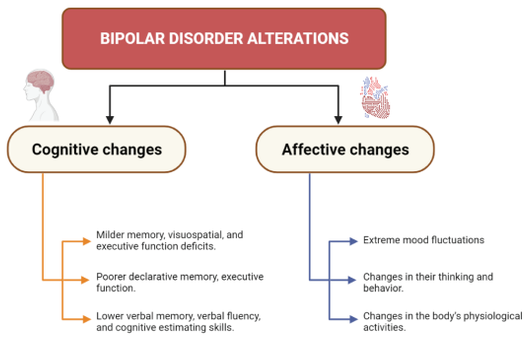


Fig. 1. A summary of behavioural changes in BD patients in terms of cognitive and affective functions.

3.1. Cognitive changes

The mental processes that allow us to receive, assess, store, process, develop, and recover knowledge gained from external stimuli are referred to as brain cognitive functions. This process improves our capacity to grasp and relate to the world. Cognitive alterations have been widely observed in bipolar patients [3]. Many studies have been carried out that describe as well as explain these abnormal changes. These changes can be typically listed as:

Milder memory, visuospatial, and executive function deficits: Although BD patients perform numerically better than patients with SCH in terms of neurocognitive performance, bipolar adults demonstrate the same qualitative impairment pattern in both diseases. Specifically, studies indicate that some individuals with BD have also complained of memory impairment during periods of elevated and depressed emotions, as well as at other times. As a person's mood changes, they may also experience memory alterations. Therefore, memory issues might worsen when an individual's mood grows more intense [3]. Moreover, many studies have shown that unaffected relatives and children of bipolar patients exhibit mild memory, visuospatial, and executive function deficits [3]. Specifically, endophenotype biomarkers have indicated that BD is commonly inherited, with genetic factors accounting for around 80% of the disease's causes.

Bipolar depression is the most common mental condition that is handed down from generation to generation. If one parent has BD, their child has a 10% risk of developing the disease.

Poorer decision-making and declarative memory: When compared to HCs, bipolar patients display psychomotor retardation as well as poor declarative memory, decision-making, and, to a lesser extent, visual memory, and attention [3]. Scientists discovered that alterations in the brain cause memory issues in certain individuals with BD. Alteration in the prefrontal cortex can be damaging as it is involved in planning, attention, problem-solving, and memory, among other tasks [2].

Lower verbal memory, verbal fluency, and cognitive estimating skills: A study has shown that when compared to depressed and remitted patients, bipolar patients exhibit lower verbal memory, verbal fluency, and cognitive estimating skills [3]. The study also indicated that the primary symptom of bipolar depression is a loss of verbal memory. Moreover, another research showed that manic episodes were associated with more severe deficiencies in verbal and working memory, executive function/reasoning, and problem-solving [6].

Therefore, several research studies have revealed that bipolar individuals, regardless of their stage of disease, exhibit cognitive deficiencies. At the time of their release from an inpatient facility, BD patients were shown to have a worse cognitive profile than depressive patients in terms of psychomotor slowness and selective attention in one study [7]. Conspicuously, the examination of the same group of participants six to eight weeks after release found that manic patients had retained the same pattern of impaired psychomotor speed and a high tendency for the perseverative behaviours that they exhibited at the time of discharge [7].

3.2. Affective changes

Fear, pleasure, the satisfaction of all sorts, sexuality, and jealousy are among the brain activities related to emotions [8]. Affective functions are situated in the brain's most fundamental areas, particularly the limbic region, which we share with many other animals [8]. In bipolar individuals, affective functioning alterations are different among the disorder's phases. For most BD patients, these changes can be typically listed as:

Extreme mood fluctuations: A person suffering from depression or mania may experience extreme mood fluctuations (also known as mood swings). Mood fluctuations may catch them off guard at first. However, BD patients may notice trends or symptoms that they are approaching a phase of mania or depression over time since the patterns of mood fluctuations mainly depend on the type of BD [8].

Changes in their thinking and behaviour: In most cases, BD patients exhibit changes in their thinking and behaviour. Frequently, BD individuals may exhibit pressurized speech and racing thoughts during manic episodes. In some cases, patients with BD are delusive in certain critical circumstances. For instance, they may believe that they have supernatural powers. In addition, bipolar individuals also experience depressed episodes along with euphoric moods, which are comparable to major depressive disorder (so-called unipolar depression) [8].

3.3. Changes in the body's physiological activities

Aside from changes in mood, BD patients are also observed to have changes in their physiological activities. Many studies have pointed out some typical lifestyle changes observed in manic patients that could be the result of physiological changes. For instance, in a study on women with mental disorders [9], in the instance BD, women are more likely to engage in reckless sexual conduct. In another study, researchers studied 63 female outpatients with a diagnosis of bipolar illness from the bipolar program at Favaloro University [10]. The remaining 63 women had no psychiatric history and maintained a healthy lifestyle. The 63 bipolar women were not in committed monogamous partnerships and engaged in casual sexual activity. They also had personal relationships with people whose HIV or other sexually transmitted illness status were unknown. In addition, they were infected with two or more sexually transmitted diseases. The authors determined that women with bipolar illness were more prone than healthy women (women without BD) to engage in hazardous and promiscuous conduct [10]. According to the American Psychiatric Association, BD is characterized by fundamental sleep disturbances. The diagnostic criteria also imply that during manic periods there may be a decreased demand for sleep, but during depressive episodes, insomnia or hypersomnia may occur virtually every day.

The physiological alterations of BD patients are explained in Table 1.

Table 1. Physiological changes in BD patients.

Physiological changes in BD patients	
Sleep pattern	People with bipolar illness report lower-quality sleep in general between cycles, including increased night waking, an overall impression of less sleep, and occasional insomnia.
Energy level	Sleep difficulties are connected to lower energy levels and, as a result, a lower probability of engaging in healthy habits in people with bipolar (e.g., socializing, shopping, cooking, and exercising).
Alcohol and drug use	Self-medicating with drugs and alcohol is prevalent among individuals with BD. At first, these medications appear to alleviate the symptoms of manic and depressive episodes, which explains why so many people with bipolar illnesses become addicted.
Sex drive	Sexual activities may be affected by the illness. Patients may suffer from hypersexuality during a manic episode. This might put them at a greater risk of harmful activities, such as developing a sexually transmitted infection (STI).

4. EEG biomarkers in BD diagnosis

4.1. EEG microstates

What are EEG microstates? EEG microstates are fleeting, or quasi-stable, patterned states [8]. These microstates typically last between milliseconds and seconds and represent the most fundamental manifestations of human brain processes; they are hence dubbed "the atoms of cognition". Initially, microstates

estimations and analysis were performed using alpha-band activity, but nowadays, wider bandwidth EEG bands are commonly used. Due to the quasi-stability of microstates, global EEG topography is stable, but intensity and polarity may fluctuate [11].

How is an EEG microstate obtained? In one observational study, researchers aimed to apply electroencephalographic microstates to classify SCH and BD [11]. Twenty SCH patients (mean age 25.2±6.8, n=15 females), thirty-four BD patients (mean age 22.8±4.12, 13 female), and thirty-five HCs (mean age 24.9±6.2, n=25 females) were recruited. EEG data of the patients were recorded for 3 min at a resting state while they closed their eyes and sat on a chair comfortably. Fig. 2 illustrates the process of EEG data collection and processing. Specifically, data were acquired using a 64-channel device (Brain Product) and sampled at 5,000 Hz. The impedances were maintained under 5 Kohm. MATLAB 2013b and the MATLAB EEGLAB toolbox were employed to pre-process the data. A 48-52 Hz Parks-McClellan stop-band notch filter was employed to reduce electric interference from the 50 Hz line. Then, the data were band-pass filtered (0.1-40 Hz) and downsampled to 250 Hz. The data then were segmented into epochs of 2 s and poor epochs were discarded with evident muscle activation after being verified visually. Besides this, the researchers also employed independent component analysis (ICA) to eliminate the oculomotor component when necessary. Finally, the data were referred to as the average reference and bandpass filtered to 2-20 Hz. Microstate analysis was performed using MATLAB 2013b. Figure 3 depicts the four microstate classes for SCH patients, BD patients, and HCs. The 4 microstate classes explained 76.2, 74.1, and 76.0 percent of the global variation in SCH, BD, and HC patients, respectively, while one-way ANOVA revealed no significant differences (p=0.126) between SCH, BD, and HC patients. To exclude the effect of the varied template maps in different epochs, the global field power (GFP) of the EEG data was estimated for each participant with the longest length. In the first stage, all GFP peaks from all participants were clustered to generate the template maps. In the second step, all GFP peaks from the first step were clustered to identify which class each subject's template maps at each time point belonged to. When clustering, the modified k-means method with k restricted to 100 was utilized. In order to maintain continuity with earlier investigations, the EEG data were separated into four microstate topographies. The goal of this study is to analyse the four EEG microstates; we acquire four typical microstates for three distinct groups and all participants, independent of their medical condition. Across all groups, a global segmentation was performed, and one set of template maps was fitted to the initial data. These four all-group template maps as shown in Fig. 3 were then utilized to derive the microstate properties. Based on the microstate transition, the chance of transition between every two microstates back and forth was estimated. Twelve observed possible transitions were considered [11].

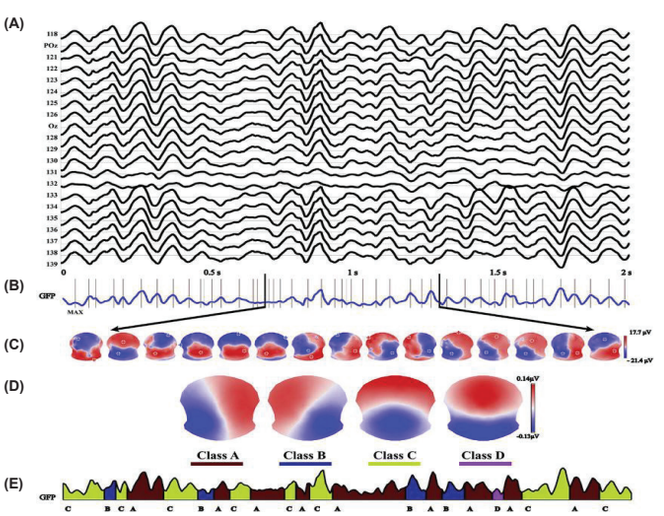


Fig. 2. A illustration of the EEG microstate data collection and processing [9].

EEG microstates in BD: As mentioned before, EEG microstates are considered one of the most state-of-the-art techniques in diagnosing many serious psychiatric disorders. In some studies, this even alters the conventional questionnaire diagnostic tools. Although many psychological studies have been conducted to diagnose SCH, few studies use this innovative technique to distinguish BD from other psychiatric disorders. According to a study, the researchers investigated the microstate features of EEG in patients with SCH, BD, and HCs to give an electro-directional physiological explanation and any mutual mechanism for the observed brain dysfunction of SCH and BD [11]. Another study, employed EEG microstates to investigate any potential resting state temporal dynamics abnormalities in BD. The findings are stated in Fig. 3 [12].

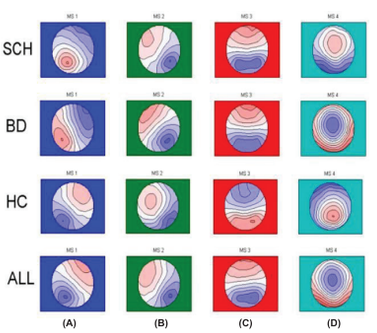


Fig. 3. EEG microstates (A, B, C, D) in SCH, BD, HCs, and ALL for all types of subjects [8].

EEG microstate data were transformed into four types of microstate topographies (A, B, C, D). From the experimental results, the authors concluded that BD patients exhibited a greater presence of microstates A compared to that of SCH [11]. Regarding microstates B, BD patients exhibit a greater presence of those microstates than that of SCH [11], and a reduced presence of microstates B in BD compared to HCs [12]. BD patients displayed less presence of microstates C than that of SCH [11]. Finally, BD patients displayed a greater presence of microstates D than that of SCH [8]. Table 2 illustrates all the findings from [11] and [13].

As can be seen from Table 2, bipolar patients exhibit greater frequent switching between microstates A and C, D and vice versa compared to SCH patients. Therefore, in clinical practice, microstate analysis can play a crucial role in diagnosing patients more accurately and treating them properly.

Table 2. EEG microstate findings summary in BD and SCH.

Microstate	BD	SCH
 A	BD patients exhibit a greater presence of microstate class A [11].	SCH exhibits less presence of microstate classes A and B [13].
 B	BD patients exhibit a greater presence of microstate class B [8]. However, when compared to HCs, BD patients indicated a reduction in microstate B [13].	
 C	BD patients exhibit less presence of microstate class C compared to the SCH group [11].	SCH patients indicate a greater presence of microstate class C and class D compared to the BD group [11].
 D	BD patients exhibit less presence of microstate class D compared to the SCH group [11].	

4.2. EEG resting states

What are EEG resting states? Recently, neuroimaging technologies have emerged as crucial tools for evaluating the brain status of individuals with consciousness issues (DOC). The recording and analysis of resting-state EEG have been extensively used by physicians owing to their comparatively cheap cost and portability [13]. EEG represents the electrical activity of the underlying neurons and includes information on neuronal population oscillations, the channel of information flow, and brain activity networks. Several characteristics obtained from EEG signal processing techniques have been presented to characterize the electrical properties of the brain with DOC. The calculation of these characteristics is difficult for physicians who are attempting to interpret the related physiological implications and then use them clinically. Resting EEG data indicate spontaneous neural activity that is pertinent to the underlying brain state. In contrast, relevant characteristics generated from resting-state EEG may be useful for monitoring the brain status of DOC patients and contributing to care-related decisions. For that reason, EEG resting states are well-investigated to enhance our knowledge of the EEG features of DOC [13].

EEG resting states in BD: Another method, also one of the most promising EEG features in diagnosing BD, is the resting state. Resting-state EEG evaluations are designed to determine intrinsic brain activity not induced by a task. In an attempt to use EEG to diagnose BD, studies have been carried out [14-17].

D.J. Kim, et al. (2013) [14] employed disturbed resting state EEG synchronization to diagnose BD. It is hypothesized that a crucial aspect of BD is the disruption of functional connectivity, which represents disturbances of synchronization and oscillations within brain networks. During this project, the authors aimed to discover if individuals with BD have different synchronization or network characteristics in their resting EEG. In another study, resting EEG was once again utilized to distinguish between individuals with first-episode SCH, bipolar psychosis, and their first-degree relatives [15]. The researchers also employed resting EEG methods to investigate whether SCH and bipolar patients suffer from altered cortical functional networks [16]. Last, but not least, researchers also examined the relationships between cortical network indices and psychiatric, clinical, or cognitive measures to broaden current knowledge about both brain disorders.

Additionally, to differentiate between the following psychotic disorders SCH, BD and methamphetamine-induced psychotic disorder, F.M. Howells, et al. (2018) [17] used electroencephalographic delta/alpha frequency activity. Regarding their methodology, EEG was recorded under three different testing conditions: resting eyes open, resting eyes closed, and while performing cognitive tasks (visual continuous performance task). From those aforementioned studies, EEG data were extracted and processed before presenting the results. The EEG is a system that measures the frequency of brain waves. Generally, the raw EEG is described in terms of frequency bands: delta, alpha, gamma, and theta.

- For *delta*, when compared to multiple personality disorder (MPD) for the right hemisphere, bipolar patients exhibited lower delta/alpha frequency activity in resting eyes closed. However, when compared to HCs, bipolar patients displayed higher delta/alpha frequency activity during resting eyes opened [17].

- For *alpha*, BD patients exhibited a decrease in mean synchronization [13]. Also, in another study, there was a reduction in alpha activity in BD patients and reduced peak alpha frequencies [15].

- For *gamma*, gamma frequency band activities increased in bipolar patients when compared to first-episode SCH patients [15].

- For *theta*, theta frequency band activities increased in bipolar patients when compared to first-episode SCH patients [15]. Table 3 below summarizes the main findings from the aforementioned research.

Table 3. A brief summary of brain wave changes in BD patients compared to other psychiatric conditions.

Brain wavebands	Frequency range	Main findings	References
Alpha	8-12 Hz	Patients with BD demonstrated a reduction in mean synchronization. In addition, there was a decrease in alpha activity and peak alpha frequencies in BD patients.	[14, 15]
Gamma	>30 Hz	Gamma frequency band activities increased in BD patients when compared to first-episode SCH patients.	[15]
Theta	4-8 Hz	Theta frequency band activities increased in BD patients when compared to first-episode SCH patients.	[15]
Delta	0.5-4 Hz	In contrast to individuals with MPD for the right hemisphere, bipolar patients displayed decreased delta/alpha frequency activity in the resting/eyes closed state. However, compared to HCs, bipolar patients exhibited more delta/alpha frequency activity during eyes-open rest compared to CHs.	[17]

4.3. Other EEG features

Qualitative EEG (qEEG) is well-known as an optimal diagnostic tool in psychiatric diagnosis. With that in mind, the following study was carried out by comparing the differences and similarities between women with attention deficit hyperactivity disorder (ADHD), women with BD, and female HCs using qEEG [18]. In other words, they examined the similarities in EEG spectral power abnormalities during rest and cognitive function between women with ADHD and women with BD. The researchers discovered that during resting state conditions, the post-hoc tests displayed significantly higher absolute theta power in the ADHD group compared to HCs, but not during the Cue (CPT-OX). Their statistical technique also demonstrated that the BD group had significantly higher absolute theta power than the HCs, although not during the Cue (CPT-OX). Finally, during resting state or CPT-OX, no significant changes in absolute theta power between the ADHD and BD groups were found. While HCs demonstrated a task-related increase in absolute theta activity from resting to cognitive task, neither the BD nor the ADHD groups showed any significant changes. These findings might indicate evidence for commonalities in brain dysfunction between ADHD and BD. In both illnesses, absolute theta power may also serve as a marker of neurobiological processes.

Domain analysis, state-of-the-art techniques, and neural networks: A. Khaleghi, et al. (2015) [19] wants to utilize various domain analyses, state-of-the-art methodologies, and neural networks to classify the EEG signal of adolescents with BD I and BD II. Four alternative feature selection methods: mutual information maximization (MIM), conditional mutual information maximization (CMIM), fast correlation-based filter (FCBF), and double input symmetrical relevance (DISR) were used to determine the most discriminative features. Finally, for each feature selection technique, the extracted features were fed into a multi-layer perceptron (MLP) neural network as a classifier, and the system's performance was calculated. For the first time, the researchers used quantitative EEG processing techniques to classify BD subtypes. It was discovered that the PSD (power spectrum density) of the F3 and F4 channels of BD I and BD II were almost identical, although the spectrograms varied noticeably. As a result, spectrum-based features were more effective for distinguishing sub-types. The results demonstrate that combining DISR with MLP yields the greatest accuracy of 91.83%, which is acceptable for a medical diagnostic system. Furthermore, the following order of best to the worst performance in four distinct feature selection algorithms was provided as follows: DISR > CMIM > MIM > FCBF (average accuracy rate for test data set: 91.83>89.67>86.33>84.61).

The error-related negativity (ERN) and event-related brain potential (ERP): Some other EEG features that captured the attention of neurologists are the ERN and ERP. As far as we know, ERN is an electrical measure believed to reflect changes in dopamine when individuals make errors while performing cognitive activities, while ERP is defined as the brain's measurable reaction to a sensory, cognitive, or motor event. A study recorded EEG data from 16 patients with BD (euthymic state) and 14 matched HCs while executing a speeded two-choice reaction-time paradigm (flanker task) while EEG measurements were obtained to explain lower performance monitoring in BD using EEG signals [20]. The two groups' behavioural and ERP measures were then compared. As a result, S. Kim, et al. (2020) [16] found that, despite being euthymic, individuals with BD suffered more depressive symptoms than CHs. While no behavioural abnormalities were detected, when residual mood effects were considered, individuals with BD had lower ERN amplitudes than CHs. By observing lower ERN amplitudes, the researchers concluded that the reduction in ERN values in the BD group represented less performance monitoring that can expand our current knowledge of executive functioning in BD.

EEG power, cordance, and coherence: Researchers utilized EEG power, cordance, and coherence differences between unipolar and bipolar depression in this study [20]. A total of 25 bipolar and 56 unipolar depression patients were enrolled in the research. In addition to the resting state EEG, sociodemographic and clinical variables were obtained. When parametric assumptions were satisfied, data was examined using multivariate and repeated analyses of variance. No variations in EEG absolute power or frontal asymmetry indices were found between UD and BD in the research [18]. In terms of cordance, there were significant group differences in right theta cordance values ($p=0.031$). In terms of coherence, BD showed significantly higher central-temporal theta ($p=0.003$), parietal-temporal alpha ($p=0.007$), and theta ($p=0.001$) coherence than UD patients. Finally, the right frontal-central ($p=0.007$) and central inter-hemispheric ($p=0.019$) regions of the brain showed weaker alpha coherence in BD.

EEG entropy: As far as we know, EEG entropy is an EEG features that applies the idea of entropy to time series like EEG to provide a means to measure, in a statistical sense, the level of uncertainty or unpredictability in the pattern. The goal of this recent study [21] was to investigate if a deficit in EEG entropy might be applied as a biomarker for BD and SCH patients with impaired function. The researchers examined EEG spectral entropy modulation during a P300 task in 79 SCH patients (31 first episodes), 29 BD, and 48 HCs. Then, they performed a 3-tone auditory oddball task (P300 task), which is a well-established cognitive activity and event-related potential with a known spatial distribution. Meanwhile, C. Tas, et al. (2015) [22] discovered that patients with chronic or first-episode SCH and patients with BD had substantial spectral entropy modulation impairments with task performance, despite no significant pre-stimulus spectral entropy variations. MRI results also indicated the structural connectivity values were unrelated to the modulation of spectral entropy. Finally, the abnormalities were independent of treatment dosages, and there was no difference in spectral entropy modulation between individuals receiving antipsychotics, lithium, benzodiazepines, or antidepressants. As a result, in our review, Fig. 4 below illustrates important information of EEG features in bipolar diagnosis.

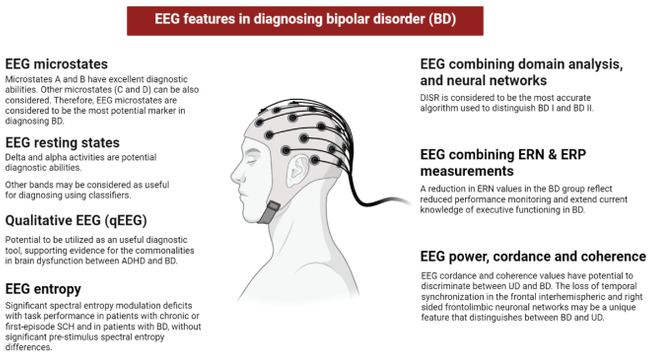


Fig. 4. A summary of EEG features in diagnosing BD.

5. Conclusions

In conclusion, this review presented a systematic mapping of EEG BD biomarkers, as well as a number of recent research works each with a brief explanation and comparison as well as a discussion of their state-of-the-art achievements. In this review, we categorized potential biomarkers into seven categories and concluded that the microstate features in EEG results would be the most potential marker in diagnosing BD because of the abnormal presence of microstate B in EEG topography. Additionally, the abnormalities in delta and alpha signals from EEG resting states were considered to be potential markers for diagnosing BD. Moreover, using other EEG features such as qualitative EEG (qEEG), EEG combining domain analysis and neural networks, EEG combining ERN and ERP measurements, EEG power, cordance and coherence, and EEG entropy could also potentially broaden our understanding of mental health problems like BD.

CRediT author statements

Tien Dat Nguyen, Van Toi Vo, Thi Thanh Huong Ha: Data curation, Software, Visualisation, Investigation, Writing original draft preparation.

ACKNOWLEDGEMENTS

This research is funded by Vietnam National University, Ho Chi Minh city under grant number NCM2020-28-01.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] I. Grand, M. Berk, B. Birmaher, et al. (2016), "Bipolar disorder", *The Lancet*, **387(10027)**, pp.1561-1572, DOI: 10.1016/S0140-6736(15)00241-X.

[2] The National Alliance on Mental Illness (2017), Bipolar Disorder, https://nami.org/About-Mental-Illness/Mental-Health-Conditions/Bipolar-Disorder, accessed 14 February 2022.

[3] M. Sanches, I.E. Bauer, J.F. Galvez, et al. (2015), "The management of cognitive impairment in bipolar disorder: Current status and perspectives", *American Journal of Therapeutics*, **22(6)**, pp.477-486, DOI: 10.1097/MJT.0000000000000120.

[4] R. Degabriele, J. Lagopoulos (2009), "A review of EEG and ERP studies in bipolar disorder", *Acta Neuropsychiatrica*, **21(2)**, pp.58-66, DOI: 10.1111/j.1601-5215.2009.00359.x.

[5] B. Murray, J. Gottlieb, M.P. Hidalgo, et al. (2020), "Chronobiologic assessment methods for bipolar disorders: Current research protocols and clinical applications", *Bipolar Disorders Special Issue, The 22nd Annual ISBD Conference*, **22(S1)**, pp.41-42, DOI: 10.1111/bdi.12934.

[6] M. Vrabie, V. Marinescu, A. Talaşman, et al. (2015), "Cognitive impairment in manic bipolar patients: Important, understated, significant aspects", *Annals of General Psychiatry*, **14(1)**, DOI: 10.1186/s12991-015-0080-0.

[7] S. Gruber, K. Rathgeber, P. Braunig, et al. (2007), "Stability and course of neuropsychological deficits in manic and depressed bipolar patients compared to patients with major depression", *Journal of Affective Disorders*, **104(1-3)**, pp.61-71, DOI: 10.1016/j.jad.2007.02.011.

[8] R.E. Kelly, W. Mansell, A.M. Wood, et al. (2011), "Extreme positive and negative appraisals of activated states interact to discriminate bipolar disorder from unipolar depression and non-clinical controls", *Journal of Affective Disorders*, **134(1-3)**, pp.438-443, DOI: 10.1016/j.jad.2011.05.042.

[9] H. Dinc, N.E. Boyacioglu, N.K. Ozcan, et al. (2019), "Reproductive and sexual health in women with bipolar disorder: A comparative study", *Dusunen Adam The Journal of Psychiatry and Neurological Sciences*, **32**, pp.23-32, DOI: 10.14744/DAJPNS.2019.00004.

[10] E. Marengo, D.J. Martino, A. Igoa, et al. (2015), "Sexual risk behaviors among women with bipolar disorder", *Psychiatry Res.*, **230(3)**, pp.835-838, DOI: 10.1016/j.psychres.2015.10.021.

[11] F. Wang, K. Hujjaree, X. Wang (2021), "Electroencephalographic microstates in schizophrenia and bipolar disorder", *Frontiers in Psychiatry*, **12**, DOI: 10.3389/fpsyt.2021.638722.

[12] M.I. Tomescu, T.A. Rihs, R. Becker, et al. (2014), "Deviant dynamics of EEG resting state pattern in 22q11. 2 deletion syndrome adolescents: A vulnerability marker of schizophrenia?", *Schizophrenia Research*, **157(1-3)**, pp.175-181, DOI: 10.1016/j.schres.2014.05.036.

[13] F. Vellante, F. Ferri, G. Baroni, et al. (2020), "Euthymic bipolar disorder patients and EEG microstates: A neural signature of their abnormal self experience?", *Journal of Affective Disorders*, **272**, pp.326-334, DOI: 10.1016/j.jad.2020.03.175.

[14] D.J. Kim, A.R. Bolbecker, J. Howell, et al. (2013), "Disturbed resting state EEG synchronization in bipolar disorder: A graph-theoretic analysis", *NeuroImage: Clinical*, **22(2)**, pp.414-423, DOI: 10.1016/j.nicl.2013.03.007.

[15] B.A. Clementz, S.R. Sponheim, W.G. Iacono, et al. (1994), "Resting EEG in first□episode schizophrenia patients, bipolar psychosis patients, and their first□degree relatives", *Psychophysiology*, **31(5)**, pp.486-494, DOI: 10.1111/j.1469-8986.1994.tb01052.x.

[16] S. Kim, Y.W. Kim, M. Shim, et al. (2020), "Altered cortical functional networks in patients with schizophrenia and bipolar disorder: A resting-state electroencephalographic study", *Frontiers in Psychiatry*, **11(661)**, DOI: 10.3389/fpsyt.2020.00661.

[17] F.M. Howells, H.S. Temmingh, J.H. Hsieh, et al. (2018), "Electroencephalographic delta/alpha frequency activity differentiates psychotic disorders: A study of schizophrenia, bipolar disorder and methamphetamine-induced psychotic disorder", *Translational Psychiatry*, **8(1)**, DOI: 10.1038/s41398-018-0105-y.

[18] A.S. Rommel, G.L. Kitsune, G. Michelini, et al. (2016), "Commonalities in EEG spectral power abnormalities between women with ADHD and women with bipolar disorder during rest and cognitive performance", *Brain Topography*, **29(6)**, pp.856-866, DOI: 10.1007/s10548-016-0508-0.

[19] A. Khaleghi, A. Sheikhani, M.R. Mohammadi, et al. (2015), "EEG classification of adolescents with type I and type II of bipolar disorder", *Australasian Physical and Engineering Sciences in Medicine*, **38(4)**, pp.551-559, DOI: 10.1007/s13246-015-0375-0.

[20] A.M. Morsel, M. Morrens, A. Temmerman, et al. (2014), "Electrophysiological (EEG) evidence for reduced performance monitoring in euthymic bipolar disorder", *Bipolar Disorders*, **16(8)**, pp.820-829, DOI: 10.1111/bdi.12256.

[21] V. Molina, A. Lubeiro, R.L. Garcia, et al. (2020), "Deficits of entropy modulation of the EEG: A biomarker for altered function in schizophrenia and bipolar disorder?", *Journal of Psychiatry and Neuroscience*, **45(5)**, pp.322-333, DOI: 10.1503/jpn.190032.

[22] C. Tas, M. Cebi, O. Tan, et al. (2015), "EEG power, cordance and coherence differences between unipolar and bipolar depression", *Journal of Affective Disorders*, **172**, pp.184-190, DOI: 10.1016/j.jad.2014.10.001.

Effects of settling time on the flocculation progress and treatment performance in the co-culture of microalgae-activated sludge photobioreactor

Hong Hai Nguyen[1, 2], Hong Ngoc Luong[1, 2], Ngoc Kim Qui Nguyen[1, 2], Le Phuong Uyen Nguyen[1, 2], Phuong Thao Nguyen[1], Cong Sac Tran[1], Xuan Thanh Bui[1, 2*]

[1]Key Laboratory of Advanced Waste Treatment Technology, Ho Chi Minh City University of Technology, Vietnam National University - Ho Chi Minh City, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam

[2]Ho Chi Minh City University of Technology, Vietnam National University - Ho Chi Minh City, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam

Received 30 July 2021; revised 8 October 2021; accepted 14 October 2021

Abstract:

The application of microalgae-based wastewater treatment systems using wastewater as a source of nutrients has been successfully developed in recent years and has brought about positive results in wastewater treatment and microalgal biomass recovery while producing valuable products. This study presents the application of a microalgae and activated sludge (AS) co-culture in a low agitation photobioreactor (aPBR), which could reduce energy usage. In addition, the results demonstrate the role of settling time on co-culture flocculation progress and wastewater treatment performance. The average Chemical Oxygen Demand (COD) removal was up to 76.1% of co-culture and was achieved using a photobioreactor system with a low agitation speed of 80 rpm. Moreover, a high biomass growth was observed to be coupled with high nutrient removal. After 63 days of operation, heavy flocs with excellent settling ability were dominant in the photobioreactor. The findings also suggest that the short settling time of 30 min could enhance the bio-flocculation of the microalgae-AS biomass. Besides, short settling times can be considered as an effective and easy strategy in co-culture operations to promote heavy floc formation. This would be a preliminary step for activated algae granulation in a co-culture system.

Keywords: **activated sludge, co-culture, flocculation, microalgae, wastewater treatment.**

Classification number: **5.3**

1. Introduction

Nowadays, biological processes are mainly applied to wastewater treatment plants, especially the activated sludge process (ASP). However, there are still many drawbacks of biological processes, for example, the need for pre-treatment, the high energy requirements for aeration and stirring, and low nutrient removal capacity [1, 2]. Moreover, the removal of phosphorus in wastewater, which is usually done by chemical and physical processes like flocculation, comes with high cost and low treatment efficiency. Due to these existing shortcomings, the application of microalgae-based wastewater treatment systems using wastewater as a source of nutrients has been successfully developed in recent years and brought about positive results to the treatment of nitrogen (N), phosphorus (P), and algal biomass recovery [3-5]. The advantage of using algae corresponds to the ability to remove nitrogen and phosphorus in wastewater through the assimilative process of biomasses, which has been studied as a post-treatment process [6], in urine [7], aquaculture wastewater [8], and livestock wastewater [9]. In addition, utilizing microalgae to convert carbon emissions into biomass is regarded as one of the most cost-effective methods of CO_2 reduction, and the obtained biomass of algae could become a high-profit product like biofuels, food sources, or nutraceuticals [10, 11]. However, the widespread adoption of microalgae-based technology is constrained due to high cost of operation and biomass harvesting. It is estimated that about 20-30% of operating costs during the cultivation of microalgae are from microalgal biomass harvesting through coagulation/flocculation, centrifugation, or flotation [12].

The symbiotic co-culture of microalgae and AS has been increasingly taken more advantage of due to the consistent ability of carbon dioxide to produce oxygen, which decreases the need for aeration, increases nutrient removal rates, and promotes high microalgae recovery [13-15]. On the other hand, a phenomenon called "bio-flocculation" appears when microalgae and other microorganisms are associated in a co-culture. Extracellular polymer compounds such as polysaccharides and proteins will be secreted into the medium to form flocs that cause bio-flocculation when co-cultivating microalgae and AS [16, 17]. Through the flocculation process, the size of the flocs increase by cell aggregation, which enhances the settling rate of the biomass. Bio-flocculation effectively eliminates the need for chemical flocculants, and thus represent an inexpensive, non-feasible, and toxic alternative for effective biomass harvesting. During flocculation, the sizes of the floc cells are increased by an aggregation of cells that can enhance the biomass settling rate. However, bio-flocculation is currently not widely applied in harvesting steps because of the high production cost of bio-flocculants and the challenge of controlling the process on an industrial scale. An ideal bio-flocculant should be inexpensive, nontoxic, and effective at low concentrations and it should preferably be derived from non-fossil fuel sources, thus being sustainable and renewable.

In this study, microalgae and activated sludge as co-cultures were experimented in a low aPBR to reduce the effort of energy usage. To promote an easy way to harvest the microalgae by gravity settling, two settling times of 3 h and 0.5 h were investigated in terms of wastewater treatment ability and bio-flocculation process of the co-culture system.

*Corresponding author: Email: bxthanh@hcmut.edu.vn

2. Materials and methods

2.1. Microalgae-activated sludge co-culture

The co-culture of microalgae and activated sludge at the ratio 5:1 (%w:%w) was chosen for this experiment due to the higher nutrient removal efficiencies (5-40%) and biomass growth rate comparing to other inoculation ratios. The microalgae strain *Chlorella Vulgaris* was used in this study due to its well-known benefits (e.g., biofuels, medical application, etc.) [18]. The strain was taken from Aquaculture Research Institute 2, Ministry of Agriculture and Rural Development, Vietnam. The aerobic AS was taken from a membrane bioreactor (MBR) treating supermarket wastewater in Ho Chi Minh city. The initial biomass concentration of co-culture was 400 mg/l.

2.2. Synthetic wastewater

In this study, the co-culture was cultivated by synthetic wastewater with characteristics of 400±20 $mg.l^{-1}$ of COD. Total nitrogen (TN) was present in the form of ammonium at 40±1.4 $mg.l^{-1}$, and nitrate and nitrite nitrogen were not detected. The N/P mass ratio was about 10. In addition, 1 ml of trace elements was prepared following Bold's basal medium (BBM) [19, 20] per liter of feeding wastewater. The initial pH of synthetic wastewater was controlled by $NaHCO_3$ (10%) in the range 7-7.5 for the biological growth of microorganisms.

2.3. Experimental system set-up

The tubular PBR system consisted of a closed cylindrical acrylic column with a height of 41.75 cm and diameter of 17 cm. The water level was 30 cm with a total volume of 10 l and a working volume of 7 l. The PBR was operated at a low agitation speed of 80 rpm without aeration, which is called an PBR. The PBR system was installed in a wooden box with a thickness of 10 mm to prevent loss of light and temperature fluctuations (maintained around 27-32°C). Light was provided by 3800-4000 lux LED lights (SMD 5050, China) with a light-dark cycle of 14:10 (h).

2.4. Operating conditions

The PBR system was operated with synthetic wastewater in a sequencing batch process of 24 h each cycle with a volume exchange ratio of 50%. Initially, one cycle consisted of 15 min influent addition, 20.5 h reaction phase (with agitation), and 3 h settling. The settling time was reduced based on the settling velocity at the steady state as follows: day 1-day 63 was 3 h, then day 63-day 90 was 0.5 h. The reaction phase correspondingly increased and a 15-min effluent withdrawal of the suspended biomass was done while the settled biomass was remained for use in the next batch. The effluent was withdrawn by an effluent valve and stored in an Erlenmeyer flask for water quality analysis within 1 d of sampling.

2.5. Analysis method

The samples of each reactor were collected daily. The influent was prepared weekly and analysed. For characterization of the microalgae and AS biomass, the mixed liquid-suspended solids were collected via the middle valve in the reaction phase. Then, the COD, ammonia, NO_3-N, NO_2-N, total phosphorus (TP), and Mixed liquor suspended solids (MLSS) concentrations were analysed in accordance with standard methods [21].

In this research, microalgal biomass evaluation was conducted based on Chlorophyll-a (Chl-a) in the mixture of AS and microalgae [22]. Chl-a concentration was used to represent the growth of the algal biomass [23]. Chl-a in mixture of AS and microalgae was extracted by acetone solution [24] and a 20-ml well-mixed biomass sample was taken from the reactor to analyse sludge and algae characteristics. The biomass concentration and Chl-a concentration correlation was done through the following calibration curve of Chl-a and MLSS: y=7.7191x+759.04.

3. Results and discussion

As denoted in Fig. 1, after 11 days of operation, the total biomass concentration increased from 0.4 to 1.14 g/l for the co-culture system i.e., microalgae:AS inoculum ratio of 5:1 was strongly decreased to 1:2.33. It was observed during that time the flocculation process had yet not started (before day 63), the AS was dominant (microalgae fraction was below 10%), and the microalgae concentration decreased from 0.5 to around 0.105±0.041 g/l. The biomass fractions of microalgae and activated sludge also changed following the formation of bio-flocculants (Fig. 1). It is important to note that a major change of biomass fraction was observed between the initial and final stage of the experiment: the fraction of activated sludge remained stable at 84% in the first 30 day, then it decreased gradually to around 18% after day 63 when bio-flocculation occurred.

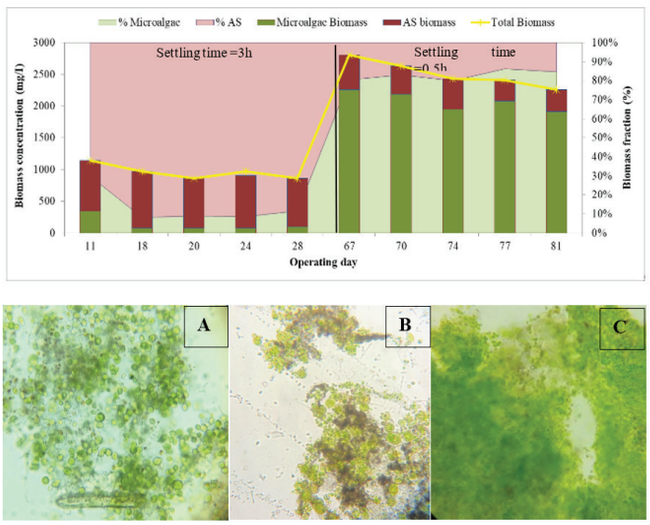


Fig. 1. Growth of microalgae and AS in the PBR at different settling times under low agitation speed (top) and microscopic observation on (A) day 14, (B) day 53, and (C) day 83 (x40) (bottom).

Another study evaluated the growth of co-cultures in different microalgae where AS inoculum ratios were tested from 1:1 to 9:1, with a stirring speed of 100 rpm, and a low COD/N wastewater ratio of 2.5:1 and showed the opposite result [25]. At the end of the experiment, the microalgae fraction inside all the co-culture systems were above 70% with the dominance of microalgae. The reason being that the composition of feeding wastewater for microalgae and AS comprised a high COD/N ratio (10:1). As it is known that microalgae exhibit shorter lifetimes than AS, there might have been a struggle to compete with AS for the nutrient uptake [26] and this high COD/N ratio could elevate the biomass growth rate of AS rather than that

of microalgae due to sufficient COD in wastewater. Besides, under conditions of total biomass concentration over 1.0 g/l (before Day 63) with high AS density, a stirring speed of 80 rpm may not be sufficient to help microalgae access to light and overcome the dense cover of AS, thereby slowing their growth.

3.1. Biomass growth and flocculation

Despite growth being adversely affected, bio-flocculation was promoted and began being recognized clearly after day 63. In addition, reducing the settling time from 3 to 0.5 h enforces the bio-flocculation process by increasing the amount of small-sized biomass being washed out of the system. This will help classify the groups of microorganisms that tend to create high-density flocs, which are retained in the aPBR, and remove the low-density groups of microorganisms that tend to be suspended, which are released after discharge phase.

When appearing together in one co-culture, microorganisms can secrete extracellular polymer compounds, such as polysaccharides and proteins, to form bio-flocculants with microalgae. These results lead to a strong increase of microalgal biomass [27]. In the formed flocs, AS tends to act as a core to adhere the externally suspended algal cells (see Fig. 1B). This can be explained by the fact that microalgae are photoautotrophs, and the way they appear around the border of flocs could give them easier access to light outside the aPBR and overcome the covering phenomena by the AS. In addition, after reducing the settling time, the dense structure of microalgae-AS biomass (Fig. 1C) with higher settling rates in heavy flocs occupy less space inside the PBR but contain more microorganisms, which help to increase light penetration inside the co-culture and show positive results in enhancing microalgae growth.

With an agitation speed of 80 rpm, the microalgal growth significantly prevailed over AS growth. In fact, the bio-flocculant of algal-bacteria can accelerate the microalgal growth rates, enhance wastewater removal efficiency, boost the carbohydrates and lipid content in microalgae, promote microalgal flocculation processes, and facilitate microalgal cell wall disruption [28].

After reducing the settling time, aggregation of the biomass could already be observed, which also obviously influenced the settling velocity. Indeed, the co-culture flocculant obtained after day 63 exhibited a higher settling velocity of 3.56 m/h than only microalgae culture, which was less than 0.0036 m/h [29]. Therefore, after the flocculation process, the settling time was efficiently reduced from 3 to 0.5 h without any negative affect to the co-culture system. These results indicate that the effort to harvest microalgae could be reduced if the co-culture in PBR is operated under an appropriate operating condition by stimulating the flocculation process. However, bio-flocculation is a complex process that combines microalgae self-aggregation (i.e., filamentous microalgae), the flocculation of produced algae, and bio-flocculant-producing bacterial species (i.e., EPS) [30]. Agitation is responsible for creating cell-to-cell contact between microalgae and bacteria without cell stress or lysis over a period of time [31]. According to O. Tiron, et al. (2017) [29] and M. Xu, et al. (2015) [32], under certain hydrodynamic conditions, the formed flocs are strong and similar in size as the weaker ones break and re-flocculate until they are strong enough to resist the shear stress, i.e., turbulence, which can promote bio-flocculation by increasing the collision frequency while the shear stress can promote microbial cell aggregation.

3.2. Nitrogen removal

In terms of nitrogen pollutants, poor ammonium removal in co-cultures with low agitation speed was due to low hydraulic retention time and high ammonia concentration in feed water [33]. Fig. 2 has shown the ammonium removal performance during the whole experiment with initial concentrations of 40 mg/l for a 2-day hormone replacement therapy (HRT). After 1-2 days of co-cultivation, although the nitrifiers and microalgae did not present a good adaptation ability in the new environment in co-culture, heterotrophs in the activated sludge showed good ability in assimilated COD. During that time, when the ammonium removal only reached higher than 20%, nitrate concentration also showed a negligible value (<0.05 mg/l). However, after this beginning period, the nitrifiers showed good nitrification process in the following days. Nitrate concentration in the effluent has shown that when the first period had a dominance of bacteria in the microbial consortium, it showed better nitrification than the second one. Accordingly, the average removal efficiency of ammonium in the 3 h settling period was 42%, which is significantly higher than the 28% of the 0.5 h settling period. In the 3-h settling period, if ammonium removal took place by nitrification process of AS, it is uptake into biomass by microalgae in the 0.5 h settling time. In microalgae-bacteria systems, microalgae mainly provide oxygen for bacteria to remove organic substances and actively uptake nutrients [34]. However, without atmospheric CO_2, the photosynthesis of microalgae was limited as the only CO_2 supply was from AS, which leads to longer time for ammonium accumulation into microalgal biomass. At the end of the experiment, ammonium removal efficiency

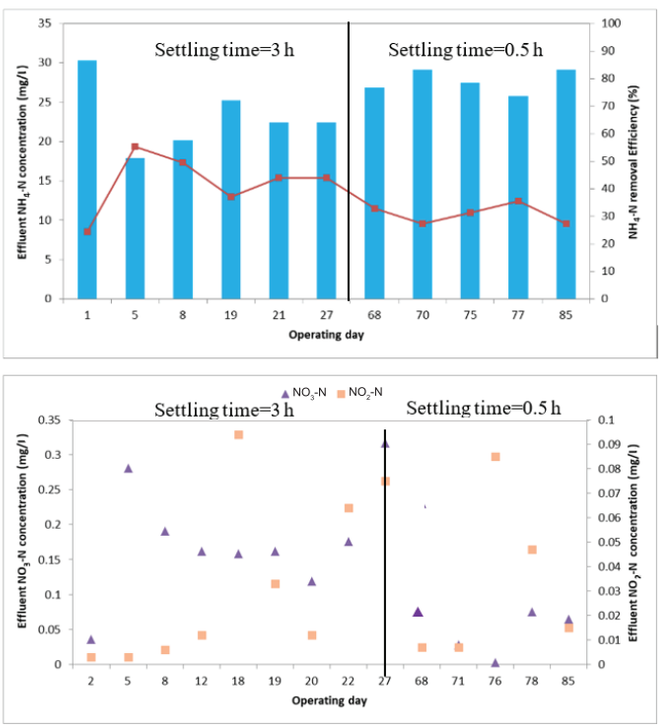


Fig. 2. Ammonia removal (top) and nitrogen compounds (bottom) in the PBR with different settling time.

reduced to lower than 30%, and this demonstrated a solution to increase the ammonium removal ability of co-culture that should be investigated in future studies.

On the other hand, the study by C.S. Lee, et al. (2015) [24] showed that nitrogen removal efficiency was positively related to light intensity and microalgal biomass. In this study, total biomass reached its highest concentration of 2.509±0.212 g/l on day 67 of the operation but tended to decrease to approximately 2.200 g/l by the end of the experiment. An increase of biomass concentration could lead to reduced light transmission, which may significantly reduce nutrient removal performance.

3.3. Phosphorus removal

Phosphorous in wastewater is primarily removed by biomass uptake and precipitation [35]. Throughout our experiment, the pH value remained in the range of 6.5-8.5, which means that phosphorous removal by precipitation was negligible. It can thus be confirmed that the phosphorus removal mechanism principally depends on biomass uptake capacity. Fig. 3 shows that after only 2 days of experiment, TP removal efficiency reach over 90%, which indicated that co-culture in PBR systems have a good ability for P uptake without aeration provision. During the whole experiment, TP removal efficiency varied from 90-97% without any significant fluctuation implying that there was no strong relationship between TP removal and settling time. Moreover, it seemed that bio-flocculation did not affect the phosphorus and nitrogen removal either. One possible explanation is due to the fact that nutrient removal is dependent on the O_2 generation rate and coupling with CO_2 from AS biodegradation, which was initially proportional to the algal density within a certain range. However, when the biomass concentration inside the PBR overcame the maximum supply of light, the nutrient removal efficiency rapidly reached its limit. After the microalgae concentration reached that value, the nutrient uptake rate possibly remained at a certain level, without any increase, following the growth of microalgae [36]. The remark from the results was the bio-flocculation could have occurred under stress condition (light-shading, high COD/N ratio, without aeration) and not coupling with high nutrient removal efficiency.

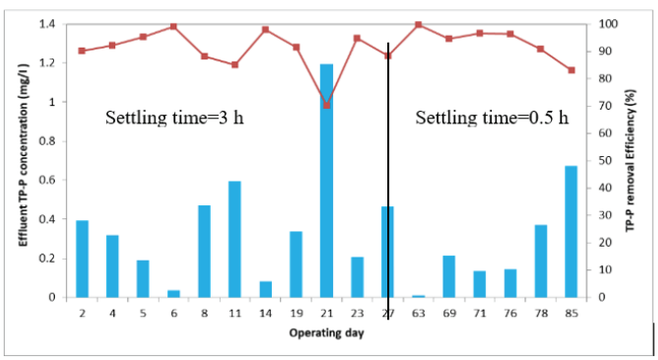


Fig. 3. Total phosphorus removal performance in the PBR with different settling time under low agitation speed.

3.4. COD removal

Concluding comprehensibly, AS treats organic matter in wastewater and produces CO_2 and inorganic nutrients required by microalgae, while microalgal photosynthesis generates O_2 back for the AS thus forming the synergistic relationship in co-culture systems [37]. The analysis of COD in effluent was done during the experiment (Fig. 4) with initial concentrations of 400 mg/l for a 2 day HRT. After 4 days of experiment, microbes in the co-culture removed 70% of COD, which showed that bacteria in activated sludge has a good ability to adapt to the new environment with microalgae. When the settling time was 3 h, there was a small fluctuation in COD removal efficiency of co-culture system. However, the COD removal efficiencies during that time were always above 70%, and gradually increased to over 90% when the settling time was reduced from 3 to 0.5 h. In another previous study, microalgae were demonstrated to play a dominant role in nitrogen removal via biological assimilation while activated sludge was responsible for improving COD removal [25]. In the first period in which AS was dominant, the COD removal efficiency was lower than when microalgae became dominant as bio-flocculant. These results imply that the COD removal by co-culture is accomplished by symbiotic interactions between activated sludge and microalgae, which help each other to enhance overall COD removal. The possible interactions are enhanced while the bio-flocculation was configured, and bacteria concentration remained low as well. At the end of experiment, while the bacteria took a small fraction in total biomass, the co-culture still had a good ability to remove COD up to 95.3% (day 85), which indicated the potential of applying co-culture in wastewater treatment using aPBR - an energy-saving system.

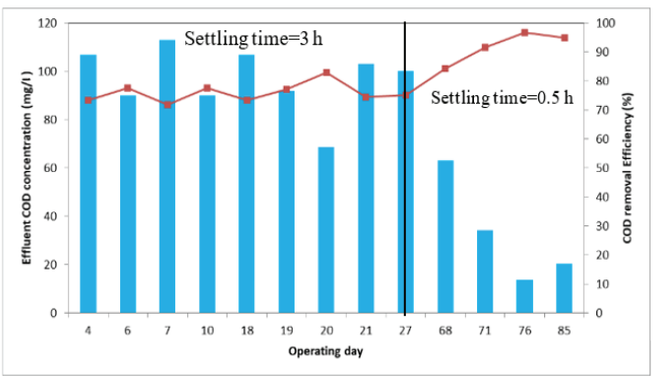


Fig. 4. COD removal performance in the PBR with different settling time.

4. Conclusions

Under the following typical conditions such as COD/N (10:1) in raw wastewater and a light:dark cycle of 14:10, the results obtained indicate the vital role of microalgae-AS co-culture in treating wastewater with a low agitation photobioreactor. The findings also suggest that the short settling time of 30 min could enhance the bio-flocculation of the microalgae-AS biomass. Besides, short settling times can be considered as an effective and easy strategy in co-culture operations to promote heavy floc formation. This would create an advantage in microalgae harvesting through bio-flocculation. However, due to the low nitrogen removal in this study, further research should be taken to overcome this current disadvantage.

CRediT author statements

Hong Hai Nguyen, Hong Ngoc Luong, Ngoc Kim Qui Nguyen: Data curation, Software, Visualisation; Le Phuong Uyen Nguyen, Phuong Thao Nguyen, Cong Sac Tran: Investigation, Writing - Original draft preparation; Xuan Thanh Bui: Conceptualisation, Methodology, Supervision.

ACKNOWLEDGEMENTS

This research is funded by Vietnam National University, Ho Chi Minh city (VNU-HCM) under grant number NCM2021-20-01. Hong Hai Nguyen was funded by Vingroup Joint Stock Company and supported by the Domestic Master/ PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Vingroup Big Data Institute (VINBIGDATA), code VINIF.2020.ThS.77.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] S. Longo, B.M. d'Antoni, M. Bongards, et al. (2016), "Monitoring and diagnosis of energy consumption in wastewater treatment plants. A state of the art and proposals for improvement", *Applied Energy*, **179**, pp.1251-1268, DOI: 10.1016/j.apenergy.2016.07.043.

[2] G. Quijano, J.S. Arcila, G. Buitrón (2017), "Microalgal-bacterial aggregates: Applications and perspectives for wastewater treatment", *Biotechnology Advances*, **35(6)**, pp.772-781, DOI: 10.1016/j.biotechadv.2017.07.003.

[3] A.R. Martinez, N.M. Garcia, I. Romero, et al. (2012), "Microalgae cultivation in wastewater: Nutrient removal from anaerobic membrane bioreactor effluent", *Bioresource Technology*, **126**, pp.247-253, DOI: 10.1016/j.biortech.2012.09.022.

[4] M.L. Menegazzo, G.G. Fonseca (2019), "Biomass recovery and lipid extraction processes for microalgae biofuels production: A review", *Renewable and Sustainable Energy Reviews*, **107**, pp.87-107, DOI: 10.1016/j.rser.2019.01.064.

[5] F.G.A. Fernández, C.G. Serrano, J.M.F. Sevilla (2018), "Recovery of nutrients from wastewaters using microalgae", *Frontiers in Sustainable Food Systems*, DOI: 10.3389/fsufs.2018.00059.

[6] R. Honda, Y. Teraoka, M. Noguchi, et al. (2017), "Optimisation of hydraulic retention time and biomass concentration in microalgae biomass production from treated sewage with a membrane photobioreactor", *Journal of Water and Environment Technology*, **15(1)**, pp.1-11, DOI:10.2965/jwet.15-085.

[7] T.T. Nguyen, X.T. Bui, H.H. Ngo, et al. (2021), "Nutrient recovery and microalgae biomass production from human urine source by membrane photobioreactor at low biomass retention time", *Science of The Total Environment*, **785**, DOI: 10.1016/j. scitotenv.2021.147423.

[8] R. Xing, W.W. Ma, Y.W. Shao, et al. (2018), "Growth and potential purification ability of Nitzschia sp. benthic diatoms in sea cucumber aquaculture wastewater", *Aquaculture Research*, **49(3)**, DOI: 10.1111/are.13722.

[9] D. Garcia, C. Alcantara, S. Blanco, et al. (2017), "Enhanced carbon, nitrogen and phosphorus removal from domestic wastewater in a novel anoxic-aerobic photobioreactor coupled with biogas upgrading", *Chemical Engineering Journal*, **313**, pp.424-434, DOI: 10.1016/j.cej.2016.12.054.

[10] Y. Chisti (2007), "Biodiesel from microalgae", *Biotechnology Advances*, **25(3)**, pp.294-306, DOI: 10.1016/j.biotechadv.2007.02.001.

[11] P.M. Schenk, S.R.T. Hall, E. Stephens, et al. (2008), "Second generation biofuels: High-efficiency microalgae for biodiesel production", *BioEnergy Research*, **1**, pp.20-43, DOI: 10.1007/s12155-008-9008-8.

[12] I. Rawat, R.R. Kumar, T. Mutanda, et al. (2011), "Dual role of microalgae: Phycoremediation of domestic wastewater and biomass production for sustainable biofuels production", *Applied Energy*, **88(10)**, pp.3411-3424, DOI: 10.1016/j.apenergy.2010.11.025.

[13] O. Tricolici, C. Bumbac, V. Patroescu, et al. (2014), "Dairy wastewater treatment using an activated sludge-microalgae system at different light intensities", *Water Science and Technology*, **69(8)**, pp.1598-1605, DOI: 10.2166/wst.2013.752.

[14] S. Zhu, L. Qin, P. Feng, et al. (2019), "Treatment of low C/N ratio wastewater and biomass production using co-culture of *Chlorella vulgaris* and activated sludge in a batch photobioreactor", *Bioresource Technology*, **274**, pp.313-320, DOI: 10.1016/j.biortech.2018.10.034.

[15] C.C. Tang, W. Zuo, Y. Tian, et al. (2016), "Effect of aeration rate on performance and stability of algal-bacterial symbiosis system to treat domestic wastewater in sequencing batch reactors", *Bioresource Technology*, **222**, pp.156-164, DOI: 10.1016/j.biortech.2016.09.123.

[16] T.L. Pham, M.H. Bui (2020), "Removal of nutrients from fertilizer plant wastewater using *Scenedesmus* sp.: Formation of bioflocculation and enhancement of removal efficiency", *Journal of Chemistry*, **2020(809272)**, pp.1-9, DOI: 10.1155/2020/8094272.

[17] C. Liu, Y. Hao, J. Jiang, et al. (2017), "Valorization of untreated rice bran towards bioflocculant using a lignocellulose-degrading strain and its use in microalgal biomass harvest", *Biotechnology for Biofuels*, **10(1)**, DOI: 10.1186/s13068-017-0780-6.

[18] C. Safi, B. Zebib, O. Merah, et al. (2014), "Morphology, composition, production, processing and applications of *Chlorella vulgaris*: A review", *Renewable and Sustainable Energy Reviews*, **35**, pp.265-278, DOI: 10.1016/j.rser.2014.04.007.

[19] H.C. Bold (1949), "The morphology of *Chlamydomonas chlamydogama*, Sp. Nov.", *Bulletin of The Torrey Botanical Club*, **76(2)**, pp.101-108, DOI: 10.2307/2482218.

[20] H.W. Bischoff, H.C. Bold (1963), *Some Soil Algae from Enchanted Rock and Related Algal Species*, University of Texas Publications, 95pp.

[21] American Water Works Association (2012), *Standard Methods for Examination of Water and Wastewater*, 1796pp.

[22] C.C. Tang, Y. Tian, H. Liang, et al. (2018), "Enhanced nitrogen and phosphorus removal from domestic wastewater via algae-assisted sequencing batch biofilm reactor", *Bioresource Technology*, **250**, pp.185-190, DOI: 10.1016/j.biortech.2017.11.028.

[23] R.E. Hecky, H.J. Kling (1981), "The phytoplankton and protozooplankton of the euphotic zone of Lake Tanganyika: Species composition, biomass, chlorophyll content, and spatio-temporal distribution", *Limnology and Oceanography*, **26(3)**, pp.548-564, DOI: 10.4319/lo.1981.26.3.0548.

[24] C.S. Lee, S.A. Lee, S.R. Ko, et al. (2015), "Effects of photoperiod on nutrient removal, biomass production, and algal-bacterial population dynamics in lab-scale photobioreactors treating municipal wastewater", *Water Research*, **68**, pp.680-691, DOI: 10.1016/j.watres.2014.10.029.

[25] T.T.D. Nguyen, T.T. Nguyen, Q.A. Binh, et al. (2020), "Co-culture of microalgae-activated sludge for wastewater treatment and biomass production: Exploring their role under different inoculation ratios", *Bioresource Technology*, **314**, DOI: 10.1016/j. biortech.2020.123754.

[26] A.L. Gonçalves, J.C.M. Pires, M. Simoes (2016), "Wastewater polishing by consortia of *Chlorella vulgaris* and activated sludge native bacteria", *Journal of Cleaner Production*, **133**, pp.348-357, DOI: 10.1016/j.jclepro.2016.05.109.

[27] I.D. Godos, J.L. Mendoza, F.G. Acien, et al. (2014), "Evaluation of carbon dioxide mass transfer in raceway reactors for microalgae culture using flue gases", *Bioresource Technology*, **153**, pp.307-314, DOI: 10.1016/j.biortech.2013.11.087.

[28] R. Ramanan, B.H. Kim, D.H. Cho, et al. (2016), "Algae-bacteria interactions: Evolution, ecology and emerging applications", *Biotechnology Advances*, **34(1)**, pp.14-29, DOI: 10.1016/j.biotechadv.2015.12.003.

[29] O. Tiron, C. Bumbac, E. Manea, et al. (2017), "Overcoming microalgae harvesting barrier by activated algae granules", *Scientific Reports*, **7(1)**, DOI: 10.1038/s41598-017-05027-3.

[30] X. Liu, S. Sun, B. Ma, et al. (2016), "Understanding of aerobic granulation enhanced by starvation in the perspective of quorum sensing", *Applied Microbiology and Biotechnology*, **100(8)**, pp.3747-3755, DOI: 10.1007/s00253-015-7246-1.

[31] A.K. Lee, D.M. Lewis, P.J. Ashman (2009), "Microbial flocculation, a potentially low-cost harvesting technique for marine microalgae for the production of biodiesel", *Journal of Applied Phycology*, **21**, pp.559-567, DOI: 10.1007/s10811-008-9391-8.

[32] M. Xu, P. Li, T. Tang, et al. (2015), "Roles of SRT and HRT of an algal membrane bioreactor system with a tanks-in-series configuration for secondary wastewater effluent polishing", *Ecological Engineering*, **85**, pp.257-264, DOI: 10.1016/j.ecoleng.2015.09.064.

[33] Z.C. Wang, M.C. Gao, Y. Ren, et al. (2015), "Effect of hydraulic retention time on performance of an anoxic-aerobic sequencing batch reactor treating saline wastewater", *International Journal of Environmental Science and Technology*, **12(6),** pp.2043-2054, DOI: 10.1007/s13762-014-0594-z.

[34] N.G.A.I. Karya, N.P.V.D. Steen, P.N.L. Lens (2013), "Photo-oxygenation to support nitrification in an algal-bacterial consortium treating artificial wastewater", *Bioresource Technology*, **134**, pp.244-250, DOI: 10.1016/j.biortech.2013.02.005.

[35] I.D. Godos, S. Blanco, P.A. García-Encina, et al. (2009), "Long-term operation of high rate algal ponds for the bioremediation of piggery wastewaters at high loading rates", *Bioresource Technology*, **100(19)**, pp.4332-4339, DOI: 10.1016/j.biortech.2009.04.016.

[36] Y. Su, A. Mennerich, B. Urban (2012), "Synergistic cooperation between wastewater-born algae and activated sludge for wastewater treatment: Influence of algae and sludge inoculation ratios", *Bioresource Technology*, **105**, pp.67-73, DOI: 10.1016/j.biortech.2011.11.113.

[37] G. Mujtaba, K. Lee (2017), "Treatment of real wastewater using co-culture of immobilized *Chlorella vulgaris* and suspended activated sludge", *Water Research*, **120**, pp.174-184, DOI: 10.1016/j.watres.2017.04.078.

Table of contents in 2022

Vietnam Journal of Science, Technology and Engineering (VJSTE): Manuscript Template

General Instructions on using this template and submitting a manuscript to VJSTE: Thank you for preparing a manuscript for submission to VJSTE. Using this template, or following the guidelines below, will help us in processing your article.

The structure of different types of articles may vary according to the type of study, and the content of the article.

The following are the detailed main components of the original research and review articles submitted to the journal.

Original research

The original research should have three following parts:

Preliminary section: It includes title, list of all contributing authors with their institutional affiliation, and abstract (160-250 words) with at least 3 keywords.

Title: How to format a science article (replace with your real title)

Author(s): List the author name(s) separated by commas. Use superscript numbers to link affiliations, and symbols *†‡ for author notes. For example, Y. Obsie[1*], S.A. Adem[1, 2]

**Corresponding author(s):* Include the email addresses or telephone numbers of the corresponding author(s). Please use the asterisk (*) symbol for the corresponding author information. This information is at the bottom of the manuscript's first page.

Author's affiliation addresses:

[1]Affiliations should be preceded by superscript numbers corresponding to the author list

[2]For any institutions, please provide the institution name, street, district, city, country

Received ; revised ; accepted

Abstract:
Start your abstract here…
Keywords: **keyword 1, keyword 2, keyword 3…**
Classification number(s): **1.1, 3.4, 5.3, etc.**

Abstract: The abstract should range between 160 and 250 words and be structured in a single paragraph following keywords in a separate line. The abstract should represent the main part of the research manuscript. It should reflect a brief and comprehensive background following the aim of the study (main questions to be addressed), methods used for investigating the problem, a summary of significant findings, main conclusions, and implications of the results.

Keywords: They should come immediately after the abstract and contain 3-7 specific and relevant keywords. They should reflect the research study and the relevant field of research.

Classification numbers: Depend on the Journal's Classification Scheme.

Author(s) are encouraged to write a "Highlights" that summarises the most exciting, interesting points from your manuscript.

Manuscript section: It is the main body of the scientific manuscript including: 1) Introduction; 2) Material(s) and method(s); 3) Results and discussion (separation of these two sections is acceptable); 4) Conclusions; 5) Recommendations/Suggestions/ Implications/Study limitations...(if any). These sections may slightly vary depending on the type of research approach and design used in the research. Each section should be numbered.

All of the Figures, Tables, and Equations should be cited in numerical order.

Figures

Fig. #: (Begin each figure caption with a label, "**Fig. 1.**" for example, as a new paragraph).

Fig. 1. The figure caption should begin with an overall descriptive statement of the figure followed by additional text. They should be immediately after each figure. Figure parts are indicated with capital letters **(A)**. If you prefer, you can place both the actual figures and captions logically through the text near where they are cited rather than at the end of the file (but not both). If a paragraph in the main text begins with the name of a figure, write out "Figure" in full (e.g., <para> "Figure 1 shows...").

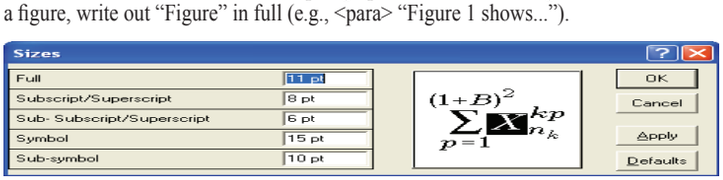


Fig. 1. The figure caption.

Tables

Table #: (Begin each table caption with a label "**Table 1.**", etc. as a new paragraph).

Table 1. Start this caption with a short description of your table. Format tables using the Word Table commands and structures. Do not create tables using spaces or tabs characters.

Table 1. Caption.

a	b	c	d	e
f				
g				
h				

Equations

$$\int_a^b x dx \tag{1}$$

Equations can be included. We do not recommend using the native Word 2007, 2008, 2010 or 2011 equation editor. This can in some cases produce less reliable MathML, the online markup language we use, which may result in display errors. Instead, use the legacy equation editor in Word (Insert menu; select insert object; select word equation) or use MathType (recommended). If you enter equations in simple LaTeX, check that they will convert accurately (Word 2007 and higher can convert simple LaTeX equations).

Closing section: It includes appendices/supplementary materials/data availability statement (if applicable), CRediT author statement (for multiple authors), acknowledgements (if applicable), competing interests, ethics (if any) and a complete list of references corresponding to all the citations given in the text of the manuscript. The appendices/supplementary materials may include data files and other material relevant to the research that should be submitted in separate files if required.

APPENDICES (if any)

The appendices should ideally be under 1000 words (approximately 2 A4 pages) for printing version. With longer appendices, they are only accepted for online publication.

CRediT author statement (for multiple authors)

CRediT offers authors the opportunity to share an accurate and detailed description of their diverse contributions to the published work.

ACKNOWLEDGEMENTS (if applicable)

This section is for the researchers who were funded for the research projects, and it should enclose all sources of financial support.

COMPETING INTERESTS

The author(s) declare(s) that there is no conflict of interest regarding the publication of this article.

REFERENCES

For in-text citation: References should be cited in the bracket with a number [1]. Multiple reference citations are separated by commas [2, 3] or if a series, dashes [4-6]. When citing sources with the author(s) and the year, they should be shortened to the first author's name, followed by et al. (with more than one author) and the year.

For reference list: With less than or equal to three authors, list all the author(s). With more than three authors, the first three should be listed and followed by et al.

References are listed in numerical order, and in the same order in which they are cited in text. The reference list should include all and only those references you have cited in the text. The reference list appears at the end of the article.

The reference consists of: Initial(s) and first/last name of the author(s), author' names separated by commas (Year of publication), "Titles of journal articles or chapter of book", *Title of Journal, Book or Other Publication*, **Volume(Number)** (if any), page numbers (if any), DOI (if any), URL and accessed date (if any) (in its original language if not English).

To cite a website and online/electronic resources, please include the accessed date (day month year).

Example of a reference list:

[1] T. Binzoni, T.S. Leung, D. Boggett, et al. (2003), "Non-invasive laser Doppler perfusion measurements of large tissue volumes and human skeletal muscle blood RMS velocity", *Phys. Med. Biol.*, **48(1)**, pp.2527-2549, DOI: 10.1088/0031-9155-48-15-318.

[2] H.W. Bischoff, H.C. Bold (1963), *Some Soil Algae from Enchanted Rock and Related Algal Species*, University of Texas Publications, 95pp.

[3] R. Busi (2016), "Weedy rice in the Philippines and Vietnam", https://www.aciar.gov.au/publication/technical-publications/weedy-rice-philippines-and-vietnam-final-report, accessed 15 July 2018.

Review articles

They should be different from the original research as they would contain different sections on the literature review. Authors may structure this part as per nature, background, and available literature on the issue argued in the manuscript.

For more detailed instructions, please visit our website: vietnamscience.vn or vietnamscience.vjst.vn

This manuscript template has been used from 2021 onwards.

P-ISSN 2525-2461 • E-ISSN 2615-9937

Vietnam Journal of Science, Technology and Engineering

DECEMBER 2022 • VOLUME 64 NUMBER 4

EDITORIAL OFFICE

113 Tran Duy Hung Str., Cau Giay Dist., Hanoi, Vietnam

Tel: (84.24) 39436793; Fax: (84.24) 39436794

Email: vjste@most.gov.vn

Website: vietnamscience.vn or vietnamscience.vjst.vn

PUBLICATION LICENCE

No. 459/GP-BTTTT 20th July 2021