

PHÁT TRIỂN CÁC PHỨC CHẤT MỚI GIỮA THIOSEMICARBAZONE VÀ CÁC ION KIM LOẠI SỬ DỤNG PHƯƠNG PHÁP MÔ HÌNH HÓA QSPR

Nguyễn Hoàng Minh¹, Phạm Văn Tất², Nguyễn Minh Quang^{1*}

¹Khoa công nghệ hóa học, Trường Đại học Công nghiệp TpHCM

²Khoa khoa học sức khỏe, Đại học Hoa Sen

*Email: nguyenminhquang@iuh.edu.vn

Ngày nhận bài: 23/3/2022; ngày hoàn thành phản biện: 14/3/2023; ngày duyệt đăng: 4/4/2023

TÓM TẮT

Hàng số bền của 20 phức chất mới (ML₂) giữa thiosemicarbazone với các ion kim loại được tính toán từ mô hình quan hệ định lượng cấu trúc-tính chất (QSPR). Các mô hình QSPR này được xây dựng bằng kỹ thuật hồi quy tuyến tính đa biến (QSPR_{MLR}) và mạng thần kinh nhân tạo (QSPR_{ANN}). Các biến của mô hình nhận được qua phép tính lượng tử bán thực nghiệm PM7 và PM7/sparkle. Kết quả 10 mô hình QSPR_{MLR} xây dựng được có cùng bốn biến, tuy nhiên qua kỹ thuật đánh giá ngoại với một tập dữ liệu độc lập thì mô hình thứ tư (QSPR_{MLR4}) nhận được kết quả với sự đáp ứng tốt nhất những tiêu chuẩn thống kê và các giá trị này là $R^2_{\text{train}} = 0,885$; $Q^2_{\text{LOO}} = 0,860$; $SE = 1,527$. Hơn nữa, mô hình mạng QSPR_{ANN} I(4)-HL(10)-O(1) được tìm thấy khi phát triển từ các mô tả của mô hình QSPR_{MLR4} và cũng cho kết quả tốt như $R^2_{\text{train}} = 0,961$; $Q^2_{\text{test}} = 0,969$; $Q^2_{\text{validation}} = 0,978$. Các kết quả từ nghiên cứu này có thể sử dụng để phát triển các dẫn xuất mới khác nhằm ứng dụng trong lĩnh vực hóa học, phân tích và môi trường.

Từ khóa: Hàng số bền $\log\beta_{12}$, hồi quy đa biến, mạng thần kinh nhân tạo, QSPR, thiosemicarbazone.

1. MỞ ĐẦU

Thiosemicarbazone được biết đến như một dẫn xuất phổ biến trong hóa học hữu cơ và chúng được sử dụng rộng rãi trong các lĩnh vực liên quan [1,2]. Đây là lý do mà chúng được các nhà khoa học quan tâm và được nghiên cứu cũng như tổng hợp rất nhiều trong thực nghiệm [1]. Thiosemicarbazone cũng rất dễ dàng tạo phức chất với các ion kim loại do đặc trưng liên kết chức các phối tử nito và lưu huỳnh trong cấu trúc [1]. Phức chất của chúng cũng được tìm thấy nhiều ứng dụng liên quan trong nhiều công trình nghiên cứu, đặc biệt là trong lĩnh vực hóa học phân tích [2], trong đó hàng

số bền là đại lượng đặc trưng cho khả năng tạo phức này của các ligand tương ứng. Do đó, việc tìm kiếm các dẫn xuất mới của thiosemicarbazone với các ưu điểm vượt trội và đa ứng dụng luôn là vấn đề các nhà hóa học quan tâm.

Trong nghiên cứu này, chúng tôi xây dựng một lộ trình đầy đủ từ sự kết hợp của hóa học lượng tử với phương pháp bán thực nghiệm mới (PM7 và PM7/sparkle) [3], các công cụ tin học/toán học thống kê để xây dựng các mô hình trên đối tượng phức chất của thiosemicarbazone và các ion kim loại nhằm tìm kiếm các mô hình tốt nhất từ đó phát triển mới các hợp chất liên quan. Việc xây dựng các mô hình này tuân thủ chặt chẽ các nguyên tắc và sự hướng dẫn của tổ chức OECD [4] và các chỉ số của Tropsha [5].

Nghiên cứu này xây dựng thành công hai mô hình QSPR với phương pháp hồi quy tuyến tính đa biến (QSPR_{MLR}) trên tập dữ liệu bao gồm 66 giá trị thực nghiệm của phức chất của thiosemicarbazone với các ion kim loại Ni²⁺, Co²⁺, Mn²⁺, Zn²⁺, Hg²⁺, Cu²⁺, Cd²⁺ and Mg²⁺ trong dung dịch nước (Bảng 1). Tiếp đó, chúng tôi cũng đã phát triển mô hình mạng thần kinh nhân tạo (QSPR_{ANN}) từ các mô tả của mô hình QSPR_{MLR} nhằm tìm kiếm các mô hình dự báo tốt hơn. Chúng tôi đã sử dụng các mô hình này để phát triển tính toán giá trị hằng số bền của 20 phức chất trên một số dẫn xuất thiosemicarbazone thiết kế mới bằng cách thêm các nhóm chức đã được tổng hợp gắn vào các vị trí tìm năng của khung cấu trúc gốc (Bảng 6).

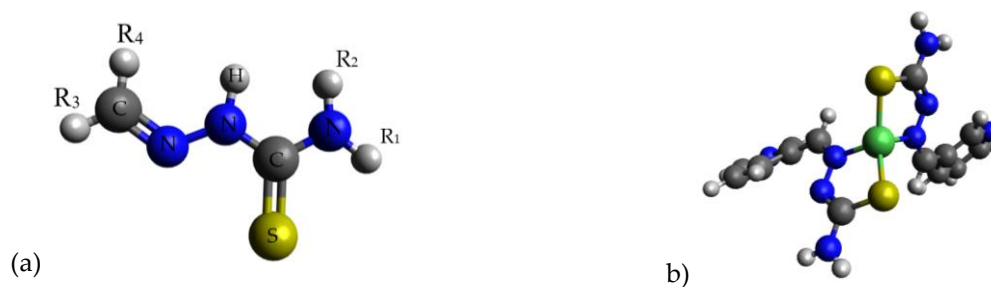
2. PHƯƠNG PHÁP TÍNH TOÁN

Nghiên cứu này phải được bắt đầu bằng một tập dữ liệu thực nghiệm, do đó, việc khai thác dữ liệu là bước đầu tiên trong nghiên cứu này [4].

2.1. Thu thập dữ liệu nghiên cứu

Đối tượng sử dụng trong nghiên cứu là phức chất dạng đơn nhân hai càng (ML₂) của thiosemicarbazone (L) và các ion kim loại (M). Như vậy, cấu trúc của ligand và phức chất này được mô tả như Hình 1. Do đó, hằng số bền (β_{12}) được tính theo công thức sau:

$$\beta_{12} = \frac{[ML_2]}{[M][L]^2} \quad (1)$$



Hình 1. (a) Cấu trúc tổng quát của thiosemicarbazone; (b) Phức chất giữa Ni^{2+} với 3-formylpyridine thiosemicarbazone [6]

Kỹ thuật khai thác dữ liệu được thực hiện như sau: đầu tiên, một tập dữ liệu lớn được thu thập từ các công trình nghiên cứu thực nghiệm, sau đó sử dụng cách phân loại cụm (clustering) với phương pháp phân cụm theo phân cấp tập hợp AHC (Agglomerative Hierarchical Clustering) để phân chia các tập dữ liệu nhỏ hơn [16]. Kết quả một tập dữ liệu bao gồm 66 giá trị hằng số bền của phức đã được sử dụng để xây dựng các mô hình QSPR trong nghiên cứu này được trình bày trong bảng 1.

Bảng 1. Giá trị lớn nhất ($\log\beta_{12,max}$) và nhỏ nhất ($\log\beta_{12,min}$) hằng số bền thực nghiệm của các phức chất trong nghiên cứu

TT	Thiosemicarbazone				Ion kim loại	Số lượng phức, n	$\log\beta_{12,min}$	$\log\beta_{12,max}$	Tài liệu
	R ₁	R ₂	R ₃	R ₄					
1	H	H	H	- C ₅ H ₄ N	Ni ²⁺	1	8,980	8,980	[6]
2	H	H	H	- C ₅ H ₄ N	Mn ²⁺	1	7,850	7,850	[6]
3	H	H	H	- C ₅ H ₄ N	Co ²⁺	1	8,580	8,580	[7]
4	H	H	H	- C ₅ H ₄ N	Zn ²⁺	1	8,890	8,890	[7]
5	H	H	H	- C ₆ H ₄ OH	Cu ²⁺	1	10,773	10,773	[8]
6	H	H	H	-C ₄ H ₃ O	Cd ²⁺	1	5,510	5,510	[9]
7	H	H	H	- C ₉ H ₈ N	Hg ²⁺	16	10,2493	10,992	[10]
8	H	H	H	-C ₁₀ H ₆ OH	Mg ²⁺	4	5,5800	6,700	[11]
9	H	H	H	-C ₁₀ H ₆ OH	Cd ²⁺	4	9,9200	11,420	[11]
10	H	H	H	-C ₁₀ H ₆ OH	Zn ²⁺	3	12,700	14,370	[11]
11	H	H	-	- C ₉ H ₇ NO	Zn ²⁺	7	12,591	15,056	[12]
12	H	H	-	- C ₉ H ₇ NO	Cd ²⁺	7	11,384	13,837	[12]
13	H	H	H	-C ₆ H ₄ NH ₂	Cu ²⁺	2	19,490	20,750	[13]
14	H	H	H	-C ₆ H ₄ NH ₂	Ni ²⁺	2	19,750	23,700	[13]
15	H	H	H	-C ₆ H ₄ NH ₂	Co ²⁺	2	16,970	22,290	[13]

TT	Thiosemicarbazone				Ion kim loại	Số lượng phức, n	$\log\beta_{12,\min}$	$\log\beta_{12,\max}$	Tài liệu
	R ₁	R ₂	R ₃	R ₄					
16	H	H	H	-C ₆ H ₄ NH ₂	Mn ²⁺	2	17,130	20,700	[13]
17	H	H	H	-C ₆ H ₄ NH ₂	Zn ²⁺	2	15,230	20,780	[13]
18	H	H	H	-C ₆ H ₄ NO ₂	Pr ³⁺	2	19,170	20,240	[14]
19	H	H	H	-C ₆ H ₄ NO ₂	Cd ²⁺	2	17,550	18,320	[15]
20	H	H	H	-C ₆ H ₄ NO ₂	Pb ²⁺	2	16,580	17,670	[15]
21	H	H	H	-C ₆ H ₄ NO ₂	Cr ³⁺	2	16,000	21,160	[15]
22	H	H	H	-C ₆ H ₄ NO ₂	Fe ³⁺	1	18,490	18,490	[15]

2.2. Tính toán bộ mô tả

Bộ mô tả của mô hình được sử dụng trong nghiên cứu mô hình hóa QSPR được hiểu là các biến số của phương trình mô hình xây dựng được, nó bao gồm mô tả lượng tử và mô tả phân tử (0D-3D). Mô tả lượng tử được tính toán từ quá trình tối ưu hóa cấu trúc phức chất, còn các mô tả phân tử đặc trưng cho cấu trúc của các phân tử trong nghiên cứu QSPR và được tính toán từ công cụ QSARIS 1.1 [17] từ những cấu trúc đã phức chất đã được tối ưu hóa.

Sau khi tính toán, quá trình loại bỏ các biến không phù hợp để tạo ra kết quả sẽ nhận được một bộ cơ sở dữ liệu bao gồm các phức chất thực nghiệm tương ứng với các giá trị hằng số bền ($\log\beta_{12,\text{exp}}$) và các biến là các tham số cấu trúc được tính toán. Cuối cùng, chúng tôi sử dụng cơ sở dữ liệu này để phát triển các mô hình hồi quy tuyến tính đa biến và mạng thần kinh nhân tạo.

2.3. Xây dựng mô hình hồi quy tuyến tính đa biến

Nghiên cứu sử dụng phương pháp hồi quy tuyến tính đa biến để xây dựng mô hình QSPR. Đây là một trong những phương pháp thống kê được sử dụng kinh điển trong nhiều công trình nghiên cứu. Phương pháp này xác mức độ quan hệ tuyến tính giữa biến độc lập (giá trị dự báo) và nhiều hơn một biến phụ thuộc. Trong trường hợp này, giá trị hằng số bền của phức giữa thiosemicarbazone và các ion kim loại ($\log\beta_{12}$) là biến độc lập và bộ mô tả là những biến phụ thuộc. Mô hình của phương pháp này được biểu diễn theo phương trình sau: [16,18]

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_k \cdot X_k + \varepsilon \quad (2)$$

Trong đó, Y là biến phụ thuộc; $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ là các tham số hồi quy của mô hình, X_i tương ứng với biến giải thích thứ i (với $i = 1$ đến k) và ε là sai số ngẫu nhiên [16,18].

2.4. Xây dựng mô hình mạng thần kinh nhân tạo

ANN dựa trên một tập hợp các đơn vị hoặc nút được kết nối được gọi là tế bào thần kinh nhân tạo, mô hình hóa lỏng lẻo các tế bào thần kinh trong não sinh học. Trong các triển khai ANN thông thường, tín hiệu tại kết nối giữa các thần kinh nhân tạo là số thực và đầu ra của mỗi thần kinh nhân tạo được tính bằng một số hàm phi tuyến tính của tổng các đầu vào của nó. Thông thường, các tế bào thần kinh nhân tạo được tập hợp lại thành từng lớp. Các lớp khác nhau có thể thực hiện các loại biến đổi khác nhau trên đầu vào của chúng [19]. Mỗi quan hệ này được thể hiện bằng hàm truyền, hai hàm truyền được sử dụng trong nghiên cứu này như sau [19]:

$$g(x) = \log \text{sig}(x) = \frac{1}{1+e^{-x}} \quad (3)$$

$$g(x) = \tan \text{sig}(x) = \frac{1-e^{-x}}{1+e^{-x}} \quad (4)$$

Bên cạnh đó, nghiên cứu này sử dụng mạng truyền thẳng nhiều lớp MLP (Multi-layer Perceptron) và giải thuật lan truyền ngược [20] - giải thuật được Rumelhart và cộng sự đề xuất năm 1986 [20]. Kiểu mạng MLP này có cấu trúc bao gồm một lớp vào, một lớp ra và một hoặc nhiều lớp ẩn. Trong khi đó, giải thuật lan truyền ngược là dạng tổng quát của giải thuật trung bình bình phương tối thiểu (LMS). Giải thuật này thuộc dạng giải thuật xấp xỉ để tìm các điểm mà tại đó hiệu năng của mạng là tối ưu. Giải thuật này được thực hiện các bước: lan truyền thẳng các tính toán trong mạng, lan truyền ngược; và cập nhật lại các trọng số và độ lệch tương ứng. Giải thuật dừng lại khi giá trị của hàm mục tiêu trở nên đủ nhỏ [21].

2.5. Kỹ thuật đánh giá mô hình

Đánh giá mô hình là bước quan trọng khẳng định tính đúng đắn của mô hình QSPR xây dựng được. Thông thường, việc đánh giá mô hình bao gồm đánh giá nội và đánh giá ngoại trên hai tập dữ liệu độc lập [5]. Trong nghiên cứu này, việc đánh giá nội được thực hiện trên tập dữ liệu luyện mô hình ban đầu với 66 giá trị thực nghiệm (Bảng 1) với chỉ số đánh giá R^2 ($> 0,6$) kết hợp với việc đánh giá chéo CV (Cross Validation) theo kỹ thuật loại bỏ dần từng trường hợp LOO (Leave One Out) với chỉ số thống kê đánh giá Q^2_{CV-LOO} ($> 0,6$) và việc đánh giá ngoại EV (External Validation) trên tập dữ liệu độc lập với tập luyện với chỉ số đánh giá là Q^2_{EV} ($> 0,5$) [5,18]. Các đại lượng này đều được tính trên cùng một công thức với các tập dữ liệu khác nhau như sau [5,18]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (5)$$

Trong đó, Y_i , \hat{Y}_i , and \bar{Y} là giá trị thực nghiệm, dự đoán và trung bình.

Sai số chuẩn (SE) đóng vai trò quan trọng trong việc nhận định dữ liệu, thông qua đó, nó sẽ đánh giá được sự biến động giá trị là tốt hay xấu. Về nguyên tắc, SE tỷ lệ nghịch với kích thước mẫu nên giá trị này sẽ giảm khi số quan sát mẫu càng lớn và giá trị càng nhỏ thì thống kê sẽ gần với thực tế. SE xác định bởi công thức sau [18]:

$$SE = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N - k - 1}} \quad (6)$$

Trong đó N và k lần lượt là số biến của tập luyện và số biến của mô hình.

Nghiên cứu này cũng sử dụng giá trị tuyệt đối trung bình của sai số tương đối MARE (%), để so sánh chất lượng của mô hình. Chúng được trình bày như sau: [22]

$$MARE, \% = \frac{1}{n} \frac{|\log \beta_{12,exp} - \log \beta_{12,pred}|}{\log \beta_{12,exp}} \cdot 100 \quad (7)$$

Với n là số lượng mẫu; $\beta_{12,exp}$ và $\beta_{12,pred}$ lần lượt là các hằng số bền thực nghiệm và tính toán.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Mô hình QSPR_{MLR}

Việc xây dựng mô hình QSPR_{MLR} được thực hiện bằng kỹ thuật loại bỏ ngược và hồi quy thuận trên hệ thống Regress [18]. Khả năng dự đoán của các mô hình QSPR_{MLR} được đánh giá chéo bằng phương pháp LOO sử dụng thống kê Q^2_{LOO} . Các mô hình hồi quy đa biến được xây dựng dựa trên tập dữ liệu thu thập ban đầu và quá trình xây dựng chia thành tập luyện (80%) và tập kiểm tra (20%). Chất lượng của các mô hình được đánh giá bằng các giá trị thống kê R^2_{train} , R^2_{adj} , Q^2_{LOO} , SE và F_{stat} . Kết quả nhận được mười mô hình QSPR_{MLR} đáp ứng các yêu cầu thông kê được trình bày trong Bảng 2.

Bảng 2. Mười mô hình QSPR_{MLR} (k = 4) với các giá trị thống kê tương ứng

Ký hiệu	Mô hình QSPR _{MLR} và các giá trị thống kê
MLR1	$\log \beta_{12} = -12,312 + 9,850 \times xc3 - 2,207 \times nrings + 0,629 \times SsNH2 - 2,529 \times LUMO$ $R^2 = 0,901; R^2_{adj} = 0,894; Q^2_{LOO} = 0,875; SE = 1,419; PRESS = 157,45; F_{stat} = 140,35$
MLR2	$\log \beta_{12} = -5,734 + 14,373 \times xc3 + 0,418 \times SsNH2 - 0,408 \times xvc3 - 1,882 \times xp4$ $R^2 = 0,857; R^2_{adj} = 0,847; Q^2_{LOO} = 0,822; SE = 1,703; PRESS = 223,41; F_{stat} = 92,61$
MLR3	$\log \beta_{12} = 3,209 + 2,568 \times HsNH2 + 0,189 \times SdO - 0,419 \times dipole - 0,735 \times LogP$ $R^2 = 0,886; R^2_{adj} = 0,879; Q^2_{LOO} = 0,859; SE = 1,518; PRESS = 177,32; F_{stat} = 120,70$

MLR4	$\log\beta_{12} = -18,563 + 0,909 \times SsNH2 + 0,200 \times SdO - 0,003 \times Total\ energy + 2,702 \times SssCH2$ $R^2 = 0,885; R^2_{adj} = 0,877; Q^2_{LOO} = 0,860; SE = 1,527; PRESS = 175,46; F_{stat} = 119,02$
MLR5	$\log\beta_{12} = -38,766 + 0,386 \times SsNH2 - 2,906 \times k0 + 4,501 \times nvx + 1,389 \times xp5$ $R^2 = 0,845; R^2_{adj} = 0,835; Q^2_{LOO} = 0,836; SE = 1,769; PRESS = 237,40; F_{stat} = 84,72$
MLR6	$\log\beta_{12} = -5,328 + 2,953 \times HsNH2 + 0,162 \times SdO - 1,013 \times dipole + 0,011 \times MW$ $R^2 = 0,864; R^2_{adj} = 0,855; Q^2_{LOO} = 0,836; SE = 1,659; PRESS = 205,88; F_{stat} = 98,44$
MLR7	$\log\beta_{12} = -12,857 + 2,924 \times HsNH2 + 0,215 \times SdO + 2,939 \times SssCH2 + 0,022 \times Cosmo\ area$ $R^2 = 0,854; R^2_{adj} = 0,845; Q^2_{LOO} = 0,822; SE = 1,719; PRESS = 223,19; F_{stat} = 90,66$
MLR8	$\log\beta_{12} = -12,140 + 3,264 \times HsNH2 + 0,226 \times SdO + 2,466 \times SssCH2 + 0,014 \times Cosmo\ volume$ $R^2 = 0,855; R^2_{adj} = 0,846; Q^2_{LOO} = 0,829; SE = 1,711; PRESS = 215,02; F_{stat} = 90,70$
MLR9	$\log\beta_{12} = 8,827 + 0,731 \times SsNH2 - 4,597 \times nelem + 0,404 \times SHBa - 1,133 \times SdsCH$ $R^2 = 0,879; R^2_{adj} = 0,872; Q^2_{LOO} = 0,854; SE = 1,562; PRESS = 182,72; F_{stat} = 113,04$
MLR10	$\log\beta_{12} = 3,785 + 1,437 \times HsNH2 - 1,133 \times LogP + 38,644 \times Maxneg + 1,808 \times numBHa$ $R^2 = 0,828; R^2_{adj} = 0,816; Q^2_{LOO} = 0,794; SE = 1,868; PRESS = 258,73; F_{stat} = 74,41$

Tất cả mười mô hình trong Bảng 2 về cơ bản đều thỏa mãn yêu cầu thống kê với các giá trị R^2 và Q^2_{LOO} đều lớn hơn 0,6 [4,5], tuy nhiên để chọn mô hình hoàn chỉnh cần đánh giá ngoại bằng một tập dữ liệu độc lập. Nghiên cứu sử dụng tập dữ liệu bao gồm ba mươi phức chất thực nghiệm kèm giá trị hằng số bền ($\log\beta_{1,2exp}$) được trình bày trong Bảng 4. Dựa trên kết quả dự báo hằng số bền ($\log\beta_{1,2pred}$) ở Bảng 5 và Hình 2b thông qua hệ số $Q^2_{EV-MLR4}$ giữa giá trị thực nghiệm và các giá trị dự báo giữa các mô hình $QSPR_{MLR4}$ và giá trị MARE (%) thì mô hình $QSPR_{MLR4}$ được chọn để phát triển mô hình ANN và các phức chất mới bởi đáp ứng được giá trị $Q^2_{EV-MLR4} = 0,8734$ và MARE nhận giá trị thấp nhất (14,7948 %).

3.2. Mô hình $QSPR_{ANN}$

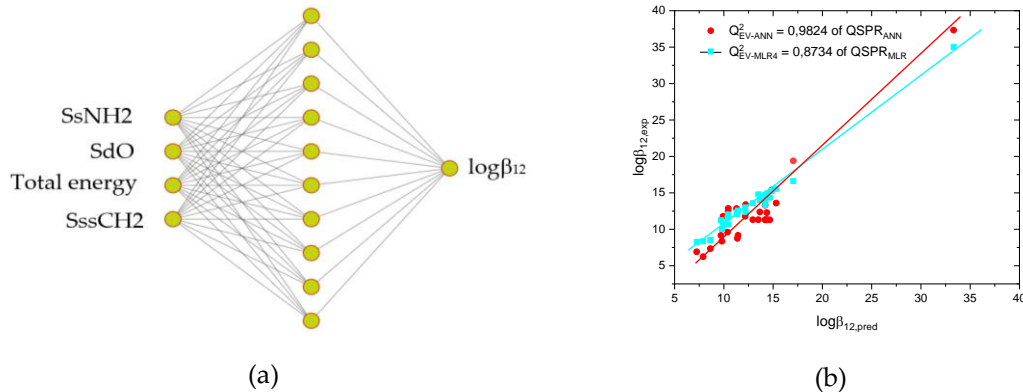
Như đã trình bày, mô hình $QSPR_{ANN}$ được phát triển dựa trên bốn biến mô tả của mô hình $QSPR_{MLR4}$ là $SsNH2$, SdO , $Total\ energy$ và $SssCH2$. Kỹ thuật phát triển lựa chọn mạng MLP và giải thuật lan truyền ngược. Do đó, kiến trúc mạng được lựa chọn trong trường hợp này là I(4)-HL(m)-O(1) với đầu vào là bốn biến mô tả, đầu ra là giá trị $\log\beta_{12,pred}$ và số nút lớp ẩn là m . Quá trình xây dựng mô hình ANN thực hiện hai bước. Đầu tiên, sử dụng tập luyện ban đầu để khảo sát các mạng MLP với các kiến trúc khác nhau (tìm lớp ẩn m). Kết quả các mạng thỏa mãn tính chất thống kê được trình bày ở Bảng 3. Tiếp đến, sử dụng tập dữ liệu đánh giá ngoại ở Bảng 4 để tìm kiếm mạng tốt nhất bằng cách dựa vào giá trị MARE (%) và chỉ số Q^2_{EV-ANN} giữa giá trị dự báo và giá trị thực nghiệm. Kết quả nhận được mạng với kiến trúc I(4)-HL(10)-O(1) với các

tham số thống kê như in đậm trong Bảng 3 và giá trị $Q^2_{EV-ANN} = 0,9824$ (Hình 2b) và giá trị MARE = 5,3034 % (Bảng 5).

Bảng 3. Kết quả khảo sát ban đầu mô hình QSPR_{ANN} với kiến trúc I(4)-HL(m)-O(1)

QSPR _{ANN}	R^2_{train}	Q^2_{test}	Q^2_{CV}	Sai số luyện	Sai số kiểm tra	Sai số đánh giá	Hàm truyền
I(4)-HL(9)-O(1)	0,9092	0,9798	0,9504	1,3509	0,9929	2,0272	hyperbolic tangent
I(4)-HL(8)-O(1)	0,9103	0,9807	0,9527	1,6227	0,6538	1,4815	hyperbolic tangent
I(4)-HL(3)-O(1)	0,9334	0,9905	0,9478	0,9574	0,8441	2,1581	hyperbolic tangent
I(4)-HL(9)-O(1)	0,9140	0,9827	0,9522	1,3566	0,6929	2,0216	log-sigmoid
I(4)-HL(10)-O(1)	0,9137	0,9808	0,9594	1,3977	0,7700	1,3120	hyperbolic tangent

Sử dụng phân tích ANOVA một yếu tố đánh giá sự khác biệt giá trị dự đoán và giá trị thực nghiệm của hai mô hình này. Kết quả cho thấy không có sự khác nhau về kết quả của hai mô hình ($F = 1,7527 < F_{0,05} = 2,6828$).



Hình 2. (a) Kiến trúc của mạng ANN I(4)-HL(10)-O(1); (b) Mối tương quan giữa giá trị thực nghiệm và giá trị dự đoán của tập dữ liệu ngoại của các mô hình QSPR

3.3. Đánh giá ngoại các mô hình QSPR

Một mô hình hoàn chỉnh cần phải được đánh giá ngoại bằng một tập dữ liệu thực nghiệm độc lập với tập dữ liệu xây dựng mô hình. Như đã đề cập, tập dữ liệu xây dựng mô hình được trình bày trong Bảng 1, còn tập dữ liệu đánh giá ngoại được trình bày trong Bảng 4 bao gồm 30 giá trị hằng số bền ($\log\beta_{12,exp}$) của các phức chất thực nghiệm. Quá trình đánh giá dựa vào hai đại lượng Q^2_{EV} và MARE (%). Kết quả dự báo từ các mô hình MLR và ANN được trình bày trong Bảng 5 và Hình 2b.

Theo quan sát kết quả từ Bảng 5, các giá trị dự báo hằng số bền ($\log\beta_{12,pred}$) của các mô hình MLR1, MLR2 và MLR5 rất kém, nhận các giá trị âm, điều này không phù hợp. Hơn nữa, dựa vào giá trị Q^2_{EV} cho thấy các mô hình MLR3, MLR6, MLR9 và MLR10

không thỏa mãn yêu cầu thống kê (đều nhỏ hơn 0,5) nên các mô hình này cũng không được chọn. Như vậy, ba mô hình còn lại bao gồm MLR4, MLR7 và MLR8 đều nhận giá trị Q^2_{EV} lớn hơn 0,5. Tuy nhiên, kết hợp với giá trị MARE (%) thì mô hình $QSPR_{MLR4}$ nhận kết quả nhỏ nhất (14,7948 %) nên được chọn. Trong khi đó, mô hình $QSPR_{ANN}$ I(4)-HL(10)-O(1) nhận các giá trị lần lượt là 5,3034 % và 0,9824 đều thỏa mãn tốt yêu cầu thống kê và kết quả cũng chỉ ra rằng mô hình ANN có khả năng dự đoán tốt hơn so với mô hình MLR4. Hơn nữa, các giá trị $\log\beta_{12,pred}$ dự đoán của mô hình ANN gần đúng với giá trị thực nghiệm $\log\beta_{12,exp}$ (Bảng 4 và Bảng 5).

Bảng 4. Ba mươi giá trị hằng số bền thực nghiệm ($\log\beta_{12,exp}$) của phức chất giữa ion kim loại và thiosemicarbazone trong tập dữ liệu đánh giá ngoại

Ký hiệu	Thiosemicarbazone				Ion kim loại	$\log\beta_{12,exp}$	Tài liệu
	R ₁	R ₂	R ₃	R ₄			
com1	H	H	H	- C ₆ H ₄ OH	Ni ²⁺	11,441	[8]
com2	H	H	H	-C ₁₃ H ₁₆ NO ₃	Fe ³⁺	33,320	[23]
com3	H	H	-CH(CH ₃) ₂	-C ₆ H ₂ Br(OH) ₂	Co ²⁺	10,4362	[24]
com4	H	H	- C ₆ H ₅	- CCH ₃ =N-OH	Ni ²⁺	9,7118	[25]
com5	H	H	H	- C ₉ H ₈ N	Ag ⁺	10,4448	[10]
com6	H	H	H	- C ₆ H ₄ OH	Ni ²⁺	11,290	[26]
com7	H	H	H	- C ₆ H ₄ OH	Cu ²⁺	12,160	[26]
com8	H	H	H	- C ₆ H ₄ OH	Zn ²⁺	10,370	[26]
com9	H	H	H	-C ₁₀ H ₆ OH	Pr ³⁺	12,940	[27]
com10	H	H	H	-C ₁₀ H ₆ OH	Nd ³⁺	13,500	[27]
com11	H	H	H	-C ₁₀ H ₆ OH	Sm ³⁺	13,800	[27]
com12	H	H	H	-C ₁₀ H ₆ OH	Eu ³⁺	14,210	[27]
com13	H	H	H	-C ₁₀ H ₆ OH	Gd ³⁺	14,380	[27]
com14	H	H	H	-C ₁₀ H ₆ OH	Y ³⁺	14,350	[27]
com15	H	H	H	-C ₁₀ H ₆ OH	Tb ³⁺	14,660	[27]
com16	H	H	H	-C ₆ H ₃ (OH)(OCH ₃)	Fe ²⁺	14,210	[28]
com17	H	H	H	-C ₆ H ₃ (OH)(OCH ₃)	Co ²⁺	14,250	[28]
com18	H	H	H	-C ₆ H ₃ (OH)(OCH ₃)	Cu ²⁺	17,050	[28]
com19	H	H	-CH ₃	-C ₆ H ₄ OH	Mn ²⁺	7,240	[29]
com20	H	H	-CH ₃	-C ₆ H ₄ OH	Cd ²⁺	7,900	[29]
com21	H	H	-CH ₃	-C ₆ H ₄ OH	Ni ²⁺	8,630	[29]
com22	H	H	-CH ₃	-C ₆ H ₄ OH	Cu ²⁺	9,810	[29]

Phát triển các phức chất mới giữa thiosemicarbazone và các ion kim loại sử dụng phương pháp ...

Ký hiệu	Thiosemicarbazone				Ion kim loại	$\log\beta_{12,\text{exp}}$	Tài liệu
	R ₁	R ₂	R ₃	R ₄			
com23	H	H	H	-C ₁₀ H ₆ OH	Mn ²⁺	9,920	[11]
com24	H	H	H	-C ₁₀ H ₆ OH	Pb ²⁺	11,370	[11]
com25	H	H	H	-C ₁₀ H ₆ OH	Co ²⁺	13,650	[11]
com26	H	H	H	-C ₁₀ H ₆ OH	Ni ²⁺	14,370	[11]
com27	H	H	H	-C ₁₀ H ₆ OH	Cu ²⁺	15,310	[11]
com28	H	H	-	- C ₉ H ₇ NO	Cu ²⁺	14,853	[12]
com29	H	H	-C ₆ H ₄ OH	-C ₆ H ₄ OH	Co ²⁺	12,215	[30]
com30	H	H	-C ₆ H ₄ OH	-C ₆ H ₄ OH	Ni ²⁺	14,194	[30]

Bảng 5. Giá trị hàng số bền dự đoán ($\log\beta_{12, \text{pred}}$) của các phức chất giữa ion kim loại và thiosemicacbazone từ mười mô hình QSPR

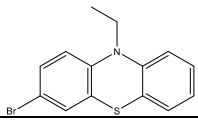
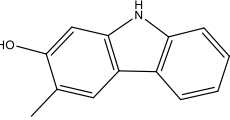
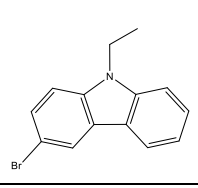
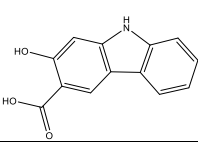
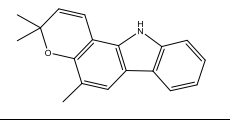
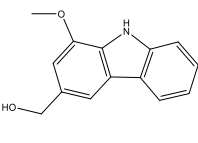
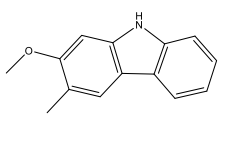
Ký hiệu	MLR1	MLR2	MLR3	MLR4	MLR5	MLR6	MLR7	MLR8	MLR9	MLR10	ANN
com1	15,4815	14,6163	12,4791	9,1579	15,1530	10,0766	9,8349	8,7834	9,4516	12,0850	9,8068
com2	17,1601	15,3536	14,6008	37,3347	11,7927	7,7659	31,6667	30,8028	28,3426	27,5604	5,1215
com3	16,5471	16,3267	10,8825	12,4310	19,1417	13,0317	9,0866	9,4922	12,5800	13,0537	15,2833
com4	15,3772	10,4325	10,3738	9,1500	16,8209	7,8309	7,0200	7,9928	12,8673	18,1056	15,6902
com5	12,0529	11,8351	10,4477	12,8631	-16,6817	9,4039	10,4069	10,9019	10,0543	10,1347	1,9878
com6	-1,4531	-7,0069	5,3001	12,8439	-13,2258	8,6498	10,9570	10,7734	-2,7829	8,7105	8,9646
com7	-0,6613	-6,9337	6,2640	11,7878	18,3794	6,6933	8,1566	8,2898	-5,5744	10,5062	6,7903
com8	-1,5657	-7,0148	7,1041	9,5997	18,2543	12,3692	8,7919	8,9385	-4,8023	9,1314	10,7772
com9	3,0667	-1,4930	10,8931	11,3075	16,4981	12,8129	11,6600	10,7466	8,4897	11,9497	4,8362
com10	3,3449	-1,4975	10,5564	11,3035	16,4981	12,0537	11,6976	10,7314	8,4897	11,9397	9,5296
com11	-7,0140	-1,5063	10,7028	11,3045	16,4981	12,5088	11,6217	10,6987	8,4897	11,9193	1,3457
com12	3,1856	-1,5107	10,2114	11,3051	16,4981	11,3580	11,6128	10,7128	8,4897	11,9080	6,2175
com13	3,3272	-1,5150	10,5540	11,3017	16,4981	12,2584	11,6332	10,6998	8,4897	11,8998	2,2149
com14	1,9115	-1,3943	10,7988	11,3087	16,4969	11,7386	11,5007	10,9109	8,4692	12,1149	3,8084
com15	3,3474	-1,5193	10,6750	11,3025	16,4981	12,5872	11,5854	10,6715	8,4897	11,8887	2,0730
com16	3,5500	-0,6104	9,0561	11,2848	14,3185	6,8239	9,5961	9,4041	12,2949	13,7162	1,3125
com17	1,4491	-0,6956	9,2045	11,3211	14,3205	7,5899	9,6750	9,4234	12,3311	13,4836	0,7502
com18	12,4061	12,2869	10,5737	19,3915	8,4067	5,3628	16,7578	17,0557	7,8176	10,0500	2,6487
com19	4,0994	-0,4401	9,1624	6,9052	14,2676	7,4641	6,4366	6,7957	7,9553	13,4705	14,1630
com20	5,5843	2,7653	7,0670	6,2315	-13,4381	3,9653	6,3507	6,9099	8,5867	11,8299	5,7772
com21	6,5270	-0,5831	8,0843	7,3391	14,2737	5,6782	6,3364	6,7031	8,0600	12,9910	1,1286
com22	1,7808	-0,5583	8,0321	8,3543	-10,1598	4,7178	6,0794	6,6557	7,9305	10,0823	2,1549
com23	2,9636	-1,2316	10,1864	11,7795	16,4907	9,3775	10,7997	10,3157	8,3617	14,3121	7,3911
com24	11,8579	9,5097	10,4498	8,7456	11,0118	10,2016	7,6384	7,7356	4,6281	14,2214	5,8276
com25	0,6012	-1,4000	9,5342	12,3667	16,4968	8,5177	10,7960	10,3108	8,4664	12,0136	2,7333
com26	5,2294	-1,4083	9,2543	12,2887	16,4968	7,8662	10,7538	10,3024	8,4664	13,8822	1,9708
com27	0,6855	-1,3894	10,9755	13,6008	-13,1858	11,2008	10,8816	10,4500	8,3280	10,9659	1,6695
com28	14,4642	13,3840	11,7168	15,3742	-15,3198	4,2618	11,9643	13,3861	14,0377	12,8841	2,5389
com29	2,9661	0,1359	11,4829	13,3793	16,3810	11,5624	9,5696	9,9573	17,8228	14,2894	1,1224
com30	8,3732	0,1291	10,3906	13,3014	16,3810	8,9793	9,3636	9,8512	17,8228	14,2528	3,4655
Q ² _{EV}	0,0580	0,0762	0,3484	0,8735	0,0083	0,0013	0,8996	0,8790	0,3194	0,4444	0,9824
MARE, %	66,3209	90,2261	23,5367	14,7948	73,9067	28,7922	19,3920	19,7907	39,4077	22,3111	5,3034

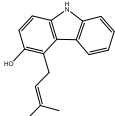
3.4. Phát triển các phức chất mới

Hai dẫn xuất phenothiazine và carbazole được chọn để thiết kế thiosemicarbazone mới và các phức chất giữa các phối tử mới với một số ion kim loại phổ biến như Cd^{2+} , Ni^{2+} , Cu^{2+} , Ag^{+} và Zn^{2+} liên quan đến bốn biến mô tả, đó là SsNH_2 , SdO , Total energy và SssCH_2 . Chúng tôi chọn các dẫn xuất này vì chúng có ứng dụng rộng rãi trong nhiều lĩnh vực và đã được tổng hợp trong các nghiên cứu thực nghiệm [31, 32]. Các thiosemicarbazone mới được hình thành bởi các nhóm phenothiazine và carbazole gắn liền tại vị trí R4 trong khi các vị trí còn lại như R₁, R₂ và R₃ của thiosemicarbazones là các nguyên tử hydro.

Một loạt các phức hợp mới được sàng lọc cẩn thận và chúng được nhúng vào dữ liệu không gian của tập luyện để kiểm tra miền ứng dụng AD (Applicability Domain) [4,5]. Kết quả có hai mươi phức chất mới đáp ứng các tiêu chuẩn của AD và chúng được dự đoán là hằng số bền từ hai mô hình $\text{QSPR}_{\text{MLR4}}$ và QSPR_{ANN} đã xây dựng. Các giá trị dự đoán của các phức chất mới ($\log\beta_{12,\text{new}}$) được đưa ra trong Bảng 6. Thêm vào đó, phương pháp ANOVA một yếu tố được sử dụng để so sánh các giá trị $\log\beta_{12,\text{new}}$ dự đoán từ hai các mô hình $\text{QSPR}_{\text{MLR4}}$ và QSPR_{ANN} . Nó chỉ ra rằng không có sự khác biệt giữa hai mô hình ($F = 1,4473 < F_{0,05} = 4,0982$).

Bảng 6. Hai mươi phức chất mới giữa ion kim loại và thiosemicabazone với giá trị $\log\beta_{12,\text{new}}$ được tính toán từ các QSPR đã xây dựng

Nhóm chức gắn vào vị trí R ₄	Ion kim loại	$\log\beta_{12,\text{new}}$		Nhóm chức gắn vào vị trí R ₄	Ion kim loại	$\log\beta_{12,\text{new}}$	
		MLR	ANN			MLR	ANN
	Cd^{2+}	18,3498	15,5887		Ag^{+}	13,5039	12,2678
	Zn^{2+}	18,3430	15,5743		Cd^{2+}	12,0050	12,2293
	Cd^{2+}	17,3734	18,9594		Cu^{2+}	13,8927	12,2008
	Ni^{2+}	18,3876	15,8605		Ni^{2+}	13,0228	12,3206
	Zn^{2+}	17,3666	18,9438		Zn^{2+}	12,0012	12,2290
	Ag^{+}	9,8381	7,5745			Ag^{+}	11,7625
	Cd^{2+}	14,2034	12,0284		Cd^{2+}	12,9629	12,3277
	Ni^{2+}	15,2188	11,6676		Cu^{2+}	14,8507	11,9910
	Zn^{2+}	14,1972	12,0283		Ni^{2+}	13,9807	12,1944

Nhóm chức gắn vào vị trí R ₄	Ion kim loại	logβ _{12,new}		Nhóm chức gắn vào vị trí R ₄	Ion kim loại	logβ _{12,new}	
		MLR	ANN			MLR	ANN
	Ag ⁺	8,3641	12,4084		Zn ²⁺	12,9591	12,3277

4. KẾT LUẬN

Nghiên cứu này đã xây dựng thành công mô hình quan hệ định lượng cấu trúc và tính chất (QSPR) bằng các phương pháp hồi quy tuyến tính đa biến (QSPR_{MLR}) và mạng thần kinh nhân tạo (QSPR_{ANN}). Các mô hình QSPR được xây dựng bằng cách sử dụng tập dữ liệu thực nghiệm với các mô tả cấu trúc thu được từ phép tính lượng tử bán thực nghiệm PM7 và PM7/sparkle. Quá trình xây dựng các mô hình QSPR đã được đánh giá một cách cẩn thận dựa trên các chỉ số thống kê như R²_{train}, Q²_{LOO}, Q²_{ex}, SE, MARE (%) và phương pháp ANOVA một yếu tố. Trong đó, hai mô hình QSPR_{MLR} và QSPR_{ANN} I(4)-HL(10)-O(1) đạt yêu cầu về khả năng áp dụng thực tế. Kết quả nhận được từ công trình này cho phép dự đoán và định hướng thực nghiệm tổng hợp các dẫn xuất thiosemicarbazone có khả năng tạo phức tốt với các ion kim loại mở ra hướng nghiên cứu mới và có nhiều hứa hẹn trong lĩnh vực phân tích, môi trường và dược học.

TÀI LIỆU THAM KHẢO

- [1]. G. Pelosi (2010). Thiosemicarbazone Metal Complexes: From Structure to Activity, *J. Open Crystallogr.*, Vol. 3, pp. 16–28.
- [2]. R. B. Singh, B. S. Garg, and R. P. Singh (1978). Analytical applications of thiosemicarbazones and semicarbazones: A review, *Talanta*, Vol. 25, 11–12, pp. 619–632.
- [3]. James J. P. Stewart (2013). Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters, *J. Mol. Model.*, Vol. 19, 1–32.
- [4]. OECD (2007), *Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships Models*, France: Organisation for Economic Co-operation and Development.
- [5]. Golbraikh A, Tropsha A (2002). Beware of q²!, *J. Mol. Graph Model.*, Vol. 20, 4, pp. 269 – 76.
- [6]. D. N. Kenie, A. Satyanarayana (2015). Protolytic Equilibria and Stability Constants of Mn (II) and Ni (II) Complexes of 3-formylpyridine Thiosemicarbazone in Sodium Dodecyl Sulphate (SDS)- Water Mixture., *Sci. Technol. Arts Res. J.*, Vol. 4, 1, pp. 74 – 79.

- [7]. D. N. Kenie, A. Satyanarayana (2015). Solution Equilibrium Study of the Complexation of Co(II) and Zn(II) with Nicotinaldehyde Thiosemicarbazone., *Sci. Technol. Arts Res. J.*, Vol. 4, 3, pp. 145 – 149.
- [8]. P. Satheesh, S. Ravichandran, V. Suryanarayanarao (2011). Spectrophotometric determination of Cu (II) and Ni (II) using 4-hydroxybenzaldehyde thiosemicarbazone., *Int. J. ChemTech. Res.*, Vol. 3, 4, pp. 2062 – 2065.
- [9]. V. Veeranna, V. S. Rao, V. V. Laxmi, T. R Varalakshmi (2013). Simultaneous Second Order Derivative Spectrophotometric Determination of Cadmium and Cobalt using Furfuraldehyde Thiosemicarbazone, *Res. J. Pharm. and Tech.*, Vol. 6, 5, pp. 577–584.
- [10]. T. Atalay, E. Ozkan (1994). Evaluation of thermodynamic parameters and stability constants of Cu(II), Ag(I) and Hg(II) complexes of 2-methylindole-3-carboxaldehyde thiosemicarbazone., *Thermochimica Acta.*, Vol. 244, pp. 291 – 295.
- [11]. Sahadev, R. K. Sharma, S. K. Sindhvani (1992). Thermal studies on the chelation behaviour of biologically active 2-hydroxy-1-naphthaldehyde thiosemicarbazone (HNATS) towards bivalent metal ions: a potentiometric study., *Thermochimica Acta*, Vol. 202, pp. 291 – 299.
- [12]. K. Sarkar, B. S. Garg (1987). Determination of thermodynamic parameters and stability constants of the complexes of p-MITSC with transition metal ions., *Thermochimica Acta*, Vol. 113, pp. 7 – 14.
- [13]. S. S. Sawhney, S. K. Chandel (1983). Solution chemistry of Cu(II)-, Co(II)-, Ni(II)-, Mn(II)- and Zn(II)-p-aminobenzaldehyde thiosemicarbazone systems., *Thermochimica Acta*, Vol. 71, pp. 209 – 214.
- [14]. S. S. Sawhney, S. K. Chandel (1984). Stability and thermodynamics of La(III)-, Pr(III)-, Nd(III)-, Gd(III)- and Eu(III)-p-nitrobenzaldehyde thiosemicarbazone systems., *Thermochimica Acta.*, Vol. 72, pp. 381 – 385.
- [15]. S. S. Sawhney, R. M. Sati (1983). pH-metric studies on Cd(II)-, Pb(II)-, Al(III)-, Cr(III)- and Fe(III)-p-nitrobenzaldehyde thiosemicarbazone systems, *Thermochimica Acta*, Vol. 66, pp. 351 – 355.
- [16]. *XLSTAT Version 2016.02.28451* (2016). USA: Addinsoft.
- [17]. *QSARIS 1.1* (2001). USA: Statistical Solutions Ltd.
- [18]. D. D. Steppan, J. Werner, and P. R. Yeater (1998). *Essential Regression and Experimental Design for Chemists and Engineers*.
- [19]. Gasteiger, J., and Zupan, J. (1993). Neural Networks in Chemistry. *Chiv. Inr. Ed. Engl.*, Vol. 32, pp. 503 – 521.
- [20]. Rumelhart D.E., Hinton G.E., Williams R. J. (1986), Learning representations by back-propagating errors., *Nature*, Vol. 323. 6088, pp. 533-536.
- [21]. *Matlab R2016a 9.0.0.341360*. (2016), USA: MathWorks.
- [22]. Pham Van Tat (2009). *Development of QSAR and QSPR*, Publisher of Natural sciences and Technique, Ha Noi.

- [23]. M. N. M. Milunovic, E. A. Enyedy, N. V. Nagy, T. Kiss, R. Trondl, M. A. Jakupec, B. K. Keppler, R. Krachler, G. Novitchi, V. B. Arion (2012). L- and D Proline *Thiosemicarbazone* Conjugates: Coordination Behavior in Solution and the Effect of Copper(II) Coordination on Their Antiproliferative Activity, *Inorg. Chem.*, Vol. 51, pp. 9309 – 9321.
- [24]. N. B. Patel, Y. J. Solanki(2016). 2,4-Dihydroxy-5-Bromo [2'Methyl] Propiophenone Thiosemicarbazone [DHBMPPT] as an Analytical Reagent: Studies on Co(II) Chelate., *J. App. Chem.*, Vol. 5, 3, pp. 654 – 660.
- [25]. K. H. Reddy, N.B.L. Prasad, T. S. Reddy (2003). Analytical properties of 1-phenyl-1,2-propanedione-2-oxime thiosemicarbazone: simultaneous spectrophotometric determination of Cu(II) and Ni(II) in edible oils and seeds, *Talanta*, Vol. 59, po. 425-433
- [26]. P. V. Bernhardt, P. C. Sharpe, M. Islam, D. B. Lovejoy, D. S. Kalinowski, D. R. Richardson (2009). Iron Chelators of the Dipyritylketone Thiosemicarbazone Class: Precomplexation and Transmetalation Effects on Anticancer Activity., *J. Med. Chem.*, Vol. 52, 2, pp. 407 – 415.
- [27]. Sahadev, R. K. Sharma, S. K. Sindhwani (1992). Potentiometric Studies on the Complexation Equilibria Between Some Trivalent Lanthanide Metal Ions and Biologically Active 2-Hydroxy-1- Naphthaldehyde Thiosemicarbazone (HNATS), *Monatshefte fur Chemie.*, Vol. 123, pp. 883 – 889.
- [28]. B. S. Garg, V. K. Jain (1989). Determination of thermodynamic parameters and stability constants of complexes of biologically active o-vanillinthiosemicarbazone with bivalent metal ions., *Thermochimica Acta.*, Vol. 146, pp. 375 – 379.
- [29]. B. S. Garg, S. Ghosh, V. K. Jain, P. K. Singh (1990). Evaluation of thermodynamic parameters of bivalent metal complexes of 2-hydroxyacetophenonethiosemicarbazone (2-HATS), *Thermochimica Acta.*, Vol. 157, pp. 365 – 368.
- [30]. F. Toribio, J. M. L. Fernandez, D. P. Bedito, M. Valcárcel (1980). 2,2'-dihydroxybenzophenone thiosemicarbazone as a spectrophotometric Reagent for the determination of copper, cobalt, nickel, and iron trace amounts in mixtures without previous separations, *Microchemical Journal.*, Vol. 25, pp. 338 – 347.
- [31]. Al-Busaidi, I. J., Haque, A., Al Rasbi, N. K., and Khan, M. S. (2019). Phenothiazine-based derivatives for optoelectronic applications: A review. *Synthetic Metals*, Vol. 257, 116189. doi:10.1016/j.synthmet.2019.116189.
- [32]. Huang, L., Feng, Z. L., Wang, Y. T., and Lin, L. G. (2017). Anticancer carbazole Alkaloids and coumarins from Clausena plants: A review. *Chinese Journal of Natural Medicines*, Vol. 15, 12, pp. 881 – 888.

DEVELOPMENT OF NEW METAL-THIOSEMICARBAZONE COMPLEXES USING QSPR MODELING METHODS

Nguyen Hoang Minh¹, Pham Van Tat², Nguyen Minh Quang^{1*}

¹ Industrial University of Ho Chi Minh City

² Hoa Sen University

*Email: nguyenminhquang@iuh.edu.vn

ABSTRACT

Based on the quantitative structure-property relationship (QSPR) models, the stability constants ($\log K$) of a series of twenty new metal-thiosemicarbazone (ML2) complexes were predicted. The QSPR models were built by the multivariate linear regression (QSPR_{MLR}) and artificial neural network (QSPR_{ANN}) methods. The variables of the models were generated from the semi-empirical quantum calculation with the new version PM7 and PM7/sparkle. The quality of the QSPR models was clearly evaluated by the statistical values of the basic principles. On the basis of those viewpoints, the ten built models produced the same four variables and the best QSPR_{MLR4} model received the results as $R^2_{\text{train}} = 0.885$; $Q^2_{\text{LOO}} = 0.860$ and $SE = 1.527$. Besides, the QSPR_{ANN} models with network architecture I(4)-HL(10)-O(1) were developed from the descriptors of QSPR_{MLR4} models and showed good results as $R^2_{\text{train}} = 0.961$; $Q^2_{\text{test}} = 0.969$ and $Q^2_{\text{validation}} = 0.978$. The suitable results from the models can be used to develop these other new derivatives for applications in the fields of chemistry, analysis and environment.

Keywords: QSPR, stability constants $\log K$, multivariate linear regression, artificial neural network, thiosemicarbazone.



Nguyễn Hoàng Minh sinh ngày 17/11/1978 tại Khánh Hòa. Năm 2001, ông tốt nghiệp Kỹ sư ngành Hóa dầu tại Trường Đại học Bách Khoa Tp.HCM. Năm 2006, ông tốt nghiệp Thạc sĩ Hóa Dầu tại Trường Đại học Bách Khoa Tp.HCM. Hiện nay, ông công tác tại Khoa công nghệ hóa học, Trường Đại học Công Nghiệp Tp. HCM.

Lĩnh vực nghiên cứu: Hóa tính toán, Hóa lý, QSPR, Phát triển các sản phẩm dầu nhờn-nhớt mới.



Phạm Văn Tất sinh ngày 30/11/1966 tại Nam Định. Năm 1989 ông tốt nghiệp ngành Hóa học tại trường đại học Tổng hợp Hà Nội; tốt nghiệp Thạc sĩ ngành Hóa học Phân tích năm 2001 tại trường đại học Đà Lạt. Năm 2006 ông bảo vệ thành công luận án Tiến sĩ tại Viện Hóa lý và Hóa lý thuyết, trường đại học Cologne, Cộng hòa Liên bang Đức. Năm 2010 ông được nhà nước công nhận chức danh phó Giáo sư chuyên ngành Hóa lý và Hóa lý thuyết. Hiện nay là trưởng Bộ môn ngành Công nghệ Kỹ thuật Môi trường, trường đại học Hoa Sen.

Lĩnh vực nghiên cứu: Hóa học tính toán lượng tử và Mô phỏng Monte Carlo các trạng thái cân bằng, Hóa học Phân tích, Hóa học Môi trường và Quản lý và Đánh giá chất lượng Môi trường bằng GIS, Viễn thám và Hệ thống Trí tuệ Nhân tạo.



Nguyễn Minh Quang sinh ngày 11/11/1977 tại Quảng Ngãi. Năm 2021, ông hoàn thành chương trình tiến sĩ tại Trường Đại học Khoa học, Đại học Huế. Hiện nay, ông là giảng viên Khoa công nghệ hóa học, Trường Đại học Công nghiệp TPHCM.

Lĩnh vực nghiên cứu: Hóa tính toán, QSPR, QSAR, Docking.

