รายงานวิจัยฉบับสมบูรณ์

โครงการ: การประยุกต์ใช้กระบวนการทางคอมพิวเตอร์วิเคราะห์

เพื่อการออกแบบและการค้นหาตัวยาใหม่

โดย อ.ดร.ชนินทร์ นันทเสนามาตร์

กรกฎาคม 2552

สัญญาเลขที่ MRG5080450

รายงานวิจัยฉบับสมบูรณ์

โครงการ: การประยุกต์ใช้กระบวนการทางคอมพิวเตอร์วิเคราะห์

เพื่อการออกแบบและการค้นหาตัวยาใหม่

ผู้วิจัย	**สังกัด**
อ.ดร.ชนินทร์ นันทเสนามาตร์	ภาควิชาจุลชีววิทยาคลินิก คณะเทคนิคการแพทย์ มหาวิทยาลัยมหิดล
ศ.ดร.วีระพงศ์ ปรัชชญาสิทธิกุล	ภาควิชาจุลชีววิทยาคลินิก คณะเทคนิคการแพทย์ มหาวิทยาลัยมหิดล

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

Abstract

Project Code: MRG5080450
Project Title: *In silico* approach for Novel Drug Discovery and Design
Principal Investigator: Dr. Chanin Nantasenamat
Mentor: Prof. Dr. Virapong Prachayasittikul
E-mail Address: mtcnt@mahidol.ac.th
Project Period: July 2007 – June 2009

This project presents computational studies of various biological systems comprising of antioxidant compounds and metallo-complexes, anti-anthrax, and anti-endotoxins. These studies utilizes quantum chemistry for providing a physicochemical description of the molecular structures and data mining techniques for correlating the relationships that exists been the molecular structures and biological/chemical activities and properties.

Antioxidants play crucial roles in scavenging oxidative damages arising from reactive oxygen species. Bond dissociation enthalpy (BDE) of phenolic O-H bond has well been accepted as an indicator of antioxidant activity since phenols donate the hydrogen atom to the free radicals thereby neutralizing its toxic effect. The BDEs from a data set of 39 antioxidant phenols were modeled using computationally inexpensive quantum chemical descriptors with multiple linear regression (MLR), partial least squares (PLS), and support vector machine (SVM). The molecular descriptors of the phenols were derived from calculations at the following theoretical levels: AM1, HF/3-21g(d), B3LYP/3-21g(d), and B3LYP/6-31g(d). Results indicated that when MLR and PLS were used as the regression methods, B3LYP/3-21g(d) gave the best performance with leave-one-out cross-validated correlation coefficients (r) of 0.917 and 0.921, respectively, while the semiempirical AM1 provided slightly lower r of 0.897 and 0.888, respectively. When SVM was used as the regression method no significant difference in the accuracy was observed for models using B3LYP/3-21g(d) and AM1 as indicated by r of 0.968 and 0.966, respectively. The quantitative structure-property relationship (QSPR) model of BDE discussed in this study offers great potential for the design of novel antioxidant phenols with robust properties.

Superoxide anions are reactive oxygen species that can attack biomolecules such as DNA, lipids and proteins to cause many serious diseases. This study reports the synthesis of copper complexes of nicotinic acid with related pyridine derivatives. The copper complexes were shown to possess superoxide dismutase (SOD) and antimicrobial activities. The copper complexes exerted SOD activity in range of 49.07-130.23 microM. Particularly, copper complex of nicotinic acid with 2-hydroxypyridine was the most potent SOD mimic with an IC(50) of 49.07 microM. In addition, the complexes exhibited antimicrobial activity against Bacillus subtilis ATCC 6633 and Candida albicans ATCC 90028 with MIC range of 128-256 microg/mL. The SOD activities were well correlated with the theoretical parameters as calculated by density functional theory at the B3LYP/LANL2DZ level of theory. Interestingly, the SOD activity of the copper complexes was demonstrated to be inversely correlated with the electron affinity, but was well correlated with both HOMO and LUMO energies. The vitamin-metal complexes described in this report are great examples of the value-added benefits of vitamins for medicinal applications.

Nicotinic acid (also known as vitamin B3) is a dietary element essential for physiological and antihyperlipidemic functions. This study reports the synthesis of novel mixed ligand complexes of copper with nicotinic and other select carboxylic acids (phthalic, salicylic and anthranilic acids). The

tested copper complexes exhibited superoxide dismutase (SOD) mimetic activity and antimicrobial activity against Bacillus subtilis ATCC 6633, with a minimum inhibition concentration of 256 microg/mL. Copper complex of nicotinic-phthalic acids (CuNA/Ph) was the most potent with a SOD mimetic activity of IC(50) 34.42 microM. The SOD activities were observed to correlate well with the theoretical parameters as calculated using density functional theory (DFT) at the B3LYP/LANL2DZ level of theory. Interestingly, the SOD activity of the copper complex CuNA/Ph was positively correlated with the electron affinity (EA) value. The two quantum chemical parameters, highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), were shown to be appropriate for understanding the mechanism of the metal complexes as their calculated energies show good correlation with the SOD activity. Moreover, copper complex with the highest SOD activity were shown to possess the lowest HOMO energy. These findings demonstrate a great potential for the development of value-added metallovitamin-based therapeutics.

Quantitative structure-activity relationship (QSAR) models were constructed for predicting the inhibition of furin-dependent processing of anthrax protective antigen of substituted guanidinylated aryl 2,5-dideoxystreptamines. Molecular descriptors calculated by E-Dragon and RECON were subjected to variable reduction using the Unsupervised Forward Selection (UFS) algorithm. The variables were then used as input for QSAR model generation using partial least squares and back-propagation neural network. Prediction was performed via a two-step approach: (i) perform classification to determine whether the molecule is active or inactive, (ii) develop a QSAR regression model of active molecules. Both classification and regression models yielded good results with RECON providing higher accuracy than that of E-DRAGON descriptors. The performance of the regression model using E-Dragon and RECON descriptors provided a correlation coefficient of 0.807 and 0.923 and root mean square error of 0.666 and 0.304, respectively. Interestingly, it was observed that appropriate representations of the protonation states of the molecules were crucial for good prediction performance, which coincides with the fact that the inhibitors interact with furin via electrostatic forces. The results provide good prospect of using the proposed QSAR models for the rational design of novel therapeutic furin inhibitors toward anthrax and furin-dependent diseases.

Bacterial lipopolysaccharides (LPS), also known as endotoxins, are major structural components of the outer membrane of Gram-negative bacteria that serve as a barrier and protective shield between them and their surrounding environment. LPS is considered to be a major virulence factor as it strongly stimulates the secretion of pro-inflammatory cytokines which mediate the host immune response and culminating in septic shock. Quantitative structure-activity relationship studies of the LPS neutralization activities of anti-endotoxins were performed using charge and quantum chemical descriptors. Artificial neural network implementing the back-propagation algorithm was selected for the multivariate analysis. The predicted activities from leave-one-out cross-validation were well correlated with the experimental values as observed from the correlation coefficient and root mean square error of 0.930 and 0.162, respectively. Similarly, the external testing set also yielded good predictivity with correlation coefficient and root mean square error of 0.983 and 0.130. The model holds great potential for the rational design of novel and robust compounds with enhanced neutralization activity.

Keywords: Drug design, data mining, machine learning, quantum chemistry, computational chemistry

Importance and Background of Research Problem

Much progress has been achieved over the centuries in prolonging life expectancy. Notable examples include sanitation and improved living condition. Drugs and compounds with therapeutic value have served as the driving force in lengthening human life span by neutralizing the negative effects of diseases. The time to market of drugs is a rather lengthy process which can take as much as 12 years (1). Furthermore, the mechanistic understanding of the biological activity of many drugs is unclear. Bioactive compounds that elicit favorable biological activity may not necessarily make good drugs as it may cause many side effects. Cheminformatics can help accelerate the drug discovery process as well as aid in the elucidation of key biological, chemical, and biophysical understanding crucial for the development of new pharmaceutical compounds. The ability to simulate the physicochemical properties of drug compounds *in silico* helps facilitate the reduction of inactive or toxic lead compounds. Consequently, the volume of bioactive molecules awaiting verification of biological activity is greatly reduced. Moreover, simulation of bioactive molecules can help provide insights on the structure-activity relationship, which facilitates the development of robust drug molecules possessing the necessary chemical substituents that is essential to biological activity.

This research project is in line with the government's aim to develop Thailand as the "Medical Hub of Asia" (2). In doing so, more and more patients from abroad would be expected to visit Thailand for medical needs therefore the demands for biomedical supplies would likely increase. Likewise, the Thailand Joint Institutional Review Board was established to oversee health research with particular emphasis on the usage of herbal medicine for humans (3). Therefore, the ability to develop novel drugs would help position Thailand to the forefront of medicine. It is expected that the accumulation of useful knowledge on the mechanistic understanding of bioactive compounds would help speed up the drug discovery process and as such reduce expenses.

Objectives

- To extract useful knowledge from the literature and biological data repository on bioactive compounds using data mining techniques.
- To apply *in silico* techniques for understanding the properties of bioactive compounds and their mechanism of action.
- To elucidate the biological, chemical, and biophysical properties of the compounds of interest.
- To elucidate the mechanism of interaction of drug-enzyme complex.
- To rationally design artificial receptors specific toward bioactive compounds of interest as well as artificial enzymes with therapeutic benefits.
- To understand the factors contributing to drug resistance.
- To apply the knowledge gained from this research project for the design of improved and robust drug compounds.

Literature Review

Drug is defined by Merriam-Webster as "a substance intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease." The therapeutic benefits of medicinal herbs have been utilized by our ancestors for centuries. Only until the last century has it became clear that the therapeutic effects of these medicinal herbs arose from bioactive molecules that are commonly known as drugs. The development of a drug can be divided into the following steps: 1) lead

discovery, 2) lead improvement, 3) preclinical studies, and 4) clinical trials. The time used to bring a new drug to market takes about 10 to 12 years with expenditures in the range of $350 millions to $450 millions (1).

In a recent article by Overington et al. (4), comprehensive analyses of all approved therapeutic drugs were performed in an attempt to better understand the properties of drugs using data mining, bioinformatics, and cheminformatics approaches. Such effort greatly plays a pivotal role to the understanding of drugs approved by the United States Food and Drug Administration (FDA). This included knowledge on the current types of drug targets, physicochemical properties of drugs, multitarget effects of drugs, localization of drug targets, and the rate of drug target innovation. In a study carried out by the National Institute for Health Care Management Research and Educational Foundation (5) on drug innovation, it was revealed that the majority of new drugs discovered between 1989 and 2000 mostly targeted known targets whereas only a small portion acted on new targets.

The early years of drug discovery started off less than a century ago made possible by the foundations laid down by chemistry and pharmacology (6). The essential principles were available for application in drug research, such as Avogadro's atomic hypothesis, theory of acids and bases, and Kekulé's theory on the structure of aromatic organic molecules (7). Most notably was Kekule's benzene theory which stimulated the birth of dye chemistry, which in turn made possible the discovery of chemoreceptors. This discovery by Paul Ehrlich was the landmark of drug research as it gave rise to chemotherapy by following the notion that different organisms have unique chemoreceptors, which can be selectively modulated (8).

Over the years, drug discovery draws on the strengths of various disciplines in the life sciences to accrue useful knowledge for the advancement of the field. Biochemistry, in particular, introduces the concepts of enzymes and receptors, which have been demonstrated to be crucial drug targets. A prominent example was the discovery of small molecule inhibitors known as sulfonamides for the metalloenzyme carbonic anhydrase (9). The discovery came as a result of the thorough description and characterization of the enzyme. Another major driving force is molecular biology which allows elucidation of the etiology of diseases at the molecular level as well as facilitates identification of molecular targets for drug intervention.

Contemporary drug discovery place great emphasis on making links between drug targets and specific disease states (10). Once the specific drug target is identified, the search for bioactive molecules known as lead compounds commences in attempts to inhibit its harmful effects. This approach is commonly known as *structure-based drug discovery*. There are also circumstances where the drug target is unknown even though the drug exerts therapeutic effect. In this situation, the drug discovery process relies on the drug as a template to yield derivatives with different substituent groups while keeping the pharmacophores (11) intact. Such approach is termed *ligand-based drug discovery*. Lead compounds are usually derived from natural products where they are subsequently functionalized to yield derivatives possessing the same or enhanced level of activity. Selection and placement of appropriate functional groups on the core structure of the lead compounds is expected to influence the bioactivity of the molecules. However, the overall contributions of the functional groups to the bioactivity require much more than that perceived visually with the naked eye.

Bringing a drug to market is a very long and costly journey involving many tedious procedures starting from searching for bioactive compounds, optimizing the compounds, and finally performing pre-clinical and clinical studies. It is estimated that the cumulative spending into drug research is $30 billion per year with only about 1 or 2 out of every 10 compounds succeeding clinical trials (12). Recently, the clinical trial for a high-density lipoprotein (HDL) raising drug called torcetrapib was cancelled due to possible side effect (13). Therefore, it is imperative that measures be developed to curb this predicament. One important aspect of this failure is the safety and efficacy of compounds reaching the clinical phase. The status of potential side effect of these lead compounds is unknown. Therefore, the determination of absorption, distribution, metabolism, excretion, and toxicity (ADMET) of lead compounds is an important task (14-16). Coupled with the limited understanding of the mechanism of actions of some drug compounds contributes to the late stage attrition. Besides the failure of lead compounds during the latter phases of the drug discovery process, other factors slowing the time-to-market rate can be attributed to the systematic testing of the compound's activity. Virtual screening (17-19) is a lucrative alternative approach to this conventional method allowing the evaluation of the activity of new compounds on the computer using established technology (20).

Bioinformatics and cheminformatics draws on the strengths of computer science and informatics to solve diverse problems in biology and chemistry, respectively. Progressions in the life sciences have amassed colossal amounts of data from the omics, including for example, genomics, proteomics, glycomics and metabolomics (21, 22). With the advent of information technology, much of the information that had once been stored in the form of encyclopedias or reference books has now been obsolete as information amasses quickly rendering it impossible to keep it updated in print form. Thus, the plethoras of biological and chemical data are stored in databases available online for quick and efficient retrieval on demand. These databases are versatile in that it allows enormous volumes of information to be stored and retrieved in an efficient manner.

Bioinformatics is defined by the National Institute of Health's (NIH) Biomedical Information Science and Technology Initiative Consortium (23) as "research, development or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, include those to acquire, store, organize, archive, analyze, or visualize such data." Bioinformatics is now routinely used by molecular biologist to search, retrieve, compare, or analyze biological sequences. Furthermore, short stretches of DNA primers could also be designed and optimized for experimentation.

Cheminformatics was defined by Frank Brown (24) in 1998 as "the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization." Besides its use in drug research, cheminformatics is applicable in a range of other fields including biochemistry, organic chemistry, physical chemistry, etc. A notable discipline in cheminformatics is quantum chemistry, which is derived from theoretical chemistry and utilizes quantum mechanics to tackle problems in chemistry. It has immense applications in the area of chemistry and biochemistry (25). The arsenal of quantum chemical techniques have been applied to the study of various chemical and biochemical problems including, for example, the correlation of structure–property relationships (26), elucidation of rapid enzymatic reactions (27), and mechanism of enzyme inhibitor interaction (28), etc. Quantum chemistry is particularly well suited for tackling questions in the life science, especially those that are difficult to determine experimentally. Take for

example, the elucidation of the mechanism of transition states cannot be investigated experimentally as they are short lived. Another example would be the identification of catalytic residues and their roles. Many questions that are not easily addressed experimentally can now be solved theoretically. Therefore, quantum chemistry is suitable for this study, particularly in obtaining insights on the mechanism of actions of drugs as well as for the generation of molecular descriptors that provides a physicochemical description of the molecules of interest. These quantum chemical descriptors are then used by the machine learning methods to derive QSPR model(s).

Molecular modeling is a general term referring to the theoretical study of molecular systems ranging from small chemical systems to large biological systems. In molecular modeling, the 3-dimensional geometries of molecules are available for visualization, manipulation, and simulation. The physicochemical properties of molecules derived from molecular simulations could be used for qualitative and quantitative structure-activity relationship studies as well as for deriving insights on chemical reactivity and mechanism of action.

The paradigm of bioinformatics and cheminformatics essentially involves the generation of information from available data and subsequently transforming it into useful knowledge. Simple data and information could easily be transformed into knowledge within a certain time manually by an expert. However, analyzing large amounts of data and information would consume substantially more time. Coupled with the inherent complexity of the data and information would likely prolong this task. Advance in artificial intelligence makes it possible to train the computer to recognize patterns and trends from biological and chemical data and, in doing so, extract useful knowledge in a timely fashion.

Bioinformatics can be simplified to include the study of genes and proteins while cheminformatics can be simplified to refer to the study of lead compounds and drugs. The interface of both disciplines makes it possible to evaluate the performance of lead and drug compounds against their respective target *in silico* via techniques such as molecular docking. Such approach facilitates the elucidation of the relative binding affinity of the prospective compounds as well as their mode of interaction within the active site.

References

1. Chow S-C, Liu J-P. Design and Analysis of Clinical Trials: Concepts and Methodologies. New Jersey: John Wiley & Sons, Inc.; 2004.
2. Developing Thailand as a Medical Hub of Asia. The Government Public Relations Department, http://thailand.prd.go.th/the_inside_viewphp?id=89. 2004.
3. Research for Human Health Getting Easier in Thailand. The Government Public Relations Department, http://thailand.prd.go.th/the_inside_viewphp?id=1663. 2006.
4. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? Nat Rev Drug Discov. 2006 Dec;5(12):993-6.
5. NIHCM. Changing Patterns of Pharmaceutical Innovation. National Institute for Health Care Management Research and Educational Foundation. 2002:1-24.
6. Drews J. Stategic trends in the drug industry. Drug Discov Today. 2003 May 1;8(9):411-20.
7. Moore FJ. A History of Chemistry. New York: McGraw-Hill; 1918.
8. Ehrlich P. Partial cell functions. Nobel Lecture; 1908.
9. Supuran CT, Scozzafava A, Casini A. Carbonic anhydrase inhibitors. Med Res Rev. 2003 Mar;23(2):146-89.
10. Adam D. Genetics group targets disease markers in the human sequence. Nature. 2001 Jul 12;412(6843):105.

11. Milne GW, Nicklaus MC, Wang S. Pharmacophores in drug design and discovery. SAR QSAR Environ Res. 1998;9(1-2):23-38.
12. Cuatrecasas P. Drug discovery in jeopardy. J Clin Invest. 2006 November 1, 2006;116(11):2837-42.
13. Pearson H. Why did a promising heart drug fail. Nature News, http://www.nature.com/news/2006/061211/full/444794ahtml. 2006.
14. Li AP. Screening for human ADME/Tox drug properties in drug discovery. Drug Discov Today. 2001 Apr 1;6(7):357-66.
15. Lin J, Sahakian DC, de Morais SM, Xu JJ, Polzer RJ, Winter SM. The role of absorption, distribution, metabolism, excretion and toxicity in drug discovery. Curr Top Med Chem. 2003;3(10):1125-54.
16. Selick HE, Beresford AP, Tarbit MH. The emerging importance of predictive ADME simulation in drug discovery. Drug Discov Today. 2002 Jan 15;7(2):109-16.
17. Hou T, Xu X. Recent development and application of virtual screening in drug discovery: an overview. Curr Pharm Des. 2004;10(9):1011-33.
18. Stahura FL, Bajorath J. Virtual screening methods that complement HTS. Comb Chem High Throughput Screen. 2004 Jun;7(4):259-69.
19. Green DV. Virtual screening of virtual libraries. Prog Med Chem. 2003;41:61-97.
20. Lengauer T, Lemmen C, Rarey M, Zimmermann M. Novel technologies for virtual screening. Drug Discov Today. 2004 Jan 1;9(1):27-34.
21. De Keersmaecker SC, Thijs IM, Vanderleyden J, Marchal K. Integration of omics data: how well does it work for bacteria? Mol Microbiol. 2006 Dec;62(5):1239-50.
22. Ge H, Walhout AJ, Vidal M. Integrating 'omic' information: a bridge between genomics and systems biology. Trends Genet. 2003 Oct;19(10):551-60.
23. National Institutes of Health Biomedical Information Science and Technology Initiative, http://www.bisti.nih.gov. 2000.
24. Brown D, Superti-Furga G. Rediscovering the sweet spot in drug discovery. Drug Discov Today. 2003 Dec 1;8(23):1067-77.
25. Cramer CJ. Essentials of Computational Chemistry: Theories and Models. second ed. ed. New Jersey: John Wiley & Sons; 2004.
26. Nantasenamat C, Naenna T, Isarankura Na Ayudhya C, Prachayasittikul V. Quantitative prediction of imprinting factor of molecularly imprinted polymers by artificial neural network. J Comput Aid Mol Des. 2005;19(7):509-24.
27. Warshel A. Computer simulations of enzyme catalysis: methods, progress, and insights. Annu Rev Bioph Biom. 2003;32:425-43.
28. Jenwitheesuk E, Samudrala R. Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations. BMC Struct Biol. 2003 Apr 1;3:2.

Methodology

This research project aims to use *in silico* approach for the study of biological and chemical mechanism of bioactive compounds. Particular emphasis is placed on understanding of key processes that could drive the discovery and design of promising lead compounds. The project is comprised of the following procedures:

Step 1

Search and retrieve relevant information of bioactive compounds from the literature.

- Key information comprises of etiology of diseases, identification of disease target, insights of the mechanism of target inhibition, current trends and perspectives of target of interest, factors contributing to drug resistance, and strategy to combat resistance.

Step 2

Construct 2-dimensional and 3-dimensional chemical structures of the compounds.

- Retrieve molecular structures of compounds from online databases such as PubChem for small molecules and Protein Data Bank for protein structures.
- Identify relevant physiological condition of biomolecules in order to properly account for its protonation state.
- Identify native isomeric form of small molecules from relevant literature.

Step 3

Perform Quantitative Structure-Activity Relationship (QSAR) study.

- Retrieve information of bioactive compounds from step 1 and prepare chemical structures as outlined in step 2.
- Calculate molecular descriptors from quantum chemical and molecular simulations.
- Construct regression or predictive model with machine learning techniques using molecular descriptors as independent variables.
- Refine regression or predictive model using intuition-based or systematic-based criteria.

Step 4

Perform mechanistic investigation using quantum chemistry.

- Employ knowledge gained from step 1 to prepare appropriate representation of the molecular structures as explained in step 2.

Results and Discussion

This research project demonstrates the utilization of computational approaches for computational design and understanding key mechanisms of various biological systems comprising of antioxidant compounds and metallo-complexes, anti-anthrax, and anti-endotoxins. As a result of this endeavor, the findings were published in international journals which encompasses the disciplines of molecular design and medicinal chemistry that is highly suitable and relevant for dissemination to the scientific community sharing the same interests and enthusiasm for drug design and development. We have demonstrated through these findings the employment of quantum chemistry for description of molecular systems in terms of their physicochemical properties (e.g. electronic energy, charge, dipole moment, energies of the highest occupied molecular orbital, energies of the lowest occupied molecular orbital, etc.). Such calculated molecular descriptors serve as information which allows us to discern the origins of biological/chemical properties as each molecule contains slightly different substituents, or functional groups which gives rise to changes in the electronic distribution within the molecule. The utilization of quantum chemical parameters in understanding the origins of antioxidant activity has been demonstrated in Papers **III** and **IV** on metal complexes of nicotinic-aromatic carboxylic acids and pyridine derivatives. The tested antioxidant activities of these complexes were confirmed and elucidated by the calculated molecular properties. The molecular descriptors were also used for the development of quantitative structure-activity relationship for modeling of biological/chemical properties of the following molecular systems: antioxidants (Paper **I**), anti-anthrax (Paper **II**), and anti-endotoxins (Paper **V**). The QSAR models was developed using data mining techniques such as artificial neural network and support vector machine. These technique allows the computer to learn from biological/chemical data by recognizing inherent patterns. Results suggests that proper representation of the protonation state of the molecules was crucial for

development of good predictive model. It was also found that the selection of suitable theoretical level could generate accurate models while consuming minimal computational resources. The identification and proper treatment of potent outliers as elucidated by statistical methods was also important for the construction of accurate QSAR models. It is also worthwhile to mention that the selection of optimal subset of molecular descriptor can also reduce the amount of computational calculation while at the same time being beneficial for the predictive performance. The computational approach proposed in this project is anticipated to be of great benefit for computational molecular design as well as for elucidating key molecular mechanisms which has successfully been demonstrated herein.

Output ที่ได้จากโครงการ

Publications

I. **Nantasenamat C**, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. Prediction of bond dissociation enthalpy of antioxidant phenols by support vector machine. *J Mol Graph Model*. 2008;27(2):188-96.

II. Worachartcheewan A, **Nantasenamat C**, Naenna T, Isarankura-Na-Ayudhya C, Prachayasittikul V. Modeling the activity of furin inhibitors using artificial neural network. *Eur J Med Chem*. 2009;44(4):1664-73.

III. Suksrichavalit T, Prachayasittikul S, Piacham T, Isarankura-Na-Ayudhya C, **Nantasenamat C**, Prachayasittikul V. Copper complexes of nicotinic-aromatic carboxylic acids as superoxide dismutase mimetics. *Molecules*. 2008;13(12):3040-56.

IV. Suksrichavalit T, Prachayasittikul S, **Nantasenamat C**, Isarankura-Na-Ayudhya C, Prachayasittikul V. Copper complexes of pyridine derivatives with superoxide scavenging and antimicrobial activities. *Eur J Med Chem*. 2009;44(8):3259-65.

V. Thippakorn C, Suksrichavalit T, **Nantasenamat C**, Tantimongcolwat T, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. Modeling the LPS neutralization activity of anti-endotoxins. *Molecules*. 2009;14(5):1869-88.

Conference Presentation

VI. Piacham T, **Nantasenamat C**, Suksrichavalit T, Isarankura-Na-Ayudhya C, Prachayasittikul V. Synthesis of molecularly imprinted nanosphere for recognition of anti-cancer agent, α-tocopherol succinate. AsiaSense: The 4th International Conference on Sensors; July 29-31, 2009, Bangkok, Thailand.

VII. **Nantasenamat C**, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. Prediction of bond dissociation enthalpy of antioxidant phenols by support vector machine. *The Sixth Princess Chulabhorn International Science Congress*, *November 25-29, 2007*, Bangkok, Thailand.

VIII. Worachartcheewan A, **Nantasenamat C**, Naenna T, Isarankura-Na-Ayudhya C, Prachayasittikul V. Modeling the activity of furin inhibitors using artificial neural network. *The Sixth Princess Chulabhorn International Science Congress*, *November 25-29, 2007*, Bangkok, Thailand.

Journal of Molecular Graphics and Modelling 27 (2008) 188–196

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Prediction of bond dissociation enthalpy of antioxidant phenols by support vector machine

Chanin Nantasenamat [a], Chartchalerm Isarankura-Na-Ayudhya [a,*], Thanakorn Naenna [b], Virapong Prachayasittikul [a,*]

[a] Department of Clinical Microbiology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[b] Department of Industrial Engineering, Faculty of Engineering, Mahidol University, Nakhon Pathom 73170, Thailand

ARTICLE INFO

Article history:
Received 10 January 2008
Received in revised form 8 April 2008
Accepted 8 April 2008
Available online 14 April 2008

Keywords:
Antioxidant
Bond dissociation enthalpy
Support vector machine
Phenols
QSPR

ABSTRACT

Antioxidants play crucial roles in scavenging oxidative damages arising from reactive oxygen species. Bond dissociation enthalpy (BDE) of phenolic O–H bond has well been accepted as an indicator of antioxidant activity since phenols donate the hydrogen atom to the free radicals thereby neutralizing its toxic effect. The BDEs from a data set of 39 antioxidant phenols were modeled using computationally inexpensive quantum chemical descriptors with multiple linear regression (MLR), partial least squares (PLS), and support vector machine (SVM). The molecular descriptors of the phenols were derived from calculations at the following theoretical levels: AM1, HF/3-21g(d), B3LYP/3-21g(d), and B3LYP/6-31g(d). Results indicated that when MLR and PLS were used as the regression methods, B3LYP/3-21g(d) gave the best performance with leave-one-out cross-validated correlation coefficients (r) of 0.917 and 0.921, respectively, while the semiempirical AM1 provided slightly lower r of 0.897 and 0.888, respectively. When SVM was used as the regression method no significant difference in the accuracy was observed for models using B3LYP/3-21g(d) and AM1 as indicated by r of 0.968 and 0.966, respectively. The quantitative structure–property relationship (QSPR) model of BDE discussed in this study offers great potential for the design of novel antioxidant phenols with robust properties.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Reactive oxygen species (ROS) are natural byproducts of normal aerobic metabolism which include oxygen ions, free radicals, and peroxides [1]. Cells are normally safeguarded from the deleterious effects of ROS by antioxidant enzymes (e.g. superoxide dismutases and catalases) and small antioxidant molecules (e.g. ascorbic acid, vitamin E, and glutathione) [2]. A delicate balance in the production and elimination of ROS is crucial to normal cell and tissue function. However, perturbation to this equilibrium by changes in environmental factors give rise to a condition known as oxidative stress [3]. This oxidative damage affects various biological macromolecules such as DNA, RNA, proteins, and membrane lipids. The highly reactive nature of ROS stems from the presence of unpaired valence shell electrons, which can easily accept or transfer an electron.

Antioxidants are substances that play important roles in combating oxidative damages that are caused by ROS [2]. Antioxidants neutralize ROS by intercepting and interacting with reactive radicals as summarized as follows:

$$RO_2^{\bullet} + AOH \rightarrow ROOH + AO^{\bullet} \quad (1)$$

where $RO_2^{\bullet}$ is the peroxy radical, AOH is the antioxidant, ROOH is the lipid hydroperoxide, and $AO^{\bullet}$ is the phenoxy radical.

Many naturally occurring antioxidants found thus far are based on phenolic compounds. Therefore, it is of great interest to be able to characterize the radical scavenging activities of these phenolic antioxidants. The use of O–H bond dissociation enthalpies (BDEs) in the characterization of phenolic antioxidants has well been documented [4–7]. Phenolic antioxidants with low BDEs are considered to have good antioxidative activity since they readily donate the hydrogen atom from the O–H bond to incoming ROS radicals. This is further supported by the fact that O–H BDE has been found to be well correlated with the activation energy of the H-abstraction reaction [8,9]. On the basis of these principles, O–H BDE has been widely used as a measure of the efficiency of radical scavenging activity [4–7,10,11]. The measurement of O–H BDE has

* Corresponding authors. Tel.: +66 2 418 0227; fax: +66 2 412 4110.
E-mail addresses: mtcis@mahidol.ac.th (C. Isarankura-Na-Ayudhya), mtvpr@mahidol.ac.th (V. Prachayasittikul).

1093-3263/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2008.04.005

traditionally been made in solutions (such as in water or DMSO) [12] and subsequently performed in the gas phase [6,7]. The phenolic O–H BDE values has been found to differ significantly from different experimental studies ranging from 84.0 to 91.6 kcal mol^{-1} [13]. Such diversity in O–H BDE values implies the uncertainty of even experimentally derived values. Therefore, there is sufficient room for improving upon the calculation methodologies of O–H BDE.

Two theoretical approaches exist for the calculation of O–H BDE values. The first approach involves two series of calculations, one for each of the two forms of the antioxidant phenols: (1) the neutral state and (2) their respective radical states. To achieve accurate calculation of thermochemical properties, multideterminantal methods that thoroughly account for the effect of electron correlation or multilevel methods entailing numerous sequential high-level calculations coupled with energy corrections are often required [14–17]. The former could be obtained from coupled clustered [CCSD(T)] [18] or quadratic configuration [QCISD(T)] [14] methods in combination with very large basis sets. The calculated results are then extrapolated to the complete basis set (CBS) limit. However, such approach is limited to only small molecule due to its computationally intensive nature. In order to study larger molecular systems, an alternative approach such as the Gaussian-*n* series (e.g. Gaussian-3 and the latest Gaussian-4 theory) [19,20] and the CBS methods [21] are typically used. Gaussian-*n* methods are composite approaches that aim to approximate the result of a more expensive calculation by performing a set of high-level correlation calculations [CCSD(T), MP4, and QCISD(T)] with moderate sized basis sets. Likewise, CBS methods are sophisticated energy computations involving several pre-defined calculations for the purpose of generating very accurate energies [17,22–24].

However, such methods can be time-consuming which contradicts with our objective of a quick and inexpensive approach towards predicting the O–H BDE. Confronted with similar computational constraints of achieving accurate C–H BDEs at the expense of high computational cost, Lewin and Cramer tackled the problem by developing rapid quantum mechanical models for the accurate estimation of C–H BDEs [25]. To meet this challenge, this study utilizes quantitative structure–property relationship (QSPR) methods for correlating the physicochemical properties of phenolic antioxidants with their respective BDEs.

Herein, the application of computationally inexpensive quantum chemical descriptors in combination with robust supervised statistical and machine learning approaches was demonstrated for the development of QSPR models. To achieve that goal, the search for suitable theoretical level in calculating the molecular descriptors and the selection of reliable supervised learning methods for constructing the QSPR model was performed. Quantum chemical descriptors were derived from geometry optimizations at the following theoretical levels: (i) semiempirical AM1 method and at the *ab initio* levels, (ii) HF/3-21g(d), (iii) B3LYP/3-21g(d), and (iv) B3LYP/6-31g(d). Three supervised learning approaches, namely (i) multiple linear regression (MLR), (ii) partial least squares (PLS) regression, and (iii) support vector machine (SVM), were used for correlating the structures of antioxidant phenols with the hydroxyl BDEs. Two models, particularly descriptors derived from AM1 or B3LYP/3-21g(d) calculations, were identified to provide equally good predictive performance with support vector machine in an economical and efficient manner.

2. Methodology

2.1. Data set

The BDEs of 39 antioxidant phenols (Table 1) were obtained from the work of Bordwell and Cheng [26]. The BDEs of phenolic O–H bonds were estimated from the oxidation potentials and pK_{HA} values of phenoxide and naphthoxide ions in DMSO. The initial geometries of the antioxidant phenols were constructed using GaussView [27].

Table 1
Data set of the antioxidant phenols

No.	Name	CAS No.	Substituent	ΔBDE (kcal/mol)
1	Phenol	108-95-2	H	0
2	*o*-Cresol	95-48-7	2-Me	−1.65
3	*m*-Cresol	108-39-4	3-Me	−0.45
4	*p*-Cresol	106-44-5	4-Me	−1.15
5	3,5-Dimethylphenol	108-68-9	3,5-Me_2	−0.75
6	2,6-Dimethylphenol	576-26-1	2,6-Me_2	−4.35
7	4-*tert*-Butylphenol	98-54-4	4-*t*-Bu	−1.15
8	2,6-di-*tert*-Butylphenol	128-39-2	2,6-*t*-Bu_2	−7.75
9	2,4,6-tri-*tert*-Butylphenol	732-26-3	2,4,6-*t*-Bu_3	−7.65
10	4-Phenylphenol	92-69-3	4-Ph	−2.25
11	2-Methoxyphenol	90-05-1	2-MeO	−3.85
12	3-Methoxyphenol	150-19-6	3-MeO	0.35
13	Hydroquinone	123-31-9	4-OH	−8.35
14	3-Aminophenol	591-27-5	3-NH_2	−1.85
15	3-Dimethylaminophenol	99-07-0	3-Me_2N	−1.95
16	4-Aminophenol	123-30-8	4-NH_2	−12.55
17	4-Dimethylaminophenol	619-60-3	4-NMe_2	−9.55
18	2-Chlorophenol	95-57-8	2-Cl	0.15
19	3-Chlorophenol	108-43-0	3-Cl	1.95
20	4-Chlorophenol	106-48-9	4-Cl	0.45
21	3,5-Dichlorophenol	591-35-5	3,5-Cl_2	4.05
22	3,4,5-Trichlorophenol	609-19-8	3,4,5-Cl_3	3.25
23	4-Bromophenol	106-41-2	4-Br	0.85
24	*m*-Trifluoromethylphenol	98-17-9	3-CF_3	3.95
25	*p*-Trifluoromethylphenol	402-45-9	4-CF_3	5.45
26	*m*-Hydroxyacetophenone	121-71-1	3-MeCO	1.95
27	*p*-Hydroxyacetophenone	99-93-4	4-MeCO	2.95
28	3-Nitrophenol	554-84-7	3-NO_2	4.45
29	4-Nitrophenol	100-02-7	4-NO_2	4.85
30	4-Methoxyphenol	150-76-5	4-OMe	−5.25
31	4-Hydroxybenzophenone	1137-42-4	4-PhCO	2.65
32	3-Methylsulfonylphenol	14763-61-2	3-$MeSO_2$	2.45
33	4-Methylsulfonylphenol	14763-60-1	4-$MeSO_2$	5.15
34	3-Cyanophenol	873-62-1	3-CN	4.05
35	4-Cyanophenol	767-00-0	4-CN	4.35
36	1-Naphthol	90-15-3	1-NpOH	−5.85
37	4-Hydroxyphenolate	20217-26-9	4-O^-	−16.85
38	2-Naphthol	135-19-3	2-NpOH	−1.85
39	6-Bromo-2-naphthol	15231-91-1	6-Br-2-NpOH	−1.35

2.2. Descriptor generation

The molecular structures of antioxidant phenols were subjected to full geometry optimizations without symmetry constraints at the semiempirical level using the Austin Model 1 (AM1) method. Likewise, calculations at the *ab initio* levels were performed with Gaussian 03 [28] using the Hartree–Fock functional with the 3-21g(d) basis set (HF/3-21g(d)) and Becke's three-parameter hybrid Lee–Yang–Parr functional with the 3-21g(d) basis set (B3LYP/3-21g(d)) and with the 6-31g(d) basis set (B3LYP/6-31g(d)). The following quantum chemical descriptors were derived from these calculations: total energy (E_{Total}), dipole moment (μ), energy of the highest occupied molecular orbital (E_{HOMO}; negative value of E_{HOMO} was used as a measure of the ionization potential), energy of the lowest unoccupied molecular orbital (E_{LUMO}; negative value of E_{LUMO} was used as an indicator of the electron affinity), atomic charge on hydroxyl hydrogen (q_H), atomic charge on hydroxyl oxygen (q_O), and lengths of O–H bond (R_{O-H}). The atomic charges were derived from Mulliken population analysis. Ionization potential and electron affinity were calculated according to Koopmans' theorem [29].

2.3. Data pre-processing

Since the independent variables were of different range, they were adjusted to comparable scale by standardization using the following equation:

$$x_{ij}^{\sin} = \frac{x_{ij} - \bar{x}_j}{\sum_{i=1}^{N} (x_{ij} - \bar{x}_j)^2 / N} \quad (2)$$

where $x_{ij}^{\sin}$ represents the standardized value, x_{ij} represents the value of each sample, $\bar{x}_j$ represents the mean of each descriptor, and N represents the sample size of the data set.

2.4. Multivariate regression

Multivariate regression was performed to correlate the independent variables, which in our case are the quantum chemical descriptors, with the dependent variable BDE. Three supervised learning approaches, namely MLR, PLS, and SVM, were used for correlating the structures of antioxidant phenols with the hydroxyl BDEs.

2.4.1. Multiple linear regression

MLR models were calculated using The Unscrambler 9.5 [30] software package to obtain equations of the following form:

$$Y = B_0 + \sum B_n X_n \quad (3)$$

where Y represents the BDEs of the antioxidant phenol compounds, B_0 represents the intercept, B_n represents the regression coefficients of descriptors X_n.

2.4.2. Partial least squares regression

PLS analysis [31,32] was also performed using The Unscrambler 9.5 software package with the PLS1 algorithm. The descriptors were subjected to pre-processing by mean-centering and auto-scaling to zero mean and unit variance according to Eq. (2). The amount of variables presented in the descriptor matrix was reduced to a small number of latent variables called PLS components (PCs), which retain the core information from the original data set. The PCs are few, orthogonal, and function as predictors of the dependent variable. The optimal number of PCs was determined according to the method of Haaland and Thomas [33] from a plot of PC versus the mean squared error (MSE) using leave-one-out cross-validation (LOO-CV). MSE was calculated using the following equation:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{N} (p_i - a_i)^2}{n} \quad (4)$$

where p_i represents the predicted output, a_i represents the experimental value, and n represents the number of antioxidant phenols presented in the data set.

2.4.3. Support vector machine

SVM calculations were performed using John Platt's Sequential Minimal Optimization (SMO) algorithm [34] with the Waikato Environment for Knowledge Analysis (Weka) version 3.4.7 [35]. SVM is a learning approach developed by Vapnik and co-workers based on the Statistical Learning Theory [36,37]. The technique has demonstrated much success in modeling biological and chemical properties as demonstrated in literatures [38–40]. Detailed accounts of SVM theory can be found in several excellent books and tutorials [36,41–46]. Here we briefly delve into the main concepts of SVM.

A training set of m compounds with known biological activity y_i (e.g. O–H BDE) and structurally derived descriptors x_i are represented as $\{(x_i, y_i)\}_{i=1}^{m}$, where correlations between structure and activities are defined by $y_i = f(x_i)$. The term $f(x_i)$ can be represented by a linear function of the form

$$f(x_i) = \langle w_i, x_i \rangle + b \quad (5)$$

where w designates the weight vector of the linear function and b corresponds to the threshold coefficient. SVM approximates the set of data with a linear function that is formulated in the high dimensional feature space with the following function:

$$y = \sum_{i=1}^{m} w_i \Phi(x_i) + b \quad (6)$$

where $\{\Phi(x_i)\}_{i=1}^{m}$ represents the features of input variables subjected to kernel transformation while $\{w_i\}_{i=1}^{m}$ and b are coefficients.

SVM is essentially a linear learning approach that was originally devised for classification problems. However, it is also amenable to regression problems through the use of ε-insensitive loss function. SVM can handle data possessing non-linear relationships via the so-called kernel trick. Kernel transformation is essentially a projection of the descriptor matrix from the input space into the higher dimensional feature space (Fig. 1). This can be described by the following equation:

$$K(x, y) = \langle \phi(x) \cdot \phi(y) \rangle \quad (7)$$

where K is a kernel function and ϕ is a mapping from input space $X \in x, y$ to the feature space F.

Several kernel functions are available for non-linear transformation of the input space. Popular kernel functions used in SVM include the variance–covariance based linear and polynomial

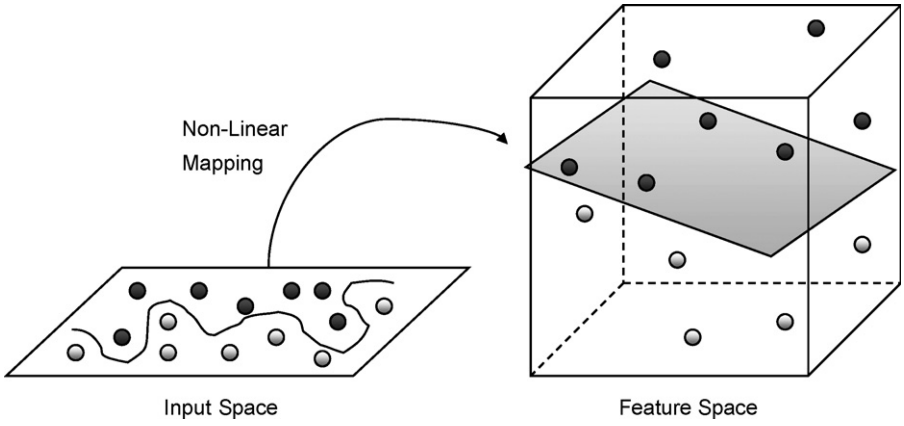


Fig. 1. Schematic of non-linear mapping of input space onto higher dimensional feature space by kernel transformation.

kernels and the Euclidean distance based radial basis function kernels.

Radial basis function kernel was used to perform this non-linear mapping as described by the following equation:

$$K(x,y) = \exp(-\gamma\|x-y\|^2) \quad (8)$$

After kernel transformation, the new feature space allows the data to be linearly separable by hyperplanes where hyperplane that maximizes the distance between the data samples was selected by the algorithm as the maximal hyperplane.

Minimization of the regularized risk function satisfies two essential properties of SVMs by means of estimating coefficients w and b: (i) define regression estimation by performing risk minimization with respect to the ε-insensitive loss function and (ii) perform risk minimization based on the SRM principle in which elements of the structure are defined by the inequality $\|w\|^2 \leq$ constant.

The regularized risk function is defined as

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(y_i, f(x_i,w)) + \frac{1}{2}\|w\|^2 \quad (9)$$

where $(C/N)\sum_{i=1}^{N} L_\varepsilon(y_i, f(x_i,w))$ is the empirical error (risk) and $(1/2)\|w\|^2$ is a measure of function flatness. The empirical error is measured by the ε-insensitive loss function $L_\varepsilon(y, f(x,w))$ in which errors below ε would not be penalized. The penalty parameter C is a regularized constant responsible for determining the trade-off between the empirical error and the model complexity.

The estimation performance of SVM regression models is determined by the ε-insensitive loss function as follows:

$$L_\varepsilon(y, f(x,w)) = \begin{cases} |y - f(x,w)| - \varepsilon & \text{for} |y - f(x,w)| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The parameter ε is referred to as the tube size, and it is defined as the approximation accuracy placed on the training data points. Essentially, the goal of support vector regression is to select a function $f(x)$ such that there is at most ε deviation from the actual value y_i for all training data while being as flat as possible. In other words, the loss function ignores errors as long as it is less than ε but would accept no significant deviation from it.

The generalization performance of the SVM model relies on the proper selection of parameters. For RBF kernels, there are two parameters involved: the complexity parameter C and the RBF kernel width γ. As the parameter C and gamma are not universally optimal to all problems an empirical parameter search is required. Therefore, determination of the optimal configuration of the SVM model was performed via a two-level grid search [47,48]. Initially, a coarse grid search was performed by exponentially adjusting the values of (C, γ). Regions affording good performance are identified and then subjected to further refinement via a local grid search.

2.5. Variable selection

MLR regression coefficients were used as a measure of the variable's importance toward the regression model. Particularly, the magnitude of the regression coefficients indicates the relative perturbation each independent variable has on the dependent variable.

2.6. Internal validation procedure

LOO-CV is an internal validation procedure used to provide estimates of the predictivity of the data sets in a cost-effective manner as all data samples were used in model development [35]. LOO-CV initiates by leaving one data sample out as the testing set while using $n-1$ samples as the training set. This procedure was carried out iteratively until all data samples were given the chance to be left out as the testing set.

2.7. Statistical analysis

The adjusted value of R^2 accounts for the number of predictors (independent variables) in the model. It is calculated according to the following equation:

$$R^2_{\text{adj}} = 1 - (1-R^2)\cdot\left(\frac{n-1}{n-p}\right) \quad (11)$$

where n represents sample size and p represents the number of predictors.

The F ratio between explained (R^2) and unexplained ($1-R^2$) variance with m and $n-m-1$ degrees of freedom is calculated according to the following equation:

$$F_{(m,n-m-1)} = \frac{R^2/m}{(1-R^2)/(n-m-1)} \quad (12)$$

where m is the number of independent variables and n is the number of compounds presented in the data set.

3. Results and discussion

The present investigation employs readily available quantum chemical descriptors in the construction of QSPR models that reliably calculates the bond dissociation enthalpies of antioxidant phenols. To achieve robust model, series of calculations were performed by looking into the performance difference as afforded by descriptors calculated at various theoretical levels followed by regression analysis with different multivariate methods. The molecular structures of the antioxidant phenols were drawn into the computer according to the structures reported by Bordwell and Cheng [26]. The substituent effect arising from variations of the functional moieties of a library of phenolic compounds was accounted for through the use of quantum chemical descriptors. Essentially, these independent variables provided quantitative description of the molecular features as well as information of their distinct functional variations. Multivariate analysis was then performed using these independent variables as input for the development of predictive models of BDEs. Prior to performing regression studies, the data was standardized according to Eq. (2) as to scale the values to comparable levels where the mean and standard deviation were adjusted to 0 and 1, respectively.

Evidently, the quality of the descriptors as well as the type of multivariate approach used for constructing the QSPR model exerts great influence on its predictive performance. Our previous investigations have demonstrated the usefulness of quantum chemical descriptors in modeling the spectral properties of the green fluorescent protein [49] and the imprinting factor of molecularly imprinted polymers [50]. The optimal theoretical level required for generating molecular descriptors that provide good predictivity was examined at the following theoretical approaches: AM1, HF/3-21g(d), B3LYP/3-21g(d), and B3LYP/6-31g(d). Of the four theoretical levels tested, it was observed that AM1 and B3LYP/3-21g(d) exhibited better accuracy than the other theoretical levels as observed from r of 0.814 and 0.895, respectively (Table 2). Therefore, these two theoretical approaches were selected for further investigations.

It was noticed that the B3LYP functional in combination with the smaller basis set 3-21g(d) unexpectedly outperformed that of the higher basis set 6-31g(d). This technical anomaly could be attributed to possible error that may arise as a result of calculations

Table 2
Predictive performance of MLR models using all compounds of the data set

Descriptors	$r_{Training}$	$r_{LOO\text{-}CV}$	R^2_{LOO-CV}	$R^2_{LOO-CV(adj)}$	F ratio[a]	$RMS_{Training}$	$RMS_{LOO\text{-}CV}$
AM1	0.932	0.814	0.663	0.633	16.201	1.834	4.687
HF/3-21g(d)	0.889	0.761	0.579	0.542	11.352	2.316	4.624
B3LYP/3-21g(d)	0.950	0.895	0.801	0.783	33.212	1.583	2.743
B3LYP/6-31g(d)	0.914	0.730	0.533	0.492	9.412	2.059	4.773

[a] Critical F value with 4 and 33 degrees of freedom is 2.659.

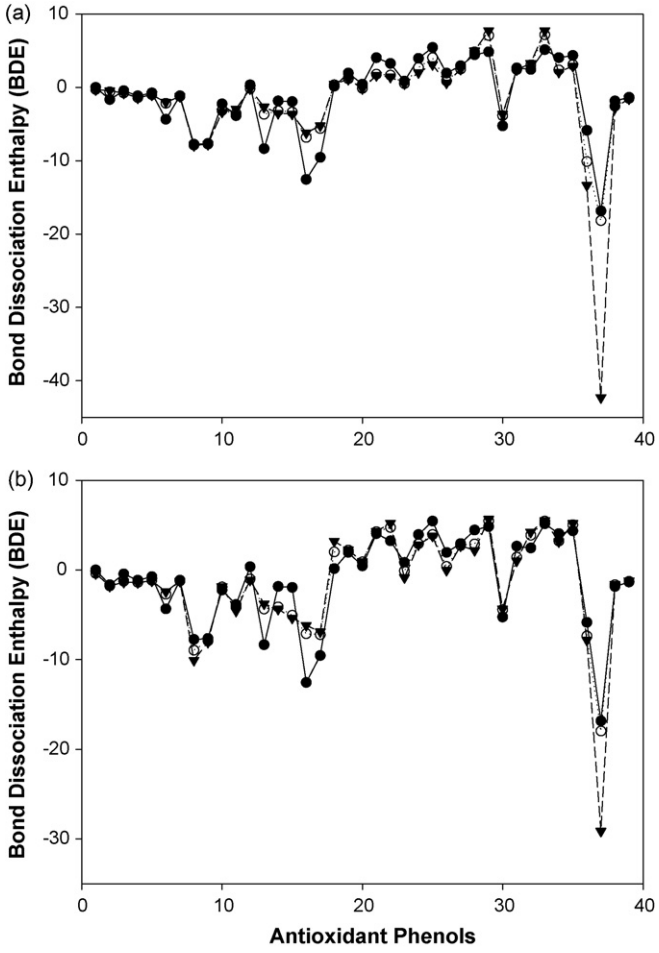


Fig. 2. Plot of the phenolic antioxidant compounds along with the experimental BDE (●), and the predicted BDE of the training (○) and cross-validated testing (▼) set. Compound 37 was identified as outlier and removed from QSPR models using molecular descriptors derived from AM1 (a) and B3LYP/3-21g(d) (b) calculations.

with the B3LYP functional. Considerable evidences in the literature strongly suggest the inability of B3LYP to accurately calculate the values of R–X BDEs as well as its failure to model hydrogen abstraction reactions [51–54]. The B3LYP functional generally suffers from systematic underestimation of the classical barrier height. Guner et al. observed similar unusual effect of performance degradation as a function of increasing basis set size in their study of hydrocarbon pericyclic reactions using the B3LYP functional [55,56].

In order to enhance the predictivity of the QSPR model, closer examination of the results indicated that one of the compound exerted great influence on the overall performance. Particularly, compound **37** was identified by The Unscrambler software package to be an outlier as also illustrated in Fig. 2 where the predicted value of compound **37** is shown to greatly differ from the experimental value. Removal of this outlying sample improved the predictive performance as illustrated in Table 3 by the boost of r from 0.895 to 0.903 for MLR model using descriptors calculated at the B3LYP/3-21g(d) level. Likewise, similar improvements in the predictivity were also observed for descriptors generated at the AM1 level where r rose from 0.814 to 0.874. Analogously, Sun et al. compared the effectiveness of molecular geometries of phenolic antioxidants calculated at the semiempirical AM1 with density functional B3LYP levels and found that AM1 geometry gave comparable accuracy with that derived from B3LYP calculations [57]. Those findings are in line with our results as AM1 shows comparable degree of performance with that of B3LYP/3-21g(d) descriptors.

To further refine the QSPR models, independent variables possessing low regression coefficient values were removed as they exert minimal perturbation to the dependent variable (Fig. 3). Such variables subjected for removal in QSPR model using B3LYP/3-21g(d) derived descriptors included E_{Total}, μ, and q_O, while the remaining descriptors, comprising of IP, EA, q_H, and R_{O-H}, were used for further investigations. This further increased the predictive accuracy to r of 0.917. Likewise, the performance for QSPR model using descriptors generated from AM1 calculations rose to 0.897 upon removal of E_{Total}, μ, and EA.

Since the objective of this study is to calculate the dissociation enthalpies of O–H bonds as a measure of their relative

Table 3
Summary of the predictive performance of the initial and refined MLR models

Descriptors	$r_{Training}$	$r_{LOO\text{-}CV}$	R^2_{LOO-CV}	$R^2_{LOO-CV(adj)}$	F ratio	$RMS_{Training}$	$RMS_{LOO\text{-}CV}$
AM1[a]	0.932	0.814	0.663	0.597	8.697[b]	1.834	4.687
AM1[c]	0.932	0.874	0.764	0.718	13.865[d]	1.604	2.188
AM1[e]	0.923	0.897	0.805	0.787	33.973[f]	1.637	1.974
B3LYP/3-21g(d)[a]	0.950	0.895	0.801	0.763	17.828[b]	1.583	2.743
B3LYP/3-21g(d)[c]	0.942	0.903	0.815	0.780	18.932[d]	1.486	1.912
B3LYP/3-21g(d)[e]	0.939	0.917	0.841	0.827	43.601[f]	1.524	1.777

[a] Initial data set.
[b] Critical F value at the 95% confidence level with 4 and 33 degrees of freedom is 2.659.
[c] Remove compound 37 as outlier.
[d] Critical F value at the 95% confidence level with 7 and 30 degrees of freedom is 2.334.
[e] Exclude three descriptors from QSPR model.
[f] Critical F value at the 95% confidence level with 7 and 31 degrees of freedom is 2.323.

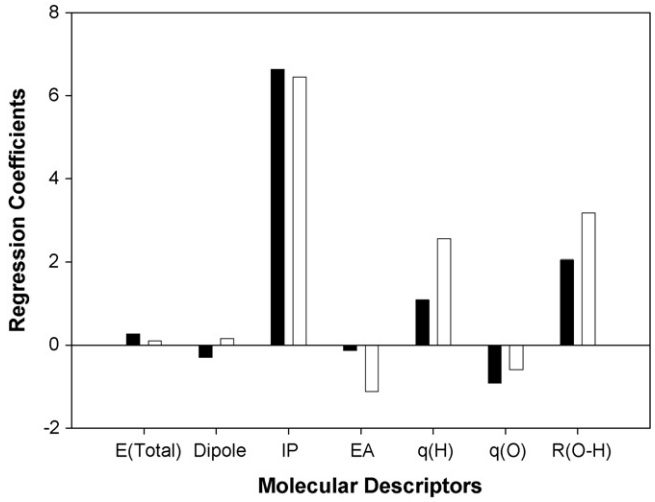


Fig. 3. Regression coefficient values of molecular descriptors derived from AM1 (■) and B3LYP/3-21g(d) (□) calculations.

antioxidative activity, therefore, it is worthy to observe whether the lengths of O–H bonds are correlated with O–H BDEs or not. Table 4 summarizes the O–H bond lengths of the phenolic antioxidants geometrically optimized at various theoretical levels. Results indicated that there is a general positive correlation of O–H bond length with their respective O–H BDE where AM1 gave the best performance of the four theoretical levels tested. Particularly, phenolic antioxidant with longer O–H bond indicates that there is weaker attraction of the hydrogen atom toward the oxygen atom. As such the hydroxyl group easily loses the hydrogen atom to ROS radicals as also indicated by its lower BDE values. It is also interesting to note that R_{O-H} was retained as molecular descriptor in both QSPR models: (1) model using AM1 descriptor and (2) model using B3LYP/3-21g(d) descriptor.

Next, the most suitable multivariate regression approach for this endeavor was investigated using two other methods besides MLR, which included PLS and SVM. It was observed that the use of molecular descriptors derived from calculation at the B3LYP/3-21g(d) level yield good predictive performance with r in excess of 0.9. Particularly, the use of typical regression methods, such as MLR and PLS, resulted in r of 0.917 and 0.921, respectively. On the other hand, models using descriptors generated at the semiempirical AM1 level provided relatively good precision but to a lesser extent than those calculated at the density functional level. This is illustrated by r of 0.897 and 0.888 when regression analysis was performed using MLR and PLS, respectively.

SVM is a supervised learning method capable of solving many complex classification problems. SVM has demonstrated success in many perplexing biological and chemical problems, such as the classification of protein folds [58] and secondary structures [59], prediction of DNA splice junction sites [60], and classification of the activity/inactivity of potential lead compounds [61]. In this

Table 4
Summary of O–H bond lengths (R_{O-H}) in phenolic antioxidants geometrically optimized at different levels of theory

Cpd. no.	AM1	HF/3-21g(d)	B3LYP/3-21g(d)	B3LYP/6-31g(d)	ΔBDE (kcal/mol)
16	0.9672	0.9639	0.9919	0.9691	−12.55
17	0.9675	0.9639	0.9921	0.9693	−9.55
13	0.9678	0.9640	0.9920	0.9694	−8.35
8	0.9655	0.9660	0.9859	0.9645	−7.75
9	0.9657	0.9660	0.9873	0.9644	−7.65
36	0.9660	0.9606	0.9895	0.9679	−5.85
30	0.9677	0.9639	0.9920	0.9693	−5.25
6	0.9678	0.9621	0.9904	0.9687	−4.35
11	0.9688	0.9640	0.9923	0.9697	−3.85
10	0.9683	0.9642	0.9924	0.9699	−2.25
15	0.9679	0.9640	0.9920	0.9695	−1.95
14	0.9680	0.9641	0.9920	0.9695	−1.85
38	0.9683	0.9644	0.9925	0.9701	−1.85
2	0.9681	0.9638	0.9916	0.9694	−1.65
39	0.9686	0.9645	0.9925	0.9702	−1.35
4	0.9680	0.9642	0.9923	0.9697	−1.15
7	0.9680	0.9642	0.9924	0.9697	−1.15
5	0.9679	0.9642	0.9924	0.9698	−0.75
3	0.9680	0.9641	0.9922	0.9697	−0.45
1	0.9680	0.9642	0.9923	0.9698	0
18	0.9689	0.9641	0.9921	0.9699	0.15
12	0.9683	0.9641	0.9921	0.9695	0.35
20	0.9684	0.9643	0.9922	0.9698	0.45
23	0.9686	0.9643	0.9922	0.9698	0.85
19	0.9684	0.9643	0.9922	0.9699	1.95
26	0.9682	0.9642	0.9924	0.9698	1.95
32	0.9684	0.9643	0.9923	0.9700	2.45
31	0.9687	0.9645	0.9927	0.9704	2.65
27	0.9688	0.9644	0.9926	0.9703	2.95
22	0.9691	0.9645	0.9922	0.9701	3.25
24	0.9686	0.9642	0.9922	0.9698	3.95
21	0.9688	0.9645	0.9922	0.9701	4.05
34	0.9686	0.9644	0.9923	0.9701	4.05
35	0.9689	0.9645	0.9925	0.9703	4.35
28	0.9689	0.9642	0.9921	0.9700	4.45
29	0.9697	0.9647	0.9925	0.9705	4.85
33	0.9697	0.9645	0.9925	0.9703	5.15
25	0.9691	0.9645	0.9924	0.9701	5.45
r with ΔBDE	0.778	0.181	0.503	0.596	

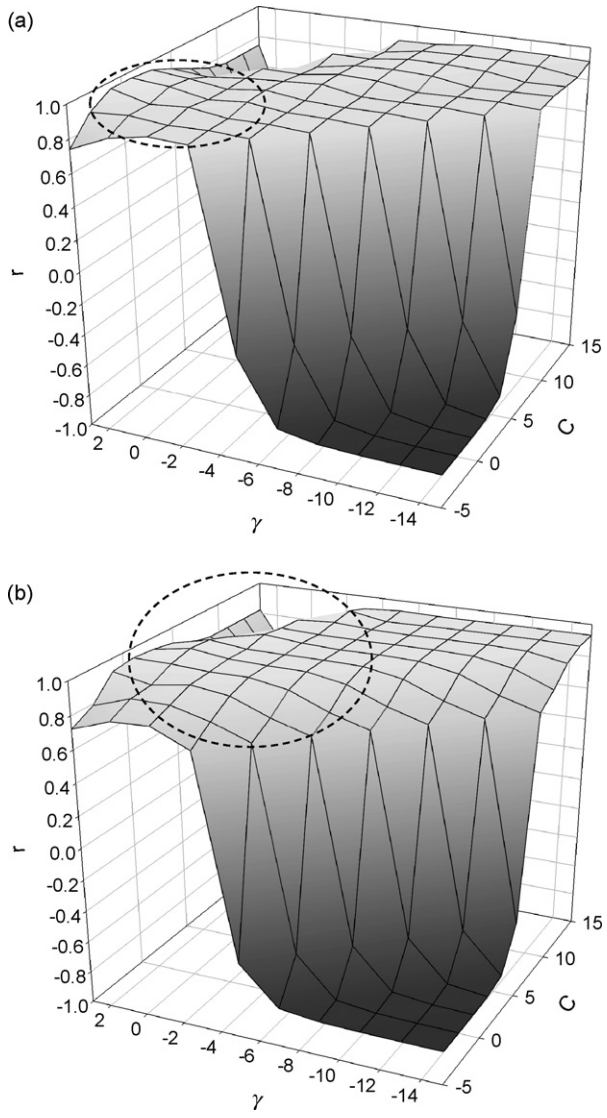


Fig. 4. Three-dimensional mesh plot of the coarse grid search of the SVM parameters for QSPR models using AM1 (a) and B3LYP/3-21g(d) (b) descriptors. The correlation coefficient (r) was plotted as a function of the SVM parameters, C and γ. Regions giving rise to good prediction accuracy is shown in the circled area.

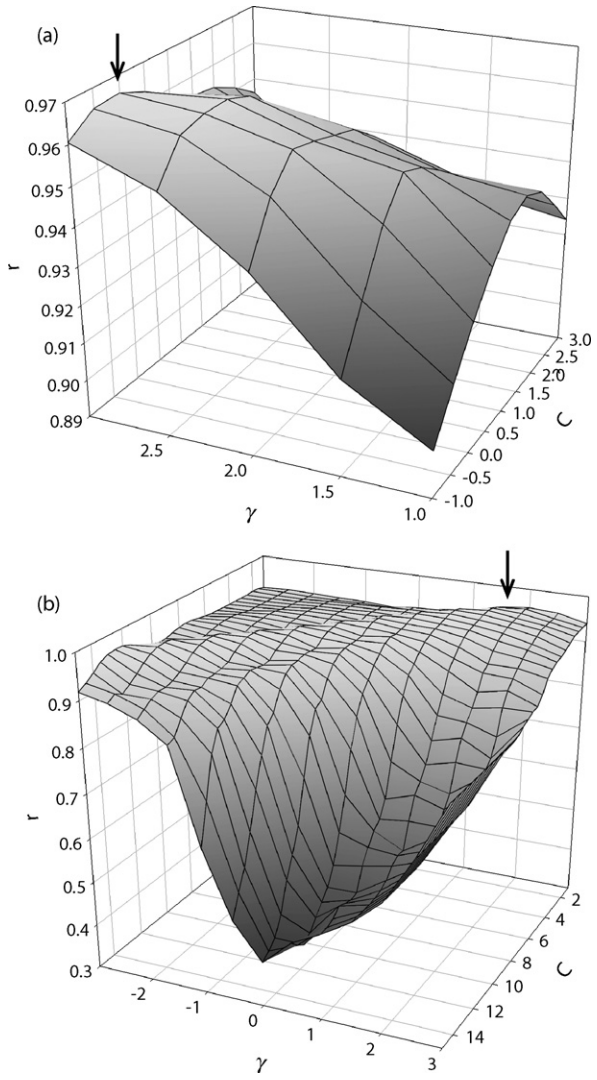


Fig. 5. Three-dimensional mesh plot of the local grid search of the SVM parameters for QSPR models using AM1 (a) and B3LYP/3-21g(d) (b) descriptors. The correlation coefficient (r) is plotted as a function of the SVM parameters, C and γ. Regions giving rise to good prediction accuracy is indicated by the arrow.

study, SVM was employed for correlating the physicochemical properties of the antioxidant phenols with their BDEs. Since SVM is a linear learning machine, non-linear data must be represented in a way that allows the algorithm to classify them in the new feature space [36]. This is achieved by first converting the descriptor matrix onto a richer feature space while conserving the original information via radial basis function kernel transformation.

To achieve an optimal set of parameters for the construction of robust SVM models, two sequential stepwise parameter searches were performed. An initial global grid search for the optimal SVM parameters, C and γ, was carried out to identify regions affording good accuracy. In this parameter search, the exponential n value of 2^n for C was varied from -5 to 15 by increments of 2 while γ was varied from -15 to 3 by increments of 2. As illustrated in Fig. 4, the optimal region was identified as indicated within the circled area. The selected region was then subjected to a more thorough search in order to locate the optimal set of parameters. Results from Fig. 5 indicated that the optimal value of C and γ was 2^0 and 2^3, respectively, for models using descriptors generated at the AM1

Table 5
Summary of the predictive performance of the various QSPR models

Descriptors	Multivariate method	$r_{Training}$	$r_{LOO\text{-}CV}$	R^2_{LOO-CV}	$R^2_{LOO-CV(adj)}$	F ratio[a]	$RMS_{Training}$	$RMS_{LOO\text{-}CV}$
B3LYP/3-21g(d)	MLR	0.939	0.917	0.841	0.827	43.601	1.524	1.777
B3LYP/3-21g(d)	PLS	0.933	0.921	0.848	0.835	46.113	2.016	2.175
B3LYP/3-21g(d)	SVM	0.987	0.968	0.937	0.931	122.752	0.752	1.122
AM1	MLR	0.929	0.897	0.805	0.787	33.973	1.637	1.974
AM1	PLS	0.915	0.888	0.789	0.770	30.765	2.243	2.491
AM1	SVM	0.984	0.966	0.933	0.927	115.172	0.813	1.247

[a] Critical F value at the 95% confidence level with 4 and 33 degrees of freedom is 2.659.

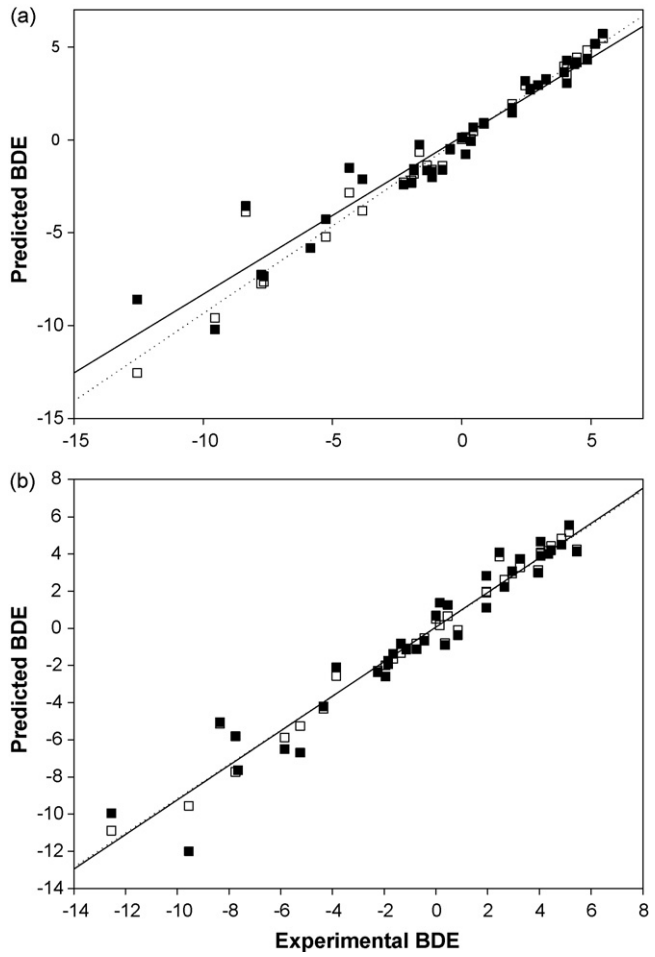


Fig. 6. Plot of the predicted BDE versus the experimental BDE for the training set (□; regression line is represented as dotted line) and testing set (■; regression line is represented as solid line).

level. Furthermore, it can also be seen that the optimal C and γ was $2^{2.5}$ and $2^{1.5}$, respectively, for models using B3LYP/3-21g(d) descriptors.

A predictive model was then constructed using the empirically derived parameters. The experimental BDEs of the antioxidant phenols were well correlated with the predicted BDEs as presented in Fig. 6. From this plot it can be seen that the SVM models using descriptors generated from the semiempirical AM1 or the density functional B3LYP method in combination with the 3-21g(d) basis set afforded similar level of high predictivity as judged by r of 0.966 and 0.968, respectively (Table 5). Furthermore, the observed F values for all models are much greater (30.765–122.752) than the critical F value (2.659) at the 95% confidence level with 4 and 33 degrees of freedom. Therefore, we can reject the null hypothesis and conclude that the independent variables correlate well with the O–H BDE values.

4. Conclusion

This study proposes the combined use of quantum chemical descriptors with supervised learning methods for the development of robust QSPR models of antioxidant phenol BDEs. The substituent effects arising from variation of the functional groups at various positions of the phenol ring could be accounted for by the QSPR models described in this study. The predicted BDEs of antioxidant phenols were found to be in good agreement with the experimental values. Of the three multivariate regression methods used in this study, SVM was found to outperform the traditional regression methods such as MLR and PLS. The methodology proposed in this study facilitates rapid evaluation of antioxidant properties using readily available quantum chemical descriptors that are inexpensive to calculate in combination with robust modeling methods such as SVM.

Acknowledgements

This research is financially supported in part by the Young Research Scholar grant (no. MRG5080450) from The Thailand Research Fund to C.N. and the annual budget of Mahidol University (B.E. 2551).

References

[1] T.E. Andreoli, Free radicals and oxidative stress, Am. J. Med. 108 (2000) 650–651.
[2] P.G. Winyard, C.J. Moody, C. Jacob, Oxidative activation of antioxidant defence, Trends Biochem. Sci. 30 (2005) 453–461.
[3] J.M. McCord, The evolution of free radicals and oxidative stress, Am. J. Med. 108 (2000) 652–659.
[4] N.M. Katarina, Mechanistic studies of phenolic antioxidants in reaction with nitrogen- and oxygen-centered radicals, J. Mol. Struct. (Theochem.) 818 (2007) 141–150.
[5] K.M. Nikolic, Theoretical study of phenolic antioxidants properties in reaction with oxygen-centered radicals, J. Mol. Struct. (Theochem.) 774 (2006) 95–105.
[6] C. Giacomelli, S. Miranda Fda, N.S. Goncalves, A. Spinelli, Antioxidant activity of phenolic and related compounds: a density functional theory study on the O–H bond dissociation enthalpy, Redox Rep. 9 (2004) 263–269.
[7] N. Nenadis, L.F. Wang, M.Z. Tsimidou, H.Y. Zhang, Radical scavenging potential of phenolic compounds encountered in *O. europaea* products as indicated by calculation of bond dissociation enthalpy and ionization potential values, J. Agric. Food Chem. 53 (2005) 295–299.
[8] K. Tanaka, S. Sakai, S. Tomiyama, T. Nishiyama, F. Yamada, Molecular orbital approach to antioxidant mechanisms of phenols by an ab initio study, Bull. Chem. Soc. Jpn. 64 (1991) 2677–2680.
[9] S. Tomiyama, S. Sakai, T. Nishiyama, F. Yamada, Factors influencing the antioxidant activities of phenols by an ab initio study, Bull. Chem. Soc. Jpn. 66 (1993) 299–304.
[10] D. Shanks, H. Frisell, H. Ottosson, L. Engman, Design principles for alpha-tocopherol analogues, Org. Biomol. Chem. 4 (2006) 846–852.
[11] Z. Velkov, E. Balabanova, A. Tadjer, Radical scavenging activity prediction of *o*-coumaric acid thioamide, J. Mol. Struct. (Theochem.) 821 (2007) 133–138.
[12] J. Lind, X. Shen, T.E. Eriksen, G. Merenyi, The one-electron reduction potential of 4-substituted phenoxyl radicals in water, J. Am. Chem. Soc. 112 (1990) 479–482.
[13] P. Mulder, H.G. Korth, D.A. Pratt, G.A. DiLabio, L. Valgimigli, G.F. Pedulli, K.U. Ingold, Critical re-evaluation of the O–H bond dissociation enthalpy in phenol, J. Phys. Chem. A 109 (2005) 2647–2655.
[14] J.A. Pople, M. Head-Gordon, K. Raghavachari, Quadratic configuration interaction. A general technique for determining electron correlation energies, J. Chem. Phys. 87 (1987) 5968–5975.
[15] K. Raghavachari, G.W. Trucks, J.A. Pople, M. Head-Gordon, A fifth-order perturbation comparison of electron correlation theories, Chem. Phys. Lett. 157 (1989) 479–483.
[16] J.W. Ochterski, G.A. Petersson, J.A. Montgomery, A complete basis set model chemistry. V. Extensions to six or more heavy atoms, J. Chem. Phys. 104 (1996) 2598–2619.
[17] G.A. Petersson, T.G. Tensfeldt, J.A. Montgomery, A complete basis set model chemistry. III. The complete basis set-quadratic configuration interaction family of methods, J. Chem. Phys. 94 (1991) 6091–6101.
[18] M.R. Hoffmann, H.F. Schaefer Iii, The treatment of triple excitations within the coupled cluster description of molecular electronic structure, J. Chem. Phys. 83 (1985) 703–712.
[19] L.A. Curtiss, K. Raghavachari, P.C. Redfern, V. Rassolov, J.A. Pople, Gaussian-3 (G3) theory for molecules containing first and second-row atoms, J. Chem. Phys. 109 (1998) 7764–7776.
[20] L.A. Curtiss, P.C. Redfern, K. Raghavachari, Gaussian-4 theory, J. Chem. Phys. 126 (2007) 084108–084112.
[21] K.A. Peterson, T.H. Dunning Jr., Benchmark calculations with correlated molecular wave functions. VIII. Bond energies and equilibrium geometries of the CH_n and C_2H_n (n = 1–4) series, J. Chem. Phys. 106 (1997) 4119–4140.
[22] M.R. Nyden, G.A. Petersson, Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions, J. Chem. Phys. 75 (1981) 1843–1862.
[23] G.A. Petersson, M.A. Al-Laham, A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms, J. Chem. Phys. 94 (1991) 6081–6090.

[24] J.A. Montgomery, J.W. Ochterski, G.A. Petersson, A complete basis set model chemistry. IV. An improved atomic pair natural orbital method, J. Chem. Phys. 101 (1994) 5900–5909.
[25] J.L. Lewin, C.J. Cramer, Rapid quantum mechanical models for the computational estimation of C–H bond dissociation energies as a measure of metabolic stability, Mol. Pharm. 1 (2004) 128–135.
[26] F.G. Bordwell, J.-P. Cheng, Substituent effects on the stabilities of phenoxyl radicals and the acidities of phenoxyl radical cations, J. Am. Chem. Soc. 113 (1991) 1736–1743.
[27] R. Dennington II, T. Keith, J. Millam, K. Eppinnett, W.L. Hovell, R. Gilliland, GaussView, version 3.09, Semichem, Inc., Shawnee Mission, KS, 2003.
[28] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.
[29] T. Koopmans, Uber die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms, Physica 1 (1934) 104–113.
[30] Camo Process AS, The Unscrambler, version 9.5, Norway, 2006.
[31] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17.
[32] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab. 58 (2001) 109–130.
[33] D.M. Haaland, E.V. Thomas, Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information, Anal. Chem. 60 (1988) 1193–1202.
[34] J.C. Platt, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, Microsoft Research, Technical Report MSR-TR-98-14, 1998.
[35] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed., Morgan Kaufmann, San Francisco, 2005.
[36] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, Cambridge, 2000.
[37] S. Abe, Support Vector Machines for Pattern Classification, Springer-Verlag London Limited, New York, 2005.
[38] Y. Sakiyama, H. Yuki, T. Moriya, K. Hattori, M. Suzuki, K. Shimada, T. Honma, Predicting human liver microsomal stability with machine learning techniques, J. Mol. Graph. Model. 26 (2008) 907–915.
[39] H.H. Lin, L.Y. Han, C.W. Yap, Y. Xue, X.H. Liu, F. Zhu, Y.Z. Chen, Prediction of factor Xa inhibitors by machine learning methods, J. Mol. Graph. Model. 26 (2007) 505–518.
[40] G. Liang, Z. Li, Scores of generalized base properties for quantitative sequence-activity modelings for *E. coli* promoters based on support vector machine, J. Mol. Graph. Model. 26 (2007) 269–281.
[41] R. Herbrich, Learning Kernel Classifiers: Theory and Algorithms, The MIT Press, Cambridge, 2002.
[42] B. Schölkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, The MIT Press, Cambridge, 2002.
[43] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222.
[44] L. Wang, Support Vector Machines: Theory and Applications, Springer-Verlag, Berlin, 2005.
[45] S. Abe, Support Vector Machines for Pattern Classification, Springer-Verlag, London, 2005.
[46] N. Chen, W. Lu, J. Yang, G. Li, Support Vector Machine in Chemistry, World Scientific Publishing Co. Pte. Ltd., Singapore, 2004.
[47] C.-W. Hsu, C.-C. Chang, C.-J. Lin, A Practical Guide to Support Vector Classification, Department of Computer Science, National Taiwan University, Technical Report, 2007.
[48] C. Staelin, Parameter Selection for Support Vector Machines, Hewlett-Packard Company, Technical Report HPL-2002-354 (R.1), 2003.
[49] C. Nantasenamat, C. Isarankura-Na-Ayudhya, N. Tansila, T. Naenna, V. Prachayasittikul, Prediction of GFP spectral properties using artificial neural network, J. Comput. Chem. 28 (2007) 1275–1289.
[50] C. Nantasenamat, C. Isarankura-Na-Ayudhya, T. Naenna, V. Prachayasittikul, Quantitative structure-imprinting factor relationship of molecularly imprinted polymers, Biosens. Bioelectron. 22 (2007) 3309–3317.
[51] B.J. Lynch, D.G. Truhlar, How well can hybrid density functional methods predict transition state geometries and barrier heights? J. Phys. Chem. A 105 (2001) 2936–2941.
[52] J.K. Kang, C.B. Musgrave, Prediction of transition state barriers and enthalpies of reaction by a new hybrid density-functional approximation, J. Chem. Phys. 115 (2001) 11040–11051.
[53] H. Basch, S. Hoz, *Ab initio* study of hydrogen abstraction reactions, J. Phys. Chem. A 101 (1997) 4416–4431.
[54] M.L. Coote, Reliable theoretical procedures for the calculation of electronic-structure information in hydrogen abstraction reactions, J. Phys. Chem. A 108 (2004) 3865–3872.
[55] V.A. Guner, K.S. Khuong, K.N. Houk, A. Chuma, P. Pulay, The performance of the Handy/Cohen functionals, OLYP and O3LYP, for the computation of hydrocarbon pericyclic reaction activation barriers, J. Phys. Chem. A 108 (2004) 2959–2965.
[56] V. Guner, K.S. Khuong, A.G. Leach, P.S. Lee, M.D. Bartberger, K.N. Houk, A standard set of pericyclic reactions of hydrocarbons for the benchmarking of computational methods: the performance of ab initio, density functional, CASSCF, CASPT2, and CBS-QB3 methods for the prediction of activation barriers, reaction energetics, and transition state geometries, J. Phys. Chem. A 107 (2003) 11445–11459.
[57] Y. Sun, D. Chen, C. Liu, Evaluation of the effectiveness of AM1 geometry used in calculating O–H bond dissociation enthalpy, J. Mol. Struct. (Theochem.) 618 (2002) 181–189.
[58] M.T. Shamim, M. Anwaruddin, H.A. Nagarajaram, Support vector machine-based classification of protein folds using the structural properties of amino acid residues and amino acid residue pairs, Bioinformatics 23 (2007) 3320–3327.
[59] J.J. Ward, L.J. McGuffin, B.F. Buxton, D.T. Jones, Secondary structure prediction with support vector machines, Bioinformatics 19 (2003) 1650–1655.
[60] C. Nantasenamat, T. Naenna, C. Isarankura-Na-Ayudhya, V. Prachayasittikul, Recognition of DNA splice junction via machine learning approaches, EXCLI J. 4 (2005) 114–129.
[61] J.C. Saeh, P.D. Lyne, B.K. Takasaki, D.A. Cosgrove, Lead hopping using SVM and 3D pharmacophore fingerprints, J. Am. Chem. Soc. 45 (2005) 1122–1133.

European Journal of Medicinal Chemistry 44 (2008) 1664–1673

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Modeling the activity of furin inhibitors using artificial neural network

Apilak Worachartcheewan [a], Chanin Nantasenamat [a], Thanakorn Naenna [b], Chartchalerm Isarankura-Na-Ayudhya [a,*], Virapong Prachayasittikul [a,*]

[a] Department of Clinical Microbiology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[b] Department of Industrial Engineering, Faculty of Engineering, Mahidol University, Nakhon Pathom 73170, Thailand

ARTICLE INFO

Article history:
Received 22 May 2008
Received in revised form 24 July 2008
Accepted 8 September 2008
Available online 1 October 2008

Keywords:
Furin
Anthrax
QSAR
Neural network

ABSTRACT

Quantitative structure–activity relationship (QSAR) models were constructed for predicting the inhibition of furin-dependent processing of anthrax protective antigen of substituted guanidinylated aryl 2,5-dideoxystreptamines. Molecular descriptors calculated by E-Dragon and RECON were subjected to variable reduction using the Unsupervised Forward Selection (UFS) algorithm. The variables were then used as input for QSAR model generation using partial least squares and back-propagation neural network. Prediction was performed via a two-step approach: (i) perform classification to determine whether the molecule is active or inactive, (ii) develop a QSAR regression model of active molecules. Both classification and regression models yielded good results with RECON providing higher accuracy than that of E-DRAGON descriptors. The performance of the regression model using E-Dragon and RECON descriptors provided a correlation coefficient of 0.807 and 0.923 and root mean square error of 0.666 and 0.304, respectively. Interestingly, it was observed that appropriate representations of the protonation states of the molecules were crucial for good prediction performance, which coincides with the fact that the inhibitors interact with furin via electrostatic forces. The results provide good prospect of using the proposed QSAR models for the rational design of novel therapeutic furin inhibitors toward anthrax and furin-dependent diseases.

© 2008 Elsevier Masson SAS. All rights reserved.

1. Introduction

Anthrax is an infectious disease caused by *Bacillus anthracis*, a rod-shaped gram-positive bacterium. Infection occurs when endospores are introduced into the host cell via abrasion, inhalation or ingestion. Afterwards, they are phagocytosed by macrophages and migrate to regional lymph nodes. They rapidly germinate and become vegetative bacteria which are then released from the macrophages, followed by their multiplication in the lymphatic system, and gaining access to the blood circulation system to cause massive septicemia and toxemia. Anthrax toxins, excreted from *B. anthracis*, are comprised of three proteins, anthrax protective antigen (PA, 83 kDa), edema factor (EF, 89 kDa), and lethal factor (LF, 90 kDa) [1–3].

Activation of anthrax toxin takes place when furin cleaves the 83 kDa PA into two subunits: a receptor-bound 63 kDa subunit and a free 20 kDa fragment [4]. Oligomerization of PA63 mediates the endocytosis of EF and LF into the cytosol which eventually leads to a cataclysmic chain of events that give rise to pathological conditions and culminating in cell death [5,6]. Therefore, furin is a lucrative target for the development of therapeutic agents.

Early approaches toward the development of furin inhibitors are mostly based on proteins and peptides, for example, α_1-antitrypsin (α_1-PDX) [7], polyarginine peptides [8], and peptidyl boronic acids. However, they inherently possess some shortcomings that may limit their therapeutic potentials; this includes their rather large molecular weight [9], low turnover [10], and potential cytotoxicity [11]. Therefore, this has prompted the development of small molecule-based furin inhibitors to solve the drawbacks of proteins/peptides. This new family of compounds is based on guanidinylated aryl 2,5-dideoxystreptamines (GADDs) which was tested against furin and was shown to effectively protect cells from furin-dependent processing of PA [12].

In the present work, artificial neural network was used for the construction of a quantitative structure–activity relationship (QSAR) model of the inhibitory activity of furin inhibitors derived from GADDs. Molecular descriptors were calculated using E-Dragon and RECON softwares, which served as input for QSAR model generation using partial least squares and back-propagation neural network. A two-step approach was carried out as follows: (i) classify whether the molecule is active or inactive and (ii) quantitatively predict the inhibition activity of the active molecules. It was discovered that by taking account of the protonation state of

* Corresponding authors. Tel.: +66 2 418 0227; fax: +66 2 412 4110.
E-mail addresses: mtcis@mahidol.ac.th (C. Isarankura-Na-Ayudhya), mtvpr@mahidol.ac.th (V. Prachayasittikul).

0223-5234/$ – see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.09.028

GADDs, which is important in interaction and inhibition of furin, provided good predictive performance as observed from improvements in the correlation coefficient of predicted and experimental pK_i from 0.759 to 0.807 when using E-Dragon descriptors. Furthermore, charge-based descriptors generated from RECON provided stellar performance as illustrated by the correlation coefficient of 0.923. The results suggested good prospect of using the proposed QSAR models as guides for designing novel compounds with robust inhibition activity.

2. Methodology

2.1. Data set

The data set used in this study was obtained from the work of Jiao et al. [12]. It is comprised of 18 furin inhibitors based on guanidinylated aryl 2,5-dideoxystreptamine derivatives. The inhibition constant (K_i) value, which is a measure of the ability to inhibit furin, was converted to pK_i by taking the negative logarithm to base 10 of K_i ($-\log K_i$). The molecular structures and their activities are shown in Table 1. Of the 18 furin inhibitors, six (**1h**, **1j**, **1k**, **1l**, **1m**, and **1r**) were discarded from the data set as they exhibit little or no inhibition against furin as judged from $K_i > 100$ μM. These molecules were labeled as inactive compounds while the remaining was designated as active compounds (Table 1).

2.2. Descriptor generation

2.2.1. Construction of molecular structure

The two-dimensional structure of furin inhibitors was drawn with ChemAxon's Marvin [13] and saved as SMILES notations which serves as input for the generation of molecular descriptors by E-Dragon, version 1.0 [14] and RECON version 5.5 [15].

2.2.2. Determination of protonation state

The appropriate protonation state of the molecules in experimental buffer at physiological pH 7.5 was identified with the Sketch module of the Marvin software package. The proposed ionization states were then introduced into the structures of furin inhibitors for generation of relevant molecular descriptors.

2.2.3. Calculation of molecular descriptors

The online version of Dragon as available on the Virtual Computational Chemistry Laboratory (VCCLAB) [16] as E-Dragon was employed for calculating 1666 molecular descriptors spanning 20 categories: Constitutional descriptors, Walk and path counts, Information indices, Edge adjacency indices, Topological charge indicates, Randic molecular profiles, RDF descriptors, WHIM descriptors, Functional group counts, Charge descriptors, Topological descriptors, Connectivity indices, 2D autocorrelation, Burden eigenvalues, Eigenvalue based indices, Geometrical descriptors, 3D MoRSE descriptors, GETAWAY descriptors, Atom-centered fragments, and Molecular properties. The definitions of these descriptors are reviewed by Todeschini et al. [17]. SMILES notations of the chemical structures served as input data for subsequent conversion to three-dimensional structures using the built-in software option, CORINA.

Similarly, SMILES notation was used as input for the generation of 248 transferable atom equivalent (TAE) descriptors using RECON. The TAE methodology was developed by Breneman and Rhem [18] based on Bader's theory of atom in molecules [19]. The algorithm is based on building a library of precomputed atomic fragments by *ab initio* calculations. The molecular properties of the compounds of interests are then reconstructed from their atomic constituents. The TAE molecular descriptors are comprised of 11 electronic surface properties [20,21] comprising Surface integral of electrostatic potential (SIEP), Electrostatic potential (EP), Electron density gradient normal to 0.002 e/au^3 electron density isosurface (Del(Rho)·N), Electronic kinetic energy density (G), Electronic kinetic energy density (K), Gradient of K electronic kinetic energy density normal to surface (Del(K)·N), Gradient of G electronic kinetic energy density normal to surface (Del(G)·N), Fukui F^+ function scalar value (Fuk), Laplacian of the electron density (Lapl), bare nuclear potential (BNP), and local average ionization potential (PIP).

2.3. Descriptor reduction

The calculated descriptors obtained from each of the software were standardized using Weka, version 3.4.5 [22] so as to adjust all of the descriptors to the same scale with zero mean and unit variance according to the following equation:

$$x_{ij}^{stn} = \frac{x_{ij} - \bar{x}_j}{\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2/N} \quad (1)$$

where x_{in}^{stn} represents the standardized value, x_{ij} represents the value of each sample, $\bar{x}_j$ represents the mean of each descriptor, and N represents the sample size of the data set.

Descriptors with constant value were discarded using the RemoveUseless feature of Weka. In order to further improve the predictive performance as well as reduce the computational time, multi-collinear and redundant descriptors were discarded using the Unsupervised Forward Selection algorithm of UFS, version 1.8 [23] as described previously [24].

2.4. Generation of training and testing sets

The data set of furin inhibitors was divided into two sets comprising training and testing sets by the leave-one-out cross-validation (LOO-CV) approach in which one sample of the data set was left out as the testing set while the remaining data samples were used as the training set. The training sets were employed to build a predictive model and predictions were made on the testing set. The process was repeated iteratively until all of the data samples were used as the testing set.

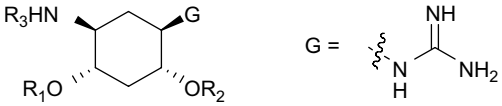
2.5. Partial least squares analysis

Partial Least Squares models were performed using the PLS1 algorithm of The Unscrambler 9.6 software package [25]. Molecular descriptors derived from E-Dragon and RECON were used as independent variables while the activity/inactivity served as the response variable. The input variables were pre-processed by autoscaling to zero mean and unit variance as previously described by Eq. (1). The number of descriptors was reduced to a smaller number of latent variables called PLS components (PC) that is able to account for the core information of the original data set. The optimal number of PCs was determined according to the method of Haaland and Thomas [26] from a plot of PCs versus the mean squared error (MSE) as calculated according to the following equation:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{n} \quad (2)$$

where p_i is the predicted output, a_i is the actual output, and n is the number of compounds presented in the data set. In this study, PLS is used in combination with cross-validation as an indicator of predictive performance. Therefore, the PLS model was calculated using LOO-CV.

Table 1
Structures of furin inhibitors and their inhibition constants against furin[a]



1

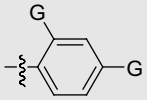
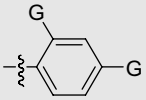
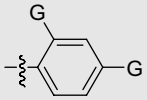
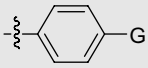
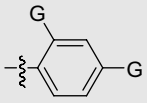
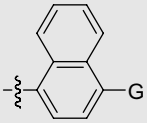
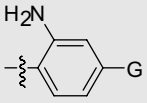
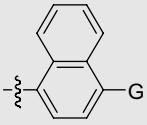
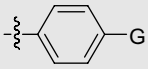
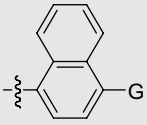
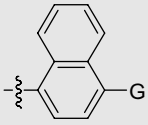
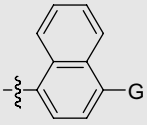
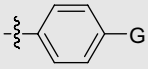
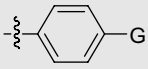
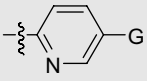
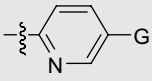
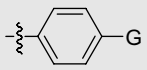
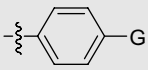
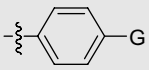
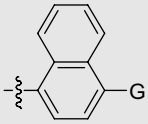
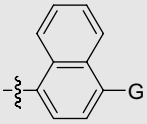
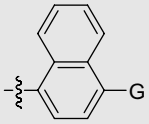
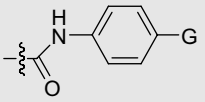
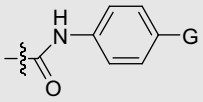
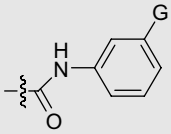
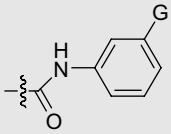
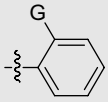
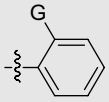
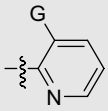
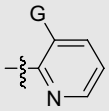
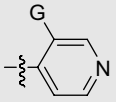
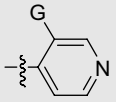
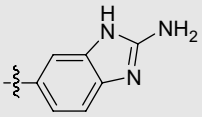
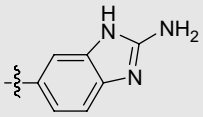
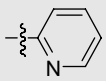
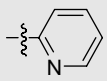
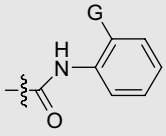
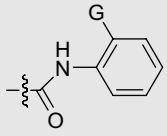
No.	Compound	R_1	R_2	R_3	K_i (μM)	pK_i [b]
A-1	**1a**				0.169	0.77
A-2	**1b**				0.089	1.05
A-3	**1c**				0.404	0.39
A-4	**1d**				0.022	1.66
A-5	**1e**				0.006	2.22
A-6	**1f**				0.069	1.16
A-7	**1g**				0.012	1.92
A-8	**1i**				0.042	1.38
A-9	**1n**				0.046	1.34
A-10	**1o**				0.423	0.37

Table 1 (*continued*)

No.	Compound	R_1	R_2	R_3	K_i (μM)	pK_i [b]
A-11	**1p**				0.812	0.09
A-12	**1q**				1.768	−0.25
I-13	**1h**				>100	–
I-14	**1j**				>100	–
I-15	**1k**				>100	–
I-16	**1l**				>100	–
I-17	**1m**				>100	–
I-18	**1r**				>100	–

[a] All data presented in the table were obtained from Ref. [12] and **1**–**12** were active compounds and **13**–**18** are inactive compounds.
[b] K_i was converted to pK_i by negative logarithm to base 10 of K_i ($-\log K_i$).

2.6. Artificial neural network calculations and model evaluation

Artificial neural network (ANN) is an interconnected feed-forward network of modeling the brain and is a popular tool in function learning due to ability to learn rather complicated function. The methods of ANN were described previously [24,27,28]. Briefly, the network has three interconnected layers: input layer, hidden layer, and output layer [29]. The molecular descriptors are sent to the input layer where they are subsequently passed on to the nodes of the hidden layer for further processing. The signals are then relayed onto the output layer. The connections between the nodes of each layer are assigned by randomized weight value. As such, the average of 10 runs of network training was used. Prior to actual prediction, ANN parameters were optimized by an empirical trial-and-error approach. These parameters included the number of descriptors, the number of nodes in the hidden layer, the learning epoch size, the learning rate (η) and the momentum (μ). Training of the neural network was performed with Weka, version 3.4.5.

Correlation coefficient (r) was used as a relative measure of the predictive performance, which is defined as the degree of correlation between the predicted and experimental values. Likewise, root mean square error (RMS) was used as a measure of predictive error. RMS was calculated according to the following equation:

$$\mathrm{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(p_i - a_i)^2}{n}} \tag{3}$$

where p_i is the predicted output, a_i is the actual output, and n is the number of compounds in the data set.

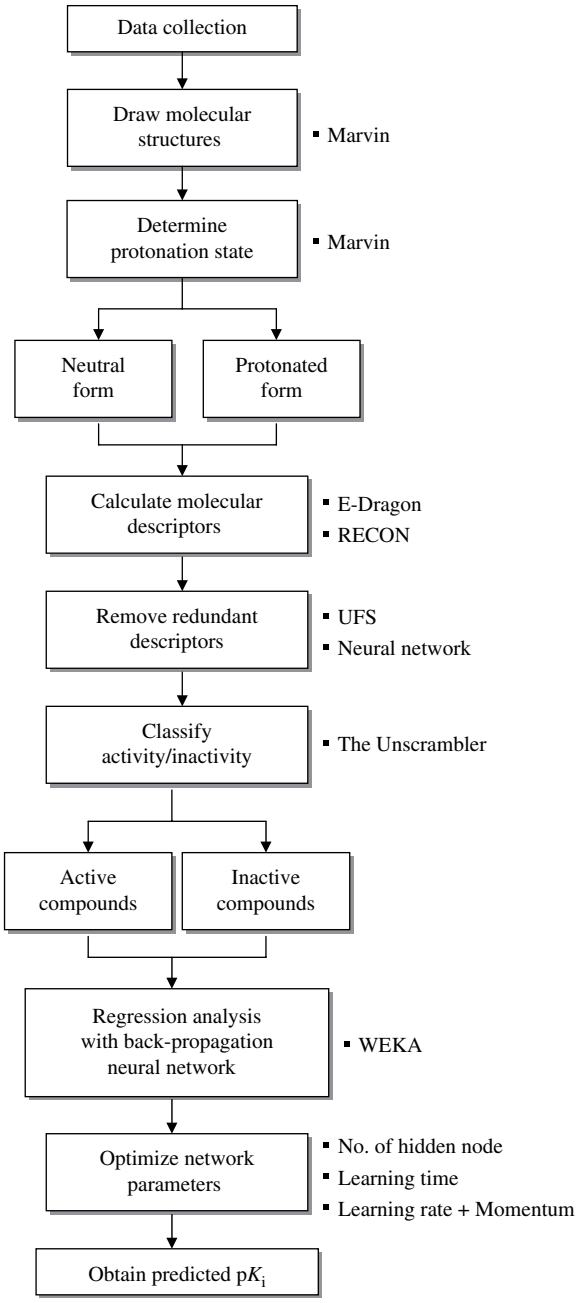


Fig. 1. Schematic representation of the prediction procedure.

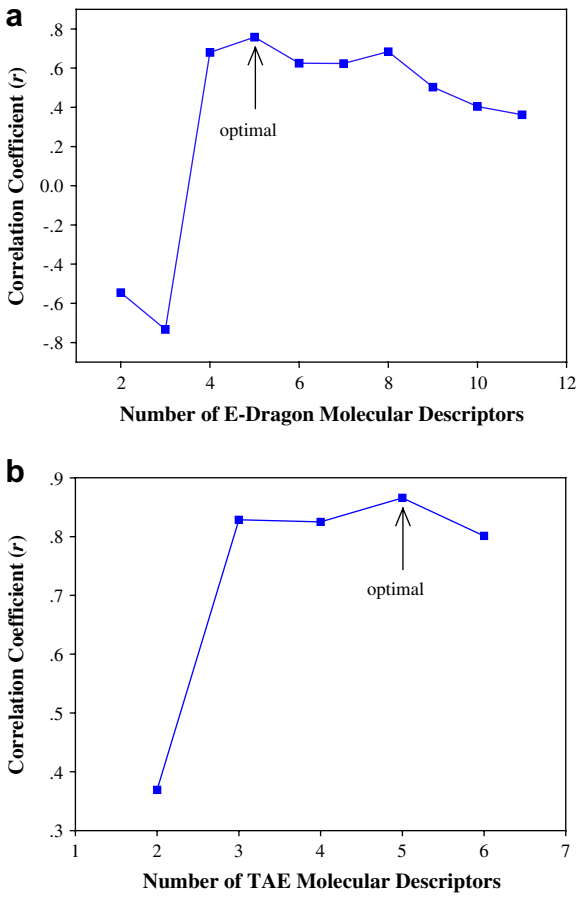


Fig. 2. Determining the optimal number of molecular descriptors for E-Dragon (a) and RECON (b) descriptors as a function of *r*.

3. Results and discussion

A data set of guanidinylated aryl 2,5-dideoxystreptamine derivatives exhibiting inhibitory activity against furin with nanomolar potency as determined from biochemical cleavage assay (Table 1) were used for the construction of a structure–property relationship model. Insights into the molecular interaction of GADDs with the furin active site have been performed *in silico* by molecular docking studies [12]. Their results indicated that the positively charged guanidinyl groups of GADD engage in strong electrostatic interaction with the negatively charged residues of furin leading to potent inhibition.

The procedures used in this study are outlined in Fig. 1. The 12 active compounds and six inactive compounds were drawn into the computer using the Marvin software package. The independent variables of this study were based on molecular descriptors calculated from two software packages comprising E-Dragon and RECON.

E-Dragon is the electronic version of the popular Dragon software, which is an application for the calculation of molecular descriptors spanning 20 categories and generating over 1666 variables. Successful applications of Dragon descriptors have previously been demonstrated for modeling the inhibitory activity of aldose reductase enzyme by flavonoid derivatives [30], non-nucleoside inhibitors of HIV-1 reverse transcriptase [31], growth hormone secretagogue agonist activities of tetrahydroisoquinoline 1-carboxamides [32], Pgp-ATPase interaction and rhodamine 123 efflux inhibitory activities of propafenone analogs [33], and selectivity index of pentachlorophenol-imprinted polymer [34].

RECON is based on the quantum theory of atoms in molecules as developed by Bader et al. [19]. The calculated descriptors are known as transferable atom equivalent (TAE), which describes molecules in terms of the electron densities, energies, and properties. The robustness of RECON has previously been demonstrated elsewhere [21,35,36] as well as in our previous investigations [24,27]. It was observed that the charge-based descriptors generated from RECON were suitable for such task as seen from the predictive performance in excess of $r = 0.9$.

Data pre-processing of the raw data was carried out in order to prepare it for further analysis by artificial neural network. Standardization of the independent variables was performed according to Eq. (1). The descriptors were subjected to removal of redundant variables using the RemoveUseless function in Weka and the Unsupervised Forward Selection (UFS) algorithm. UFS selects

Table 2
Predictive performance of ANN models[h]

Software	n[a]	r_{TestCV}[b]				RMS_{TestCV}[c]			
		1[d]	2[e]	3[f]	4[g]	1[d]	2[e]	3[f]	4[g]
E-Dragon	12	0.436	–	0.759	0.807	0.973	–	0.637	0.666
RECON	11	0.866	0.923	–	–	0.489	0.304	–	–

[a] Number of furin inhibitors in the data set.
[b] Cross-validated testing set correlation coefficient.
[c] Root mean square error for cross-validated testing set.
[d] Neutral state of furin inhibitors before optimizing parameter.
[e] Neutral state of furin inhibitors after optimizing parameter.
[f] Protonated state of furin inhibitors before optimizing parameter.
[g] Protonated state of furin inhibitors after optimizing parameter.
[h] All QSAR models were subjected to the same sequential step as detailed in Fig. 1.

maximal linearly independent variables with minimal multiple correlation from the data set, while eliminating redundant and muti-collinear variables. After applying UFS, E-Dragon descriptors were reduced to 11 from an initial number of 1666 descriptors. Similarly, TAE descriptors as calculated by RECON were reduced to 6 from an initial number of 248.

In order to determine the optimal size of the independent variables, the number of descriptors was varied as a function of the correlation coefficient using back-propagation neural network. The volume of descriptors exhibiting high correlation coefficient was identified as the optimal number. The amount of descriptors was in the range of 2–11 for E-Dragon and 2–6 for RECON. Results indicated that the optimal number of descriptors was 5 for both E-Dragon and RECON as presented in Fig. 2. The predictive performances of the generated QSAR models using the reduced set of descriptors are shown in Table 2. Brief definitions and abbreviations of the optimal descriptors derived from E-Dragon and RECON descriptors are illustrated in Table 3.

It is a widely accepted fact that QSAR models usually have local validity as they can only account for compounds which are chemically/biologically similar [37] to the training set molecules [38]. Therefore, the ability to differentiate whether a molecule is biologically active requires a QSAR model that accounts for both active and inactive molecules. However, the quantitative experimental K_i values were not available for those inactive compounds as it was reported in the original article in qualitative terms that K_i is greater than 100 μM. As a result, numerical values could not be assigned to the inactive compounds. Hence, the proposed QSAR models were developed by taking active molecules into account by filtering out inactive compounds possessing $K_i > 100$ μM. To counter such limitation, a two-step QSAR approach was performed as follows:

Table 3
The optimal descriptors used in model generation along with their respective abbreviation and definition

Software	Abbreviation	Descriptor definition
E-Dragon	MATS5p	Moran autocorrelation – lag 5/weighted by atomic polarizabilities
	R5e	R autocorrelation of lag 5/weighted by atomic Sanderson electronegativities
	nBM	Number of multiple bonds
	RDF035m	Radial Distribution Function – 3.5/weighted by atomic masses
	$nArNH_2$	Number of primary amines (aromatic)
RECON	Del(G)NA1	Gradient of B electronic kinetic energy density normal to surface
	PIP18	Local average ionization potential $PIP(r) = \sum \rho_i(r)\lvert\in_i\rvert/\rho(r)$
	Fuk1	Fukui F^+ function scalar value
	Ffuk7	Fukui F^+ function scalar value
	Ffuk9	Fukui F^+ function scalar value

Step 1. Identify whether the furin inhibitor of interest is active or inactive.

If the furin inhibitor of interest is classified as active in Step 1, send the molecule to Step 2 for further processing.

Step 2. Construct a QSAR model of active molecules.

In parallel, such approach has been demonstrated to be effective for the development of QSAR models that can reliably and accurately predict the query activity/property of interest [39]. Particularly, He and Jurs have pointed out that query compounds should be

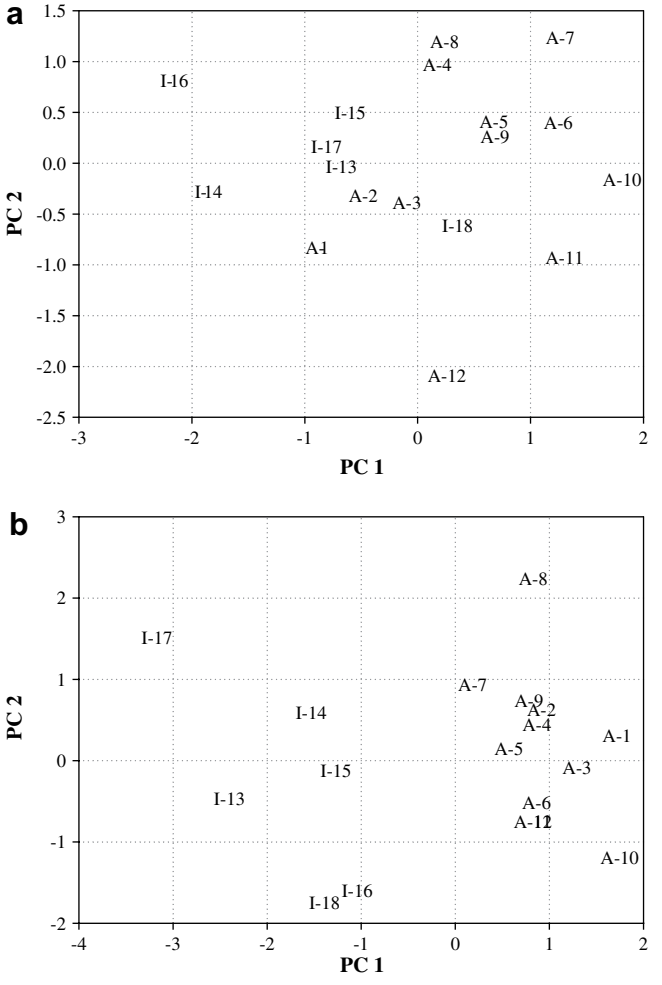


Fig. 3. Score plot of PC1 versus PC2 for models using E-Dragon (a) and RECON (b) descriptors.

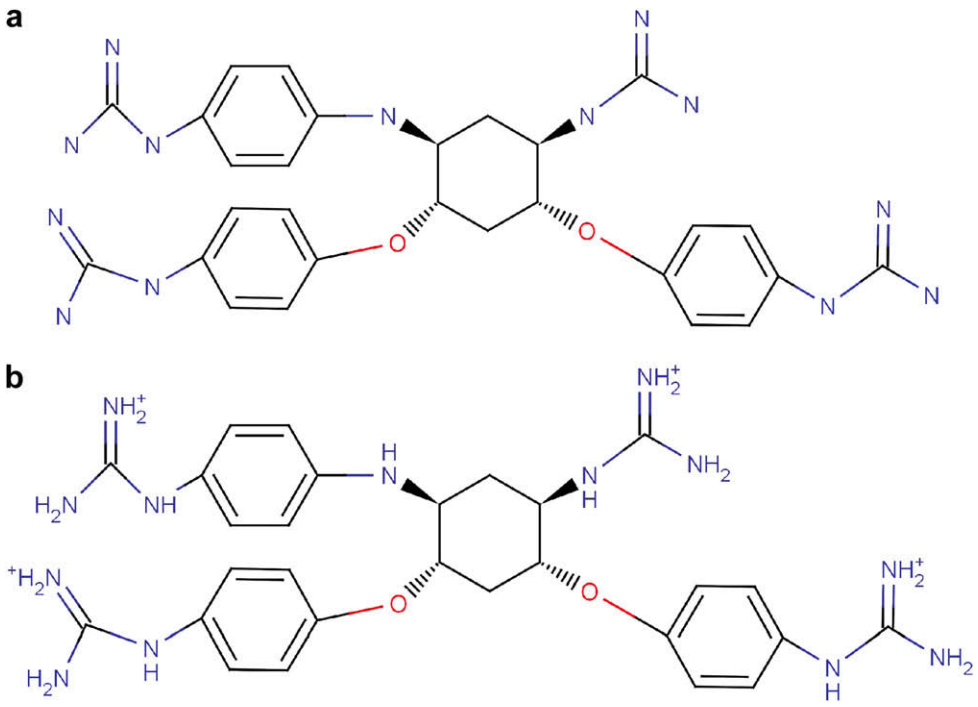


Fig. 4. The chemical structures of furin inhibitor **1n** in the neutral state (a) and protonated state (b) are shown in two-dimensions.

similar to the training set molecules used for building the QSAR model.

Therefore, the classification of active from inactive compounds by PLS was performed using the five optimal descriptors derived from E-Dragon and RECON. The descriptors were autoscaled to zero mean and unit variance prior to analysis. The score plot of PC1 versus PC2 for models using E-Dragon and RECON descriptors is depicted in Fig. 3a and b, respectively. It can be seen from both plot that the furin inhibitors could be separated into two groups: active (**1**–**12**) and inactive (**13**–**18**). PLS model using descriptors derived from RECON demonstrated better classification of the molecules as it could clearly distinguish the active from inactive molecules and could explain 69.77% of the total variance in the first two principal components. On the contrary, descriptors using E-Dragon descriptors did not yield as clear a separation as that of RECON descriptors. Particularly, the PLS model using E-Dragon descriptors could not clearly separate active compounds **1** and **2** from the inactive compounds and, likewise, could not separate inactive molecule **18** from the active molecules. Furthermore, it was observed that the first two principal components could account for only 0.959% of the total variance.

X-ray crystallography has revealed that the furin active site engages in electrostatic interaction with the inhibitors [40]. Therefore, to investigate the importance of the ionization state of the furin inhibitors on the predictive performance, comparison was made between molecular descriptors generated from the neutral state and those from the protonated state (Fig. 4). The protonated forms of the molecular structures in physiological pH were determined with the major microspecies plugin as to account for ionization in experimental buffer. The chemical structures were then modified accordingly in order to investigate the importance of protonation state on the predictive performance of QSAR models using descriptors generated from E-Dragon. Results strongly suggest that taking the protonation state into account was crucial for QSAR models using descriptors calculated from E-Dragon as shown in Table 2. This can be attributed to the fact that experimental K_i was measured in buffer of physiological pH. The TAE methodology of RECON generated charge-based descriptors for molecules of interest by reconstructing information from molecular fragments. Although the algorithm considers molecules at the neutral state, information pertaining to the molecular charge

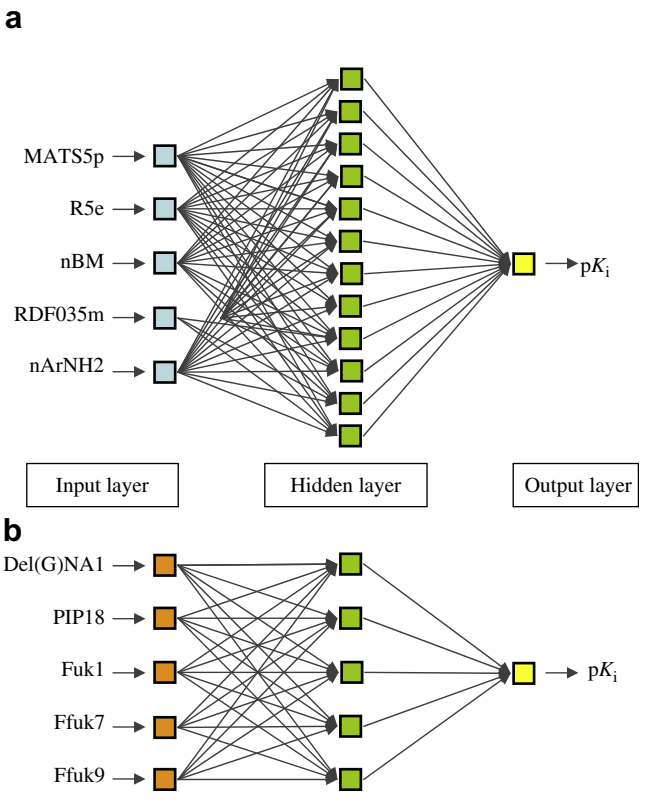


Fig. 5. Schematic representation of the ANN architecture used in this study for models using E-Dragon (a) and RECON (b) descriptors.

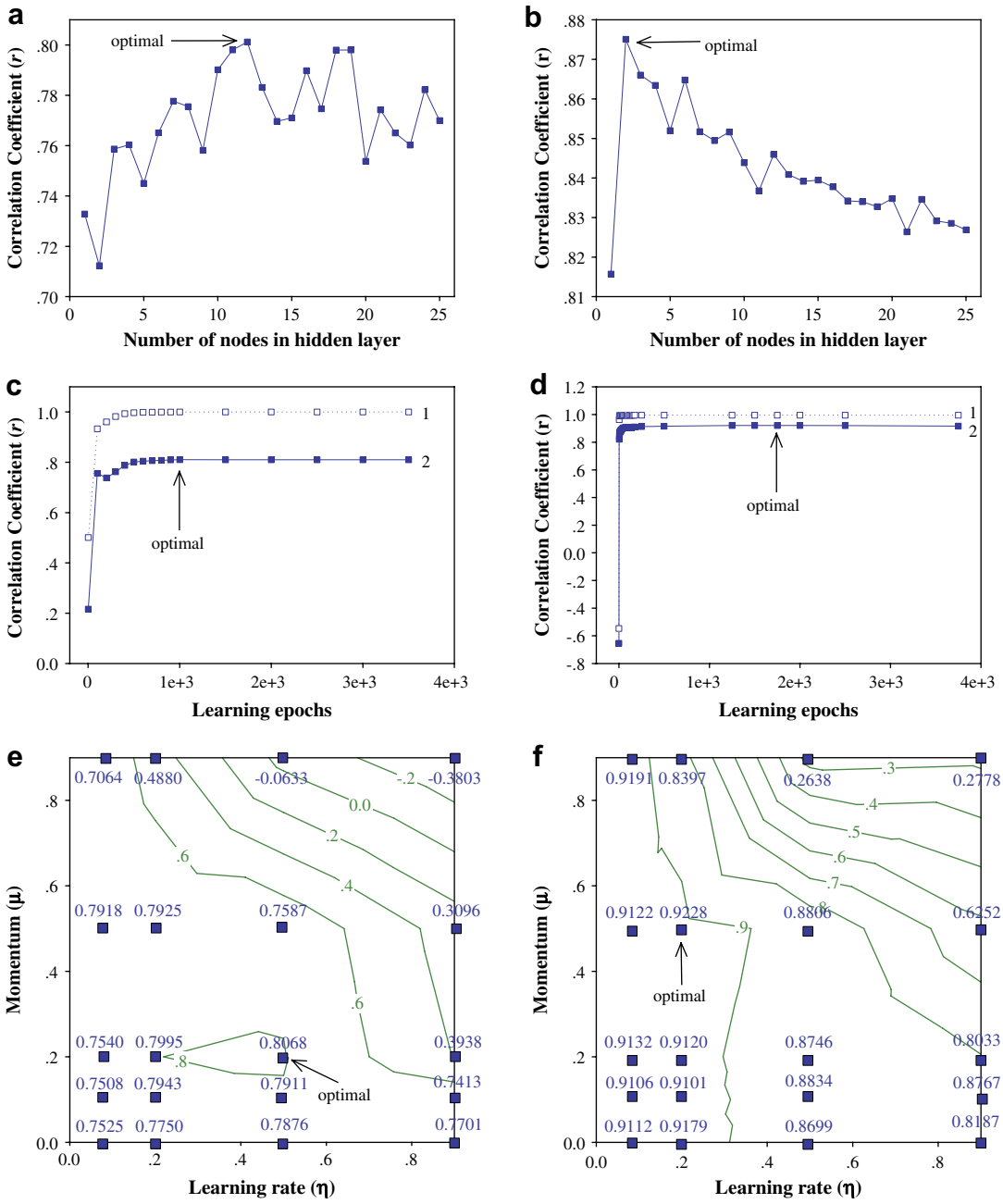


Fig. 6. Optimization of network parameters of the furin inhibitor data set using E-Dragon (a, c, e) and RECON (b, d, f). Plot of the number of hidden nodes as a function of *r* for models using E-Dragon (a) and RECON (b) descriptors. The number of learning epochs as a function of *r* for models constructed with E-Dragon (c) and RECON (d) descriptors where curves 1 and 2 represent the training set and test set, respectively. Contour plot of *r* versus the learning rate and the momentum constant for model utilizing E-Dragon (e) and RECON (f) descriptors. Each line represents constant value of *r*, while shaded boxes represent *r* values obtained from the learning procedure and fitted onto the same surface model of the contour plot.

densities was still present as the software handles these types of information effectively [41]. In Table 2, it is observed that the neutral form of the compounds exhibited low *r* and high RMS for molecular descriptors derived from E-Dragon as compared to the protonated form. This indicates that the protonated form is suitable for further study.

In practical application of the proposed QSAR approach, once molecules are classified as active compounds, they are sent for further regression analysis using ANN. The ANN architecture (Fig. 5) used in this study is made up of three layers: input layer, hidden layer, and output layer. Signals are sent from the input layer through the hidden layer and finally to the output layer in a feed-forward manner. This is followed by readjustment of weights according to the prediction error. The interconnection between the network nodes is known as weights. The initial QSAR models using descriptors derived from RECON indicated that compound **1a** was an outlier (data not shown) and so were discarded from the data set to yield a total of 11 compounds. The original data set using RECON descriptors yielded *r* and RMS of 0.763 and 0.603, respectively; upon removal of the outlier, improvements in the predictive performance were observed from increases of *r* and reduction of RMS to 0.866 and 0.489, respectively.

Before performing actual prediction of pK_i, the following ANN parameters were optimized: number of hidden nodes, size of

Table 4
Summary of the empirically derived optimal ANN parameters

Parameters	E-Dragon[a]	RECON[b]
Number of descriptors	5	5
Hidden node	12	2
Number of learning epochs	1000	70 000
Learning rate	0.5	0.2
Momentum	0.2	0.5

[a] 12 Furin inhibitors in the data set for E-Dragon used in this study.
[b] 11 Furin inhibitors in the data set for RECON used in this study.

Table 5
The experimental and predicted pK_i by ANN[a]

Compound	pK_i	E-Dragon		RECON	
		Predicted	Residual	Predicted	Residual
1a	0.77	0.11	0.46	N/A	N/A
1b	1.05	1.34	0.29	0.94	−0.11
1c	0.39	0.91	0.52	0.50	0.11
1d	1.66	1.66	0.00	1.81	0.15
1e	2.22	1.45	−0.77	1.76	−0.46
1f	1.16	1.63	0.47	1.72	0.56
1g	1.92	3.14	1.22	2.13	0.21
1i	1.38	1.98	0.60	1.16	−0.22
1n	1.34	1.15	−0.19	1.68	0.34
1o	0.37	−0.81	−1.18	0.27	−0.10
1p	0.09	0.46	0.37	−0.25	−0.34
1q	−0.25	−0.82	−0.57	0.09	0.34

Note: NA denotes not available.
[a] 12 and 11 furin inhibitors were present in the data sets modeled with E-Dragon and RECON, respectively.

learning epoch, learning rate and momentum. Correlation coefficient was used as a measure of the predictive performance, therefore, parameters giving high r were selected as optimal. The number of hidden nodes was varied from 1 to 25 for E-Dragon (Fig. 6a) and RECON descriptors (Fig. 6b). Next, the learning epoch size was tested from 1 to 3500 and 1 to 150 000 for models using E-Dragon (Fig. 6c) and RECON descriptors (Fig. 6d), respectively. Finally, the learning rate and momentum were adjusted from 0 to 1 at steps of 0.1 for E-Dragon (Fig. 6e) and RECON (Fig. 6f) descriptors. The optimal parameters as derived from the empirical search are summarized in Table 4 and used for actual prediction. Each run of network training begins with random initialization of weights to obtain a reliable calculation of the predicted outputs as derived from the average of 10 runs of network training.

The experimental and predicted activities along with their respective residuals are listed in Table 5. Comparisons of the experimental values in relation to the predicted ones are represented in Table 2 and Fig. 7. Results suggest that QSAR models computed using RECON descriptors outperformed those of E-Dragon descriptors as observed from leave-one-out cross-validated r of 0.923 and 0.807 together with RMS of 0.304 and 0.666, respectively. The stellar performance of RECON over that of E-Dragon could be ascribed to the fact that RECON provided a better representation of the charge properties of the compounds than those given by E-Dragon. This notion coincides with the reported findings by Jiao and co-workers [12] that the molecular interactions of the inhibitors with the furin active site are essentially electrostatic. The optimal descriptors derived from RECON were comprised of Del(G)NA1, PIP18, Fuk1, Ffuk7, and Ffuk9, which are all based on the electronic charge properties of molecules. The five optimal descriptors derived from E-Dragon comprise of (i) 2D-Autocorrelation indices (MATS5p), (ii) GETAWAY descriptor (R5e) (iii) Constitutional descriptor (nBM), (iv) RDF descriptor (RDF035m), and (v) functional group ($nArNH_2$). Two of the descriptors, MATS5p and R5e, describe the electronic properties of the molecules, particularly in terms of the atomic polarizability and electronegativity.

4. Conclusion

We have demonstrated the application of artificial neural network for predicting the inhibition of furin-dependent processing of anthrax protective antigen by furin inhibitors. It was found that the electronic charge properties were crucial for the prediction as observed from the importance of protonation state on the predictive performance of QSAR models using descriptors derived from E-Dragon. Descriptor derived from RECON was demonstrated to be far more superior to those using E-Dragon descriptors. This study has explored the use of PLS and ANN as efficient methodologies for classifying and predicting the biological activity of furin inhibitors. Therefore, it could be applied as robust tools for the computational design of novel furin inhibitor candidates possessing high bioactivity for future development as therapeutic compounds for furin-dependent diseases.

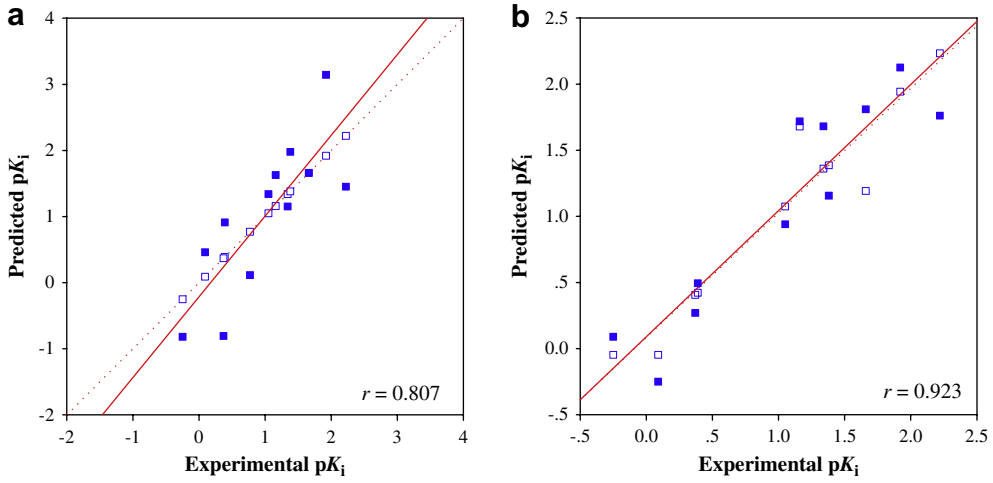


Fig. 7. Plot of the predicted versus the experimental inhibition constant (pK_i) for the training set (□; regression line is represented as dotted line) and the leave-one-out cross-validated test set (■; regression line is represented as a solid line) of models developed with E-Dragon (a) and RECON (b) descriptors.

Acknowledgements

The authors gratefully acknowledge financial support from the Young Scholars Research Fellowship (No. MRG5080450) of the Thailand Research Fund and the governmental budget of Mahidol University (B.E. 2551).

References

[1] M. Mock, A. Fouet, Annu. Rev. Microbiol. 55 (2001) 647–671.
[2] T.C. Dixon, M. Meselson, J. Guillemin, P.C. Hanna, N. Engl. J. Med. 341 (1999) 815–826.
[3] D.J. Banks, S.C. Ward, K.A. Bradley, Expert Rev. Mol. Med. 8 (2006) 1–18.
[4] J.A. Young, R.J. Collier, Annu. Rev. Biochem. 76 (2007) 243–265.
[5] F. Brossier, M. Mock, Toxicon 39 (2001) 1747–1755.
[6] P. Ascenzi, P. Visca, G. Ippolito, A. Spallarossa, M. Bolognesi, C. Montecucco, FEBS Lett. 531 (2002) 384–388.
[7] F. Jean, K. Stella, L. Thomas, G. Liu, Y. Xiang, A.J. Reason, G. Thomas, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 7293–7298.
[8] M.S. Sarac, J.R. Peinado, S.H. Leppla, I. Lindberg, Infect. Immun. 72 (2004) 602–605.
[9] Z.X. Liu, H. Fei, C.W. Chi, FEBS Lett. 556 (2004) 116–120.
[10] T. Komiyama, R.S. Fuller, Biochemistry 39 (2000) 15156–15165.
[11] G. Bolt, I.R. Pedersen, Virology 252 (1998) 387–398.
[12] G.S. Jiao, L. Cregar, J. Wang, S.Z. Millis, C. Tang, S. O'Malley, A.T. Johnson, S. Sareth, J. Larson, G. Thomas, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 19707–19712.
[13] ChemAxon Ltd., Marvin, Version 3.5.4, Budapest, Hungary.
[14] Milano Chemometrics and QSAR Research Group, E-Dragon, Version 1.0, Italy.
[15] Rensselaer Polytechnic Institute, RECON, Version 5.5, New York, USA.
[16] I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V.V. Prokopenko, J. Comput. Aided Mol. Des. 19 (2005) 453–463.
[17] R. Todeschini, V. Consonni, R. Mannhold, H. Kubinyi, H. Timmerman, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.
[18] C.M. Brenemen, M. Rhem, J. Comput. Chem. 18 (1997) 182–197.
[19] R.F.W. Bader, S.G. Anderson, A.J. Duke, J. Am. Chem. Soc. 101 (1979) 1389–1395.
[20] C.M. Breneman, C.M. Sundling, N. Sukumar, L. Shen, W.P. Katt, M.J. Embrechts, J. Comput. Aided Mol. Des. 17 (2003) 231–240.
[21] C.B. Mazza, K. Rege, C.M. Breneman, N. Sukumar, J.S. Dordick, S.M. Cramer, Biotechnol. Bioeng. 80 (2002) 60–72.
[22] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, second ed. Morgan Kaufmann, San Francisco, 2005.
[23] D.C. Whitley, M.G. Ford, D.J. Livingstone, J. Chem. Inf. Comput. Sci. 40 (2000) 1160–1168.
[24] C. Nantasenamat, T. Naenna, C. Isarankura Na Ayudhya, V. Prachayasittikul, J. Comput. Aided Mol. Des. 19 (2005) 509–524.
[25] Camo Software AS, The Unscrambler, Version 9.6, Norway.
[26] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1988) 1193–1202.
[27] C. Nantasenamat, C. Isarankura-Na-Ayudhya, T. Naenna, V. Prachayasittikul, Biosens. Bioelectron. 22 (2007) 3309–3317.
[28] C. Nantasenamat, C. Isarankura-Na-Ayudhya, N. Tansila, T. Naenna, V. Prachayasittikul, J. Comput. Chem. 28 (2007) 1275–1289.
[29] J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design, second ed, Wiley-VCH, Weinheim, 1999.
[30] M. Fernandez, J. Caballero, A.M. Helguera, E.A. Castro, M.P. Gonzalez, Bioorg. Med. Chem. 13 (2005) 3269–3277.
[31] P.R. Duchowicz, M. Fernandez, J. Caballero, E.A. Castro, F.M. Fernandez, Bioorg. Med. Chem. 14 (2006) 5876–5889.
[32] J. Caballero, F.M. Zampini, S. Collina, M. Fernandez, Chem. Biol. Drug Des. 69 (2007) 48–55.
[33] V. Tantishaiyakul, W. Wongpuwarak, J. Mol. Struct. (Theochem) 718 (2005) 183–189.
[34] C. Nantasenamat, T. Tantimongcolwat, T. Naenna, C. Isarankura Na Ayudhya, V. Prachayasittikul, Excli J. 5 (2006) 150–163.
[35] J.S. Murray, P. Politzer, G.R. Famini, J. Mol. Struct. (Theochem) 454 (1998) 299–306.
[36] B.K. Lavine, C.E. Davidson, C. Breneman, W. Katt, J. Chem. Inf. Comput. Sci. 43 (2003) 1890–1905.
[37] L. Eriksson, E. Johansson, M. Müller, S. Wold, J. Chemometr. 14 (2000) 599–616.
[38] R. Guha, P.C. Jurs, J. Chem. Inf. Model. 45 (2005) 65–73.
[39] L. He, P.C. Jurs, J. Mol. Graphics Model. 23 (2005) 503–523.
[40] S. Henrich, A. Cameron, G.P. Bourenkov, R. Kiefersauer, R. Huber, I. Lindberg, W. Bode, M.E. Than, Nat. Struct. Biol. 10 (2003) 520–526.
[41] C.M. Breneman, T.R. Thomson, M. Rhem, M. Dung, Comput. Chem. 19 (1995) 161–179.

Molecules **2008**, *13*, 3040-3056; DOI: 10.3390/molecules13123040

OPEN ACCESS

molecules

ISSN 1420-3049

www.mdpi.com/journal/molecules

Article

Copper Complexes of Nicotinic-Aromatic Carboxylic Acids as Superoxide Dismutase Mimetics

Thummaruk Suksrichavalit [1], Supaluk Prachayasittikul [2], Theeraphon Piacham [1], Chartchalerm Isarankura-Na-Ayudhya [1], Chanin Nantasenamat [1] and Virapong Prachayasittikul [1,*]

[1] Department of Clinical Microbiology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand; E-mails: rukmu47@gmail.com (T. S.), mttpc@mahidol.ac.th (T. P.), mtcis@mahidol.ac.th (C. I.), mtcnt@mahidol.ac.th (C. N.), mtvpr@mahidol.ac.th (V. P.)

[2] Department of Chemistry, Faculty of Science, Srinakharinwirot University, Bangkok 10110, Thailand; E-mail: supaluk@swu.ac.th (S. P.)

* Author to whom correspondence should be addressed; E-mail: mtvpr@mahidol.ac.th; Tel.: +66-2-418-0227; Fax: +66-2-412-4110.

Received: 25 September 2008; in revised form: 19 November 2008 / Accepted: 27 November 2008 / Published: 8 December 2008

Abstract: Nicotinic acid (also known as vitamin B3) is a dietary element essential for physiological and antihyperlipidemic functions. This study reports the synthesis of novel mixed ligand complexes of copper with nicotinic and other select carboxylic acids (phthalic, salicylic and anthranilic acids). The tested copper complexes exhibited superoxide dismutase (SOD) mimetic activity and antimicrobial activity against *Bacillus subtilis* ATCC 6633, with a minimum inhibition concentration of 256 µg/mL. Copper complex of nicotinic-phthalic acids (CuNA/Ph) was the most potent with a SOD mimetic activity of IC_{50} 34.42 µM. The SOD activities were observed to correlate well with the theoretical parameters as calculated using density functional theory (DFT) at the B3LYP/LANL2DZ level of theory. Interestingly, the SOD activity of the copper complex CuNA/Ph was positively correlated with the electron affinity (EA) value. The two quantum chemical parameters, highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), were shown to be appropriate for understanding the mechanism of the metal complexes as their calculated energies show good correlation with the SOD activity. Moreover, copper complex with the highest SOD activity were shown to

possess the lowest HOMO energy. These findings demonstrate a great potential for the development of value-added metallovitamin-based therapeutics.

Keywords: Nicotinic acid; Copper; Carboxylic acid; Superoxide dismutase; Antimicrobial activity.

Introduction

Nicotinic acid (NA) or vitamin B3 is essential for many biological processes namely for the production of energy [1], signal transduction, regulation of gene expression [2] and involvement in the synthetic pathway of lipids [3]. The oxidation of lipoproteins, low density lipoprotein and very low density lipoprotein, by free radical, particularly superoxide radical ($O_2^{\bullet -}$), causes vascular inflammation, which is involved in the early stage development of atherosclerosis, a disease affecting arterial blood vessels [4]. In developed countries such as the United States, atherosclerosis is the leading cause of illness and death accounting for 44% of deaths and morbidity [5]. Nicotinic acid has been reported to decrease the production of free fatty acids and lipoproteins [6], despite of this free radical continues to appear in the human body. To eliminate such superoxide radicals, transition metal complexes were developed as superoxide dismutase mimics. Some examples include metalloporphyrin [7] and metal-drug complexes [8].

Complexation of nicotinic acid with various metals, e.g. manganese, cobalt, nickel, copper and zinc had been previously reported [9]. Nicotinic acid-copper complex (CuNA) has been shown to exert diverse bioactivities. In particular, it can stimulate blood flow and prevent gastric congestion [10], reduce total lipids in sera hepatic tissue as well as regulate levels of alanine transaminase, aspartate transaminase, alkaline phosphatase, gamma glutamyl transpeptidase and oxidative markers such as nitric oxide (NO) and lipid peroxidation in rat models [11]. Additionally, it also exhibits SOD mimic activity in patients with hepatocellular carcinoma. The metal complexes of nicotinic acid with other small molecules such as isonicotinate and acetylacetonate were previously reported [9, 12]. The aforementioned nicotinic acid-copper complexes were formed through coordination with the carboxylate oxygen- and/or pyridine nitrogen-atoms [12, 13], which are electron donors. These electron-donating groups are commonly found in many natural compounds such as flavonoids (rutin, taxifolin, epicatechin and luteolin), phenolics (catechol and resveratrol) and drugs (aspirin, ibuprofen and oxaprozin), all of which were reported to form metal complexes with potent superoxide scavenging capacities [14-16]. For example, flavonoid copper complexes were shown to exhibit higher SOD activity than the parent free flavonoids [15]. A series of simple Cu carboxylate (acetate, salicylate or benzoate) complexes synthesized by Devereux *et al*., were reported to exhibit excellent SOD mimic activity [17]. Salicylic acid, one of the metal-chelating ligands used in this study, is a natural compound found in most fruits and vegetables and functioning as a plant growth hormone. The compound was shown to possess promising anti-cancer activity by modulating the inhibition of the P form of phenolsulphotransferase, thus preventing excessive carcinogen activation [18]. Anthranilic acid (also known as vitamin L), another ligand used in the study, is a natural ingredient essential for lactation. Its derivatives have been found to activate soluble guanylyl cyclase, a ubiquitous NO

receptor involved in vasorelaxation and hypertension [19]. In addition, copper atom is an essential nutrient involved in the catalytic function of many enzymes such as copper-zinc superoxide dismutase (CuZnSOD) [20, 21]. Copper deficiency has been reported to cause hematologic disorders, hypopigmentation, defective connective tissue cross-linking and ataxia [22, 23].

The ideal SOD-mimic should be a low molecular weight metal complex that possesses high membrane permeability. However, very few Cu-carboxylate complexes had been reported [17]. In an attempt to obtain small molecule-based metal complexes, we now report the synthesis of novel copper complexes as potent SOD mimics. Such complexes are based on the coordination of nicotinic acid-Cu with the small molecule carboxylates phthalic acid (Ph), salicylic acid (Sal) and anthranilic acid (Ant). Computational chemistry has found extensive applications for studying molecular structure and reactivity [24]. Some usage examples include studying the mechanism of protein receptor-inhibitor interaction [25], development of quantitative structure-activity/property relationship (QSAR/QSPR) [26], and elucidation of enzymatic reaction [27], etc. To shed light on the underlying mechanisms of the superoxide radical scavenging activities, quantum chemical parameters were calculated at DFT level. Results indicated that the calculated physicochemical parameters were well correlated with the experimental SOD activities. Such theoretical parameters included energies of HOMO and LUMO as well as EA. Furthermore, to confer value-added benefits to the copper complexes as well as explore the possibility of using such compounds as a multifaceted drug the antimicrobial activity of the novel complexes was also determined.

Results and Discussion

Three copper-based complexes, copper complexes of nicotinic-phthalic acids (CuNA/Ph, **1**), nicotinic-salicylic acids (CuNA/Sal, **2**) and nicotinic-anthranilic acids (CuNA/Ant, **3**), were synthesized by the reaction of 1:1:1 ratio of cupric chloride with nicotinic acid and aromatic carboxylic acids (phthalic, salicylic and anthranilic acids). The complexes were obtained in excellent yields (82-91%) as aquamarine powder, highly polar, insoluble in methanol and water, but soluble in dimethyl sulfoxide (DMSO). Their melting points (m.p.) and magnetic moments (μ_{eff}) are listed in Table 1.

Table 1. Physicochemical parameters of nicotinic acid-copper complexes and ligands.

Compound	Chemical Formula	Formula Weight ($g \cdot mol^{-1}$)	Color	Melting Point (°C)	Yield (%)	μ_{eff} (B.M.)
NA	$C_6H_5NO_2$	123.11	white	236-239	–	–
Ph	$C_8H_6O_4$	166.13	white	210	–	–
CuNA/Ph (**1**)	$C_{14}H_{10}CuNO_7$	367.78	aquamarine	>300	82	1.6944
Sal	$C_7H_6O_3$	138.12	white	158-160	–	–
CuNA/Sal (**2**)	$C_{13}H_{10}CuNO_6$	339.77	aquamarine	>300	85	1.6802
Ant	$C_7H_7NO_2$	137.14	yellow	144-148	–	–
CuNA/Ant **(3)**	$C_{13}H_{13}CuN_2O_5$	338.78	aquamarine	>300	91	1.8038

Infrared spectra

Coordination of the copper atom with the functional groups of the ligands was established from the IR spectra (Table 2). From the data of complex **1**, it can be seen that the C-O stretching vibration frequency was located at 1,298 cm^{-1}, while the free phthalic acid and nicotinic acid absorptions υ(C-O) appeared at 1,282 and 1,299 cm^{-1}, respectively. In addition, disappearance of the hydroxyl group bending vibration (δO-H) of Ph at 1,404 cm^{-1} was observed. For the carbonyl group (C=O), the stretching vibrations of both NA and Ph still exhibited intensive bands at 1,685, 1,701, and 1,718 cm^{-1}. The weak band at 1333 cm^{-1} could be ascribed to C-N stretching vibration of complex **1** as compared with the spectra of NA, which shows strong band υ(C-N) at 1324 cm^{-1}. Thus, complex **1** was formed using pyridine ring nitrogen atom of NA and two hydroxyl groups of Ph to coordinate with copper atom. Structures of ligands are shown in Figure 1.

Table 2. IR spectra of the free ligands and the copper coordination complexes.

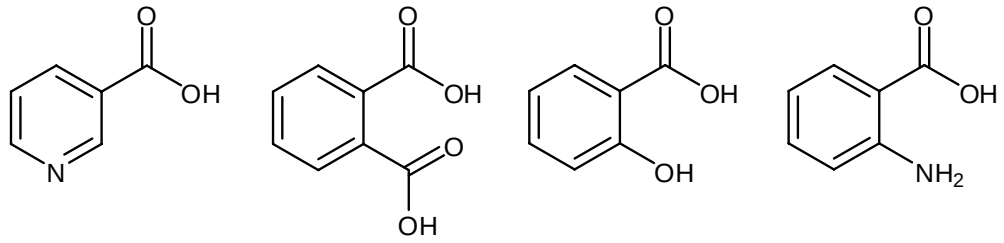
Cpd	υC=O	υC-O	υC-N	υO-H	δO-H	υNH	δNH
NA	1,718 (s) 1,700 (s)	1,299 (s)	1,324 (s)	3,072-2,449 (w)	1,418 (s)	–	–
Ph	1,700 (s) 1,685 (s)	1,282 (s)	–	3,072-2,524 (br)	1,404 (s)	–	–
1	1,718 (s) 1,701 (s) 1,685 (s)	1,298 (m)	1,333 (w)	3,080-2,565 (w)	1,418 (m)	–	–
Sal	1,662 (s) 1,655 (s)	1,296 (s) 1,250 (s) 1,211 (s) 1,156 (s)	–	3,471 (br) 3,240 (sh)	1,446 (s) 1,484 (s)	–	–
2	1,717 (s) 1,700 (s) 1,685 (s)	1,298 (s) 1,201 (w) 1,144 (w) 1,126 (w)	1,332 (m)	2,682 (sh, w) 2,567 (sh, w)	1,454 (m) 1,420 (m)	–	–
Ant	1,679 (m) 1,654 (sh, m)	1,277 (m) 1,238 (m)	1,371 (s) 1,319 (s)	2,869 (br) 2,580 (br) 2,362 (sh, w) 3,449 (br)	1,458 (m)	3,325 (sh, s) 3,239 (sh, s)	754 (vs)
3	1,718 (s) 1,700 (s)	1,298 (s)	1,386 (s) 1,332 (m)	3,400 (br) 2,681 (w) 2,565 (w)	1,458 (m) 1,420 (m)	3,275 (m)	756 (s)

Note: vs = very strong, s = strong, m = medium, w = weak, br = broad, sh = sharp.

In a similar fashion, the IR spectra of complex **2** were interpreted by comparison of its spectra with those of the free ligands. It was apparent that the hydroxyl and carbonyl absorptions of salicylic acid were at 3,471 cm^{-1} υ(O-H) and 1,655, 1,662 cm^{-1} υ(C=O), respectively, suggesting that Cu-complex

formation took place via the phenolic and carbonyl functionalities. The IR spectra of complex **2** still displayed strong absorptions at 1,700, 1,717 cm^{-1} υ(C=O), 1,298 cm^{-1} υ(C-O) and 1,420 cm^{-1} δ(O-H). However, the C-N stretching vibration of complex **2** was observed at 1,332 cm^{-1} whereas the C-N absorption of free NA appeared at 1,324 cm^{-1}. Based on the IR spectra, complex **2** was observed to be formed through the carbonyl and the phenolic hydroxyl group of Sal and the pyridine ring nitrogen atom.

Figure 1. Molecular structure of ligands: nicotinic acid (a), phthalic acid (b), salicylic acid (c), and anthranilic acid (d).



The IR spectra of complex **3** exhibited disappearance of the anthranilic acid carbonyl (C=O) absorption at 1,654 and 1,679 cm^{-1}. Additionally, the amino (NH) absorption of complex **3** appeared at 3,275 cm^{-1}, while the NH_2 absorption of Ant was observed at 3,239 and 3,325 cm^{-1}. Strong C-N stretching of Ant was displayed at 1,371 and 1,319 cm^{-1} with a very strong NH bending observed at 754 cm^{-1}. The IR spectra of complex **3** showed strong absorptions at 1,718, 1,700 (C=O), 1,298 (C-O), medium bending band of OH at 1,420 cm^{-1} and C-N stretching of ring N-atom at 1,332 cm^{-1}. The characteristic free NA C-N stretching absorption of the pyridine ring (υ 1,324 cm^{-1}) was shifted to higher frequency at 1332 cm^{-1}. The C-N stretching of NH group of complex **3** was observed at 1,386 cm^{-1}. It was suggested that complex **3** was formed through coordination of pyridine ring nitrogen atom; carbonyl and NH groups of Ant with Cu centered atom.

It is noted that these Cu-complexes were formed via monodentate interaction of the ring nitrogen atom of NA resulting in a υ(C-N) shift from 1,324 cm^{-1} to a higher frequency range of 1,332-1,333 cm^{-1}. The υ(C=O), υ(C-O) and δ(O-H) absorptions of the NA ligand of the complexes remained unperturbed. Aside from NA, other ligands coordinated with the central Cu atom via bidentate carboxylates. With regards to complex **1**, Ph is the bidentate ligand using two hydroxyls of dicarboxylic acid to coordinate with the Cu atom. This is observed by the disappearance of C-O stretching at 1,282 cm^{-1} and O-H bending at 1,404 cm^{-1}. As for complex **2**, stretching at 1,662 and 1,655 cm^{-1} (C=O) and at 3,471 and 3,240 cm^{-1} (O-H) disappeared with a shift of C-O stretching from 1,250, 1,211 and 1,156 cm^{-1} to lower frequencies. This suggested that the Cu atom formed a complex by means of a bidentate phenolic carboxylate. Similarly, the Cu atom of complex **3** was found to be coordinated with the amino carboxylate of Ant ligand, which is due to the disappearance of C=O and NH_2 stretching absorptions of Ant.

High resolution mass spectra (HRMS) measurements for complexes **1**-**3** were performed, but their molecular ions were not detected. Fragmented ions were observed at m/z of 124.0399 corresponding to

free nicotinic acid as calculated for $C_6H_5NO_2$ $[M+H]^+$. Additionally, m/z of 124.0399 was quantified for complexes **1** and **2** while 124.0396 was detected for complex **3**.

Complexes **1-3** were paramagnetic, with magnetic moments (μ_{eff}) of 1.69, 1.68 and 1.80 B.M., respectively. The μ_{eff} value indicates that the complexes were of tetrahedral conformation with the Cu(II) center. Therefore, on the basis of both IR spectra and μeff, we can confirm that complexes **1-3** were of tetrahedral geometry with the Cu(II) center.

Superoxide scavenging activity

The copper complexes were assayed for SOD-like activity using the modified method by measuring the inhibition of the photoreduction of nitro blue tetrazolium (NBT). The results showed that all the tested complexes exhibited SOD activities with IC_{50} in the range of 34.42-47.49 μM as presented in Table 3.

Table 3. Superoxide dismutase activity of the free ligands and copper complexes.

Compound*	***IC_{50} (μM)*** [a]
Sal	>594.00
Ant	>236.25
1	34.42
2	42.79
3	47.49
SOD [b]	0.0026

[a] IC_{50} was defined as fifty percent inhibition concentration of NBT reduction.

[b] Superoxide dismutase from bovine erythrocytes was a homodimeric protein.

***** Nicotinic acid and phthalic acid possessed very low activity, to the point that the IC_{50} cannot be obtained.

Complex **1** was the most potent SOD mimic with an IC_{50} of 34.42 μM. Moreover, the SOD activity of the free ligand, NA and Ph, was very low to the point that the IC_{50} could not be obtained. The SOD activity of the uncoordinated salicylic and anthranilic acids was assayed and their IC_{50} was determined to be 594 and 236.25 μM, respectively. It was observed that complex **2** was approximately 14 times more potent than the uncoordinated Sal. Similarly, complex **3** was five times more potent than the free Ant. The results support the notion that the copper atom is essential for SOD activity and that the incorporation of copper into the structure of the ligands tremendously increased the SOD activity. We attribute this to the change in oxidation state of the copper atom modulated through its coordination with the metal-chelating ligands. It is noteworthy that the coordination complexes are much smaller than the native enzymes. Moreover, an added benefit is that the compounds can be easily synthesized from simple bioactive compounds. Some of the potent SOD mimics developed thus far were based on Mn and Cu coordination compounds. Example of such SOD mimics are the Cu-drug complexes which included Cu (aspirin), Cu $(aspirin)_4$ and Cu (ibuprofenate).

Antimicrobial activity

Nicotinic acid is a widely used drug for the treatment of hyperlipidemia [3]. In addition, salicylic and anthranilic acids are bioactive aromatic carboxylic acids. The antimicrobial activity of complexes **1-3** has not been reported before. As the objective of this study is to develop value-added metallovitamins for therapeutic applications, therefore it is worthy to observe whether the compounds could also be used as antimicrobials, aside from being superoxide radical scavengers. Our results revealed that the tested complexes **1-3** can selectively inhibit the growth of *B. subtilis* ATCC 6633 with a MIC of 256 μg/mL (Table 4).

Table 4. Antimicrobial activity of the free ligands and nicotinic acid-copper complexes.

Compound	Concentration (μg/mL)	Activity
Ampicillin	25	Active[a]
NA	256	Not Active
$CuCl_2$	64*	Not Active
	32*	Not Active
Ph	256	Not Active
1	256	Active[b]
	128	Active[c]
	64	Not Active
Sal	256	Active[b]
2	256	Active[b]
	128	Active[d]
	64	Not Active
Ant	256	Not Active
3	256	Active[b]
	128	Active[c]
	64	Not Active

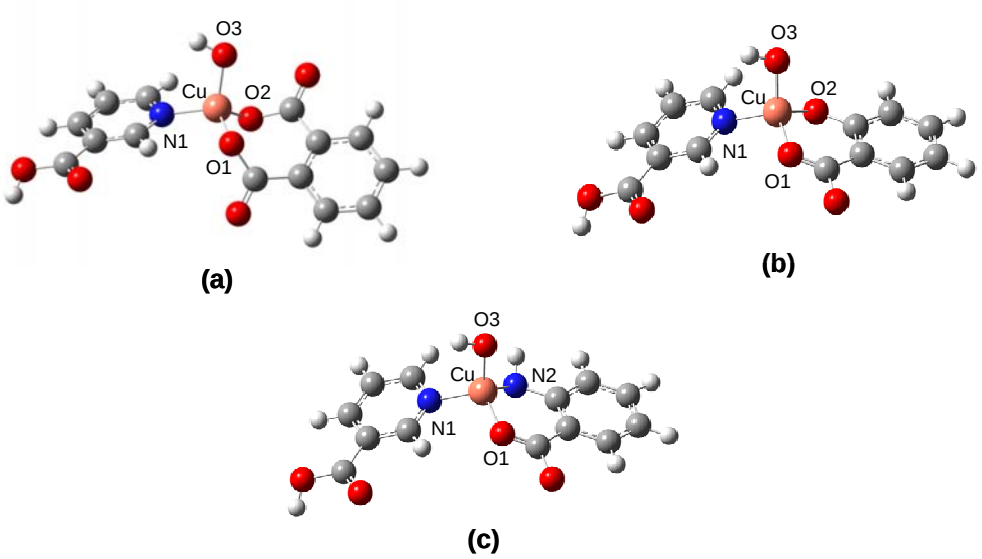
[a] Active (100% antigrowth) against *P. aeruginosa* ATCC 15442, *S. putrefaciens* ATCC 8671, *A. xylosoxidans* ATCC 2706, *S. aureus* ATCC 25923, *S. epidermidis* ATCC 12228, *E. faecalis* ATCC 29212, *B. subtilis* ATCC 6633, *S. cereviseae* ATCC 2601, *S. pyogenes* II, *S. enteritidis* type C, *P. shigelloides, L. monocytogenes*. Active against *B. subtilis* ATCC 6633 with [b] 100%, [c] 50%, [d] 25% antimicrobial activity. * $CuCl_2$ was tested against *B. subtilis* ATCC 6633.

It should be noted that salicylic acid was the only free ligand that could completely inhibit the growth of *B. subtilis* ATCC 6633 (MIC of 256 μg/mL) and that the antimicrobial activity of salicylic acid and its derivatives against *E. coli*, *B. subtilis* and *S. aureus* has been previously reported [28]. Aside from this, the other free ligands (e.g. nicotinic, phthalic and anthranilic acids) were inactive against all of the tested microorganisms. A plausible explanation for the antimicrobial activity of the copper complexes is that they may enhance bacterial killing by synergistically converting superoxide radical to hydrogen peroxide in which accumulation of hydrogen peroxide exerted harmful effect to the bacterial cells as well as participated in the subsequent formation of hydroxyl radical via the Fenton's reaction.

Molecular modeling of SOD mimics

To elucidate the mechanisms of radical scavenging activities of the copper complexes, density functional theory calculations at the B3LYP/LANL2DZ level was employed. Previous efforts have indicated that the calculation of theoretical parameters, binding energies and electron affinities, at such theoretical levels are suitable for characterizing the superoxide radical scavenging activity [29-32]. The molecular structures were constructed on the basis of IR and magnetic moment data which indicated that the coordination complex was of distorted tetrahedral conformation. Geometrically optimized structure of complexes **1-3** are presented in Figure 2.

Figure 2. Molecular structures of copper coordination complexes **1** (a), **2** (b), and **3** (c).



Electron affinity, in particular, is an appropriate theoretical parameter accounting for the electron transfer rate from superoxide anion to copper atom [30-32]. The lower the EA becomes the higher the electron transfer rate is, which correspondingly leads to higher superoxide radical scavenging activity. EA was calculated according to equation (1) by taking the difference of the total energy of Cu(II) and Cu(I) coordination complexes. The EA for complexes **1-3** as presented in Table 5 were calculated to be -125.502, -139.283 and -212.679 kcal/mol, respectively.

Table 5. Theoretical parameters of the copper complexes.

Complex	$TE_{Cu(II)}$ [a] (hartree)	$TE_{Cu(I)}$ [b] (hartree)	EA* (kcal/mol)
1	-1316.673	-1316.873	-125.502
2	-1203.223	-1203.445	-139.283
3	-1183.292	-1183.631	-212.679

[a] Total energy of the Cu(II) coordination complex

[b] Total energy of the Cu(I) coordination complex

* EA was calculated according to equation (1)

The experimental SOD activities of complexes **1-3** had an IC_{50} of 34.42, 42.79 and 47.49 μM, respectively. The results indicated that there exists a positive correlation between EA and SOD activity where high EA gives rise to high SOD activity, which is inversely correlated in previous reports [30, 31]. This is presumably due to differences in the coordination geometry of the copper complexes used in this study (tetragonally distorted) and the copper complexes (distorted square-planar and square-pyramidal) reported previously [30, 31]. Similar observation was deduced by Branco *et al*. in their studies on active site distortion of CuZnSOD [33, 34].

Additionally, the usefulness of quantum chemical descriptors such as energy of the highest occupied molecular orbital and energy of the lowest unoccupied molecular orbital in elucidating the radical scavenging activity were also investigated. Examples on the usage of HOMO and LUMO energies in accounting for chemical reactivities of molecules and their relevance to electron transfer complexes can be found in an excellent review by Karelson and Lobanov [35]. The HOMO energies of the ligands were calculated to be -5.818, -6.729 and -7.644 eV, respectively with the following order: Ant > Sal > Ph (Table 6).

Table 6. HOMO and LUMO energies of the free ligands and copper complexes.

Compound	HOMO (eV)	LUMO (eV)
Ph	-7.644	-2.252
Sal	-6.729	-1.749
Ant	-5.818	-1.553
1	-7.302	-4.898
2	-6.293	-4.595
3	-5.829	-4.086

Likewise, the calculated value of HOMO for the complexes exhibited similar trend: **3** > **2** > **1** with the corresponding values of -5.829, -6.293 and -7.302 eV, respectively. It is well established that HOMO accounts for the electron donating ability while LUMO characterizes the ability to accept electron [36]. From the frontier molecular orbital approximation, high HOMO energy value infers that the molecule or ligand can easily release electrons to the unoccupied orbital of the metal ion, indicating strong binding affinity [37]. Thus, Ant possesses the strongest interaction with copper leading to the best binding capacity. Such degree of binding capacity is found to be Ant > Sal > Ph.

In addition, the selected bond lengths and angles of the geometrically optimized structures are given in Table 7. The average bond distance at axial position for Cu-N1 of complexes **1-3** was 2.029 Å. The longest axial bond length for Cu-O1 and Cu-O2 were observed for complex **1** with values of 1.872 and 1.870 Å, respectively. The axial bond lengths of Cu-O1 and Cu-O2 for complex **2** was found to be shorter than that of complex **1** with values of 1.838 and 1.857 Å, respectively. This is explained by the inductive effect of the two carbonyl groups of complex **1** withdraws electrons from the Cu atom giving rise to low electron donating ability of the ligand as also indicated by the lower HOMO energy. Therefore, complex **1** exhibited the highest SOD activity. Finally, the shortest axial bond lengths for Cu-N2 and Cu-O1 were found in complex **3** to be 1.809 and 1.830 Å, respectively. This can be attributed to the greater electron donating ability of the amino group of anthranilic acid than the phenolic group of salicylic acid. This results in the lower HOMO energy of **2** than **3**. It was previously

reported by Li *et al*. [38] that axial bond lengths were crucial for SOD activity where long bond lengths being advantageous for the dismutation of superoxide anion. In addition, correlation between the binding capacity and the SOD activity of metal complexes was previously studied using molecular modeling and quantum chemical calculation [8]. It was found that molecules exhibiting higher metal binding affinity displayed lower SOD activity [8, 39]. This was observed for complex **3** which possessed the lowest SOD activity and the highest calculated HOMO value. On the other hand, complex **1** exhibited the highest SOD activity with the lowest calculated HOMO value (Tables 3 and 6). Furthermore, the calculated energies of HOMO and LUMO were well correlated with the SOD activity as observed from r = 0.999 and 0.953, respectively. Such results demonstrate the practical usage of HOMO and LUMO as theoretical parameters for the characterization of SOD activity in terms of charge- or electron-transfer of the complex.

Table 7. Selected bond distances (Å) and angles (°) of complexes **1-3**.

Complex	Bond Lengths (Å)		Angles (°)	
1	Cu-N1	2.035	N1-Cu-O1	111.475
	Cu-O1	1.872	N1-Cu-O2	111.235
	Cu-O2	1.870	N1-Cu-O3	107.711
	Cu-O3	1.829	O1-Cu-O2	105.661
			O1-Cu-O3	110.275
			O2-Cu-O3	110.513
2	Cu-N1	2.035	N(1)-Cu-O1	109.012
	Cu-O1	1.838	N(1)-Cu-O2	111.596
	Cu-O2	1.857	N(1)-Cu-O3	109.467
	Cu-O3	1.836	O(1)-Cu-O2	107.727
			O(1)-Cu-O3	109.929
			O(2)-Cu-O3	109.355
3	Cu-N1	2.016	N1-Cu-O1	109.926
	Cu-N2	1.809	N1-Cu-O2	109.365
	Cu-O1	1.830	N1-Cu-N2	110.023
	Cu-O2	1.835	N2-Cu-O2	109.091
			N2-Cu-O1	108.241
			O1-Cu-O2	109.581

Conclusions

The novel copper complexes with nicotinic acid and carboxylic acids were synthesized in excellent yields. The coordination compounds were characterized to be tetragonally distorted structures formed by monodentate coordination of ring N-atom of NA as well as bidentate carboxylate ligands using amino carbonyl (Ant), phenolic carbonyl (Sal) and two hydroxyls (Ph). It should be noted that from the chemical and biological point of view, these copper complexes could easily be formed through the

coordination of metal ions with simple dietary intake of vitamins or endogenous carboxylic acids and related compounds. All of the copper complexes were shown to exhibit superoxide scavenging activity as well as antimicrobial activities against *B. subtilis* ATCC 6633 with MIC 256 μg/mL. Complex **1** was the most potent SOD activity with an IC_{50} of 34.42 μM. Theoretical parameters as calculated by DFT methods were used to elucidate the mechanisms of SOD activity. Interestingly, copper complexes with higher SOD activity such as complex **1** were found to have higher EA value. Furthermore, calculated HOMO and LUMO energies displayed good correlation with the SOD activities with *r* of 0.999 and 0.953, respectively. Such results corroborate the usefulness of HOMO and LUMO energies in relation to the SOD activity as it could account for the charge- or electron-transfer of the complex. These findings suggest great potential for the development of metallovitamin-based therapeutics.

Experimental

General

Melting points of the complexes were determined on the Griffin capillary melting point apparatus and are reported uncorrected. Infrared (IR) spectra were obtained on a Perkin Elmer System 2000 FTIR using potassium bromide (KBr) pellet. Mass spectra were recorded on a Bruker Daltonics (MicroTOF) instrument. Magnetic moments were measured with a Mark 1 Magnetic Susceptibility Balance (Sherwood Scientific, Cambridge, UK). Nicotinic acid, phthalic acid, salicylic acid, and anthranilic acid were of analytical grade and purchased from Sigma-Aldrich, as was bovine erythrocyte superoxide dismutase.

Synthesis of mixed-ligand complexes of copper with nicotinic acid and phtalic acid (1)

Nicotinic acid (0.123 g, 1 mmol) dissolved in methanol (30 mL) was heated (70°C) and stirred under reflux until a clear solution was obtained. Cupric chloride dihydrate (0.170 g, 1 mmol) dissolved in methanol (2 mL) and then added dropwise to the prepared nicotinic acid solution. After heating for 45 min, the reaction mixture was added dropwise to a solution of phthalic acid (1 mmol, 0.166 g) in methanol (2 mL), then, the mixture was heated for 1 h under the same conditions. The precipitated solid was collected by filtration, washed with cold methanol and dried under vacuum over silica gel at room temperature. Complex **1** was obtained as an aquamarine powder; m.p. > 300°C; IR (KBr, cm^{-1}): 2,565-3,080 (m, OH), 1,685 (s, CO), 1,701 (s, CO), 1,718 (s, CO), 1,418 (m, OH), 1,333 (w, CN), 1,298 (m, C-O); HRMS (TOF) calculated for $C_6H_5NO_2$ $[M+H]^+$ 124.0399, found: 124.0399.

Synthesis of mixed-ligand complexes of copper with nicotinic acid and salicylic acid (2)

Complex **2** was prepared in the same way as complex **1** using nicotinic acid (0.123 g, 1 mmol), cupric chloride (0.170 g, 1 mmol) and 1 mmol (0.138 g) of salicylic acid. Product **2** was obtained as an aquamarine powder by filtration, washing with cold methanol and dried over silica gel; m.p. > 300°C; IR (KBr, cm^{-1}): 2,567 (sh, w, OH), 2,682 (sh, w, OH), 1,717 (s, CO), 1,700 (s, CO), 1,685 (s, CO),

1,454 (m, OH), 1,420 (m, OH), 1,332 (m, CN), 1,298 (s, C-O), 1,201 (w, C-O), 1,144 (w, C-O), 1,126 (w, C-O); HRMS (TOF) calculated for $C_6H_5NO_2$ $[M+H]^+$ 124.0399, found: 124.0399.

Synthesis of mixed-ligand complexes of copper with nicotinic acid and anthranilic acid (**3**)

Complex **3** was prepared by an analogous procedure, but changing from Ph and/or Sal to anthranilic acid (1 mmol, 0.137 g) as a ligand. The precipitate of complex **3** was received as an aquamarine powder; m.p. > 300°C; IR (KBr, cm^{-1}): 3,275 (m, NH), 3,400 (br, OH), 2,681 (w, OH), 2,565 (w, OH), 1,718 (s, CO), 1,700 (s, CO), 1,608 (s), 1,558 (vs), 1,458 (m, OH), 1,420 (m, OH), 1,386 (s, CN), 1,332 (m, CN), 1,298 (s, C-O); HRMS (TOF) calculated for $C_6H_5NO_2$ $[M+H]^+$ 124.0399, found: 124.0396.

Determination of superoxide dismutase (SOD)-like activity

The complexes were tested for SOD activity using the previously described method [8]. The SOD activity of copper complexes was assayed by measuring the inhibition of the photoreduction of NBT. This indirect assay is comprised of several reactions: the photochemically excited riboflavin was first reduced by methionine to a semiquinone, which donated an electron to oxygen to form the superoxide source. The superoxide readily converted NBT into a purple formazan product, which was spectrophotometrically detected at 550 nm. In this regard, the SOD activity was inversely related to the amount of formazan formed, and expressed in term of IC_{50} of NBT reduction.

Antimicrobial activity

The antimicrobial activity of the complexes was investigated against representative microorganisms using the method previously reported [40]. One milliliter of Müeller Hinton (MH) broth was mixed with the tested complexes dissolved in DMSO. The MH broth of the tested sample was then mixed with MH agar solution and placed onto the plates with the final concentrations (64, 128 and 256 μg/mL) as agar dilution. The microorganisms, cultured in the MH broth at 37 °C for 24 h, were diluted with 0.9% normal saline solution to 3×10^8 cell/mL. The organisms were inoculated onto each plate and incubated at 37 °C for 18-48 h. Complexes found to be effective against the tested strains were selected for further investigations. The inhibition of microbial cell growths was also determined. Twenty-seven strains of microorganisms were used as shown in Table 8.

Molecular modeling of superoxide dismutase mimics

The molecular models of the copper complexes were constructed with GaussView 3.09 [41] based on the IR and magnetic moment data which indicated that the coordination complexes were of distorted tetrahedral conformation. Full geometry optimizations without symmetry constraints were performed *in vacuo* using Becke's three-parameter hybrid Lee-Yang-Parr (B3LYP) [42] functional and the LANL2DZ [43-45] basis set under Gaussian 03W [46]. The LANL2DZ basis set was selected as it could handle high Z atoms [47], particularly atoms that are beyond the third row.

Electron affinity was computed by taking the difference of the total energies of the coordination complexes Cu(II) and Cu(I) as summarized by the following equation:

$$EA = TE_{Cu(I)} - TE_{Cu(II)} \quad (1)$$

where EA represents the electron affinity, $TE_{Cu(I)}$ represents the total energy of the Cu(I) coordination complex, $TE_{Cu(II)}$ represents the total energy of the Cu(II) coordination complex. HOMO and LUMO energies were derived from geometrically optimized Cu(I) coordination complex as it is the active form of the SOD mimetics.

Table 8. Organisms subjected to growth inhibition assays.

Microorganism	***Reference strain***	***Clinical isolate***
Gram-positive bacteria	*Staphylococcus aureus* ATCC 29213	*Streptococcus pyogenes* II
	Staphylococcus aureus ATCC 25923	*Bacillus cereus*
	Staphylococcus epidermidis ATCC 12228	*Listeria monocytogenes*
	Enterococcus faecalis ATCC 29212	
	Enterococcus faecalis ATCC 33186	
	Micrococcus lutens ATCC 10240	
	Bacillus subtilis ATCC 6633	
	Corynebacterium diphtheriae NCTC 10356	
Gram-negative bacteria	*Escherichia coli* ATCC 25922	*Shigella dysenteriae*
	Klebsiella pneumoniae ATCC 700603	*Salmonella enteritidis* type C
	Serratia marcescens ATCC 8100	*Morganella morganii*
	Salmonella typhimurium ATCC 13311	*Aeromonas hydrophila*
	Shewanella putrefaciens ATCC 8671	*Citrobacter freundii*
	Achromobacter xylosoxidans ATCC 2706	*Plesiomonas shigelloides*
	Pseudomonas aeruginosa ATCC 15442	
	Pseudomonas stutzeri ATCC 17587	
Yeasts	*Saccharomyces cereviseae* ATCC 2601	
	Candida albicans ATCC 90028	

Acknowledgements

We would like to thank the Faculty of Medical Technology, Mahidol University and the Faculty of Science, Srinakharinwirot University for facilities and supports as well as the Chulabhorn Research Institute for performing mass spectra measurements. T.S. is grateful for the graduate fellowship from The Ministry Staff Development Project by the Ministry of Education, Thailand. This project was also supported in part by the Young Scholars Research Fellowship from The Thailand Research Fund to T.P. (Grant No. MRG5080349) and C.N. (Grant No. MRG5080450) and the annual budget grant of Mahidol University (B.E. 2551-2555).

References

1. Depeint, F.; Bruce, W.R.; Shangari, N.; Mehta, R.; O'Brien, P.J. Mitochondrial function and toxicity: role of the B vitamin family on mitochondrial energy metabolism. *Chem. Biol. Interact.* **2006**, *163*, 94-112.
2. Hageman, G.J.; Stierum, R.H. Niacin, poly(ADP-ribose) polymerase-1 and genomic stability. *Mutat. Res.* **2001**, *475*, 45-56.
3. Carlson, L.A. Nicotinic acid: the broad-spectrum lipid drug. A 50th anniversary review. *J. Intern. Med.* **2005**, *258*, 94-114.
4. Kamanna, V.S.; Kashyap, M.L. Mechanism of action of niacin. *Am. J. Cardiol.* **2008**, *101*, S20-S26.
5. Gross, M., Lipids, Oxidation, and Cardiovascular Disease. In *Atherosclerosis and Oxidant Stress*. Holtzman, J.L., Ed. Springer: New York, **2008**; 79-95.
6. Chapman, M.J. How does nicotinic acid modify the lipid profile? *Eur. Heart. J. Suppl.* **2006**, *8*, F54-59.
7. Spasojevic, I.; Chen, Y.; Noel, T.J.; Yu, Y.; Cole, M.P.; Zhang, L.; Zhao, Y.; St Clair, D.K.; Batinic-Haberle, I. Mn porphyrin-based superoxide dismutase (SOD) mimic, MnIIITE-2-PyP5+, targets mouse heart mitochondria. *Free Radic. Biol. Med.* **2007**, *42*, 1193-1200.
8. Piacham, T.; Isarankura-Na-Ayudhya, C.; Nantasenamat, C.; Yainoy, S.; Ye, L.; Bulow, L.; Prachayasittikul, V. Metalloantibiotic Mn(II)-bacitracin complex mimicking manganese superoxide dismutase. *Biochem. Biophys. Res. Commun.* **2006**, *341*, 925-930.
9. Singh, R.; Rao, R.S.S. Synthesis and characterization of divalent manganese, cobalt, nickel, copper and zinc complexes with nicotinic acid. *J. Mol. Struct.* **1981**, *71*, 23-30.
10. el-Saadani, M.A.; Nassar, A.Y.; Abou el-Ela, S.H.; Metwally, T.H.; Nafady, A.M. The protective effect of copper complexes against gastric mucosal ulcer in rats. *Biochem. Pharmacol.* **1993**, *46*, 1011-1018.
11. Salama, R.H.; Nassar, A.Y.; Nafady, A.A.; Mohamed, H.H. A novel therapeutic drug (copper nicotinic acid complex) for non-alcoholic fatty liver. *Liver Int.* **2007**, *27*, 454-464.
12. Huh, H.S.; Lee, S.W. Copper coordination polymers containing pyridinecarboxylate and multicarboxylate: $[Cu_{1.5}(ina)(btcH)]H_2O$, $[Cu_2(ina)_2(bdc)_{0.5}(\mu_3\text{-}OH)]$, and [Cu(ina)(na)] (inaH = 4-pyridinecarboxylic acid, $btcH_3$ = 1,3,5-benzenetricarboxylic acid, $bdcH_2$ = 1,3-benzenedicarboxylic acid, naH = 3-pyridinecarboxylic acid). *J. Mol. Struct.* **2007**, *829*, 44-50.
13. Yeh, C.-W.; Suen, M.-C.; Hu, H.-L.; Chen, J.-D.; Wang, J.-C. Synthesis and structural characterization of two new chiral coordination polymers of $Cu(nicotinate)_2(H_2O)$ and $Co(nicotinate)_2$. *Polyhedron* **2004**, *23*, 1947-1952.
14. Dutta, S.; Padhye, S.; McKee, V. Structural characterization and SOD activity of copper-oxaprozinate. *Inorg. Chem. Commun.* **2004**, *7*, 1071-1074.
15. Kostyuk, V.A.; Potapovich, A.I.; Strigunova, E.N.; Kostyuk, T.V.; Afanas'ev, I.B. Experimental evidence that flavonoid metal complexes may act as mimics of superoxide dismutase. *Arch. Biochem. Biophys.* **2004**, *428*, 204-208.
16. Mahal, H.S.; Kapoor, S.; Satpati, A.K.; Mukherjee, T. Radical scavenging and catalytic activity of metal-phenolic complexes. *J. Phys. Chem. B* **2005**, *109*, 24197-24202.

17. Devereux, M.; O'Shea, D.; O'Connor, M.; Grehan, H.; Connor, G.; McCann, M.; Rosair, G.; Lyng, F.; Kellett, A.; Walsh, M.; Egan, D.; Thati, B. Synthesis, catalase, superoxide dismutase and antitumour activities of copper(II) carboxylate complexes incorporating benzimidazole, 1,10-phenanthroline and bipyridine ligands: X-ray crystal structures of $[Cu(BZA)_2(bipy)(H_2O)]$, $[Cu(SalH)_2(BZDH)_2]$ and $[Cu(CH_3COO)_2(5,6\text{-}DMBZDH)_2]$ ($SalH_2$ = salicylic acid; BZAH = benzoic acid; BZDH = benzimidazole and 5,6-DMBZDH = 5,6-dimethylbenzimidazole). *Polyhedron* **2007**, *26*, 4073-4084.
18. Harris, R.M.; Hawker, R.J.; Langman, M.J.; Singh, S.; Waring, R.H. Inhibition of phenolsulphotransferase by salicylic acid: a possible mechanism by which aspirin may reduce carcinogenesis. *Gut* **1998**, *42*, 272-275.
19. Schindler, U.; Strobel, H.; Schonafinger, K.; Linz, W.; Lohn, M.; Martorana, P.A.; Rutten, H.; Schindler, P.W.; Busch, A.E.; Sohn, M.; Topfer, A.; Pistorius, A.; Jannek, C.; Mulsch, A. Biochemistry and pharmacology of novel anthranilic acid derivatives activating heme-oxidized soluble guanylyl cyclase. *Mol. Pharmacol.* **2006**, *69*, 1260-1268.
20. Bo, S.; Durazzo, M.; Gambino, R.; Berutti, C.; Milanesio, N.; Caropreso, A.; Gentile, L.; Cassader, M.; Cavallo-Perin, P.; Pagano, G. Associations of dietary and serum copper with inflammation, oxidative stress, and metabolic variables in adults. *J. Nutr.* **2008**, *138*, 305-310.
21. Olivares, M.; Uauy, R. Copper as an essential nutrient. *Am. J. Clin. Nutr.* **1996**, *63*, 791S-796.
22. Koca, E.; Buyukasik, Y.; Cetiner, D.; Yilmaz, R.; Sayinalp, N.; Yasavul, U.; Uner, A. Copper deficiency with increased hematogones mimicking refractory anemia with excess blasts. *Leuk. Res.* **2008**, *32*, 495-499.
23. Twomey, P.J.; Reynolds, T.M.; Wierzbicki, A.S.; Viljoen, A. The relationship between serum copper and ceruloplasmin in routine clinical practice. *Int. J. Clin. Pract.* **2008**, *62*, 485-487.
24. Cramer, C.J. *Essentials of Computational Chemistry: Theories and Models*. 2nd ed.; John Wiley & Sons, Ltd.: West Sussex, U.K., **2004**.
25. Karawajczyk, A.; Drgan, V.; Medic, N.; Oboh, G.; Passamonti, S.; Novic, M. Properties of flavonoids influencing the binding to bilitranslocase investigated by neural network modelling. *Biochem. Pharmacol.* **2007**, *73*, 308-320.
26. Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Naenna, T.; Prachayasittikul, V. Prediction of bond dissociation enthalpy of antioxidant phenols by support vector machine. *J. Mol. Graph. Model.* **2008**, *27*, 188-196.
27. Cabelli, D.E.; Guan, Y.; Leveque, V.; Hearn, A.S.; Tainer, J.A.; Nick, H.S.; Silverman, D.N. Role of tryptophan 161 in catalysis by human manganese superoxide dismutase. *Biochemistry* **1999**, *38*, 11686-11692.
28. Han, P.; Chen, C.-Q.; Zhang, C.-L.; Song, K.-K.; Zhou, H.-T.; Chen, Q.-X. Inhibitory effects of 4-chlorosalicylic acid on mushroom tyrosinase and its antimicrobial activities. *Food Chem.* **2008**, *107*, 797-803.
29. de Oliveira, G.; Martin, J.M.L.; de Proft, F.; Geerlings, P. Electron affinities of the first- and second-row atoms: Benchmark ab initio and density-functional calculations. *Phys. Rev. A* **1999**, *60*, 1034-1045.
30. Ji, H.F.; Zhang, H.Y. A theoretical study on Cu(II) binding modes and antioxidant activity of mammalian normal prion protein. *Chem. Res. Toxicol.* **2004**, *17*, 471-475.

31. Ji, H.F.; Zhang, H.Y. A new strategy to combat Alzheimer's disease. Combining radical-scavenging potential with metal-protein-attenuating ability in one molecule. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 21-24.
32. Shen, L.; Zhang, H.-Y.; Ji, H.-F. Computational note on the SOD-like antioxidant potential of nicotine-copper(II) complexes. *J. Mol. Struct. (THEOCHEM)* **2007**, *817*, 161-162.
33. Branco, R.J.F.; Fernandes, P.A.; Ramos, M.J. Density-functional calculations of the Cu, Zn superoxide dismutase redox potential: The influence of active site distortion. *J. Mol. Struct. (THEOCHEM)* **2005**, *729*, 141-146.
34. Branco, R.J.F.; Fernandes, P.A.; Ramos, M.J. Cu, Zn Superoxide dismutase: distorted active site binds substrate without significant energetic cost. *Theor. Chem. Acc.* **2006**, *115*, 27-31.
35. Karelson, M.; Lobanov, V.S.; Katritzky, A.R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*, 1027-1044.
36. Guo, P.; Chen, W.; Song, J.; Cao, W.; Tian, C. A DFT study of the interaction between butein anion and metal cations (M = Mg^{2+}, Cr^{2+}, Fe^{2+}, and Cu^{2+}): Taking an insight into its chelating property. *J. Mol. Struct. (THEOCHEM)* **2008**, *849*, 33-36.
37. Sun, R.; Ge, J.; Yao, J.; Li, S.; Shen, H.; Gu, R. Predicting the binding capability of benzothiazoline-2-thione and its derivatives with gold: A DFT and FT-Raman combined studies. *Spectrochim. Acta A* **2008**, *71*, 1535-1539.
38. Li, Q.X.; Luo, Q.H.; Li, Y.Z.; Shen, M.C. A study on the mimics of Cu–Zn superoxide dismutase with high activity and stability: two copper(II) complexes of 1,4,7-triazacyclononane with benzimidazole groups. *Dalton Trans.* **2004**, 2329-2335.
39. Garbutt, J.T.; Morehouse, A.L.; Hanson, A.M. Food additives, metal binding properties of bacitracin. *J. Agr. Food Chem.* **1961**, *9*, 285-289.
40. Prachayasittikul, S.; Suksrichavalit, T.; Isarankura-Na-Ayudhya, C.; Ruchirawat, S.; Prachayasittikul, V. Antimicrobial and antioxidative activities of 1-adamantylthio derivatives of 3-substituted pyridines. *Excli J.* **2008**, *7*, 63-70.
41. Dennington II, R.; Keith, T.; Millam, J.; Eppinnett, K.; Hovell, W.L.; Gilliland, R. *GaussView, version 3.09*; Semichem, Inc.: Shawnee Mission, KS, **2003**.
42. Stephens, P.J.; Devlin, F.J.; Chabalowski, C.F.; Frisch, M.J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
43. Hay, P.J.; Wadt, W.R. *Ab initio* effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J. Chem. Phys.* **1985**, *82*, 270-283.
44. Hay, P.J.; Wadt, W.R. Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* **1985**, *82*, 299-310.
45. Wadt, W.R.; Hay, P.J. Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *J. Chem. Phys.* **1985**, *82*, 284-298.
46. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K.N.; Burant, J.C.; Millam, J.M.; Iyengar, S.S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.E.; Hratchian, H.P.; Cross, J.B.;

Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.E.; Yazyev, O.; Austin, A.J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Ayala, P.Y.; Morokuma, K.; Voth, G.A.; Salvador, P.; Dannenberg, J.J.; Zakrzewski, V.G.; Dapprich, S.; Daniels, A.D.; Strain, M.C.; Farkas, O.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Ortiz, J.V.; Cui, Q.; Baboul, A.G.; Clifford, S.; Cioslowski, J.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R.L.; Fox, D.J.; Keith, T.; Al-Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.M.W.; Johnson, B.; Chen, W.; Wong, M.W.; Gonzalez, C.; Pople, J.A. In Gaussian 03, Revision C.02, Gaussian, Inc.: Wallingford CT, **2004**.

47. Foresman, J.B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*. Gaussian, Inc.: Pittsburgh, **2000**.

Sample Availability: Contact authors.

© 2008 by the authors; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/)

European Journal of Medicinal Chemistry 44 (2009) 3259–3265

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Copper complexes of pyridine derivatives with superoxide scavenging and antimicrobial activities

Thummaruk Suksrichavalit [a], Supaluk Prachayasittikul [b,*], Chanin Nantasenamat [a], Chartchalerm Isarankura-Na-Ayudhya [a], Virapong Prachayasittikul [a,*]

[a] Department of Clinical Microbiology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[b] Department of Chemistry, Faculty of Science, Srinakharinwirot University, Bangkok 10110, Thailand

ARTICLE INFO

Article history:
Received 31 October 2008
Received in revised form
2 March 2009
Accepted 26 March 2009
Available online 5 April 2009

Keywords:
Nicotinic acid
Pyridine
Copper
Superoxide dismutase
Antioxidant activity
Antimicrobial activity

ABSTRACT

Superoxide anions are reactive oxygen species that can attack biomolecules such as DNA, lipids and proteins to cause many serious diseases. This study reports the synthesis of copper complexes of nicotinic acid with related pyridine derivatives. The copper complexes were shown to possess superoxide dismutase (SOD) and antimicrobial activities. The copper complexes exerted SOD activity in range of 49.07–130.23 μM. Particularly, copper complex of nicotinic acid with 2-hydroxypyridine was the most potent SOD mimic with an IC_{50} of 49.07 μM. In addition, the complexes exhibited antimicrobial activity against *Bacillus subtilis* ATCC 6633 and *Candida albicans* ATCC 90028 with MIC range of 128–256 μg/mL. The SOD activities were well correlated with the theoretical parameters as calculated by density functional theory at the B3LYP/LANL2DZ level of theory. Interestingly, the SOD activity of the copper complexes was demonstrated to be inversely correlated with the electron affinity, but was well correlated with both HOMO and LUMO energies. The vitamin–metal complexes described in this report are great examples of the value-added benefits of vitamins for medicinal applications.

© 2009 Elsevier Masson SAS. All rights reserved.

1. Introduction

Free radicals such as superoxide anions ($O_2^{-\bullet}$) are very reactive oxygen species (ROS), which are primarily produced in the human body and has been involved in many cellular processes, which causes various diseases [1]. The deleterious effects of ROS are counterbalanced by antioxidant enzymes such as superoxide dismutase (SOD), catalase and glutathione peroxidase as well as by small molecular antioxidants. Under oxidative stress conditions, ROS can induce molecular and cellular damages to DNA, lipids and proteins. Such events lead to the development of many serious diseases [2,3], e.g. cancer [4], atherosclerosis [5], asthma [6], and neurodegenerative disorders [7]. SOD, in particular, is the antioxidant enzyme, which can convert superoxide anion to a less-reactive form, e.g. H_2O_2 [8]. The utilization of SOD as a therapeutic agent has been reported, however there are many limitations, which hinders its application, such as its high molecular weight [9]. Therefore, small molecular weight SOD mimics as afforded by transition metal complexes are more lucrative alternatives. Some examples include metalloporphyrins [10], metal–drug complexes [11] and metallopyridine complexes [12].

Pyridine derivatives possess a diverse array of bioactivities as well as playing crucial roles for physiological functions. They have been extensively used as ligands in the formation of coordination compounds as medicinal agents. The most striking one to our interest is nicotinic acid (NA) or pyridine-3-carboxylic acid, also known as vitamin B3 (niacin), which acts as an antihyperlipidemic drug and as effective HDL cholesterol raising agent for reduction of cardiovascular risks [13,14]. Complexation of NA with various metals, e.g. manganese, cobalt, nickel, copper and zinc, had previously been reported [15], where nicotinic acid–copper complex has particularly been shown to stimulate blood flow and prevent gastric congestion [16], reduce total lipids in sera hepatic tissue as well as reducing the levels of ALT, AST, ALP, GGT and oxidative markers, e.g. nitric oxide (NO) and lipid peroxidation. Such complex was also demonstrated to exhibit SOD activity for treatment of patients with hepatocellular carcinoma [17]. Recently, we have reported promising SOD mimic activity for copper complexes of nicotinic-carboxylic acids [12]. Such copper complexes were formed by coordination of NA via the ring N-atom with bidentate carbonyl and oxy/amino groups of carboxylic acids as electron donors. Aside from NA, other interesting ligands used in this study bearing suitable N- and/or O-electron donors are bioactive pyridine derivatives and related compounds such as hydroxypyridines,

* Corresponding authors. Tel.: +66 2 418 0227; fax: +66 2 412 4110.
E-mail addresses: supaluk@swu.ac.th (S. Prachayasittikul), mtvpr@mahidol.ac.th (V. Prachayasittikul).

0223-5234/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.03.033

aminopyridines, and picolinic acid (pyridine-2-carboxylic acid). Picolinic acid, an isomeric form of NA, is a metabolite derived from tryptophan of humans and animals. The picolinic acid plays an important role in zinc transport [18,19]. Its complexes have been demonstrated to possess antiviral [20], antifungal [21] and antibacterial [22] activities as well as being involved in the induction of apoptosis [23] and immune responses [24]. 2-Hydroxypyridines (2-pyridone) have been reported to form complex with metals [25]. The complex of hydroxypyridine bases, such as 2-, 3- and 4-hydroxypyridines, with platinum or palladium was shown to possess promising anticancer activity against human ovarian cancer cells [26–28]. Moreover, 4-aminopyridines and their derivatives were reported to be involved in the inactivation of voltage-dependent K^+ channels [29,30] as well as implicated in reversing anesthesia and relaxing muscles [31,32]. The copper ion is an essential element, which possesses interesting anti-inflammatory and anti-ulcer activities [33]. As a metal cofactor in the copper–zinc superoxide dismutase, copper exerts important catalytic function in the dismutation of superoxide anions [34,35]. It is known that copper deficiency is responsible for hematological disorders, hypopigmentation, defective connective tissue cross-linking and ataxia [36,37].

The aforementioned statements have motivated us, as part of a continuous study, to synthesize low molecular weight copper complexes of biologically important pyridines as SOD mimics. The nicotinic acid was used as a primary ligand for such synthesis. Therefore, novel copper complexes of nicotinic acid with 2-substituted pyridines containing bidentate N, O and/or N, N electron donors were synthesized as SOD mimic activity. The 2-substituted pyridines are 2-hydroxypyridine, 2-aminopyridine and pyridine-2-carboxylic acid. To give insight into the cardinal mechanisms of SOD activities, quantum chemical parameters calculated at density functional theory (DFT) level were used as physicochemical description of the compounds. DFT has wide applications in the life sciences particularly in the development of quantitative structure–activity/property relationship (QSAR/QSPR) models, for the elucidation of enzymatic reactions as well as in clarification of the superoxide radical scavenging activity. Results indicated that the calculated physicochemical parameters were well correlated with the experimental SOD activities [12]. The theoretical parameters used are comprised of energies of the highest occupied molecular orbital (HOMO), energies of the lowest unoccupied molecular orbital (LUMO), and electron affinity (EA). Moreover, the antimicrobial activity of the novel complexes was also explored.

2. Materials and methods

2.1. Materials

All chemicals were of analytical reagent grade and were purchased from Sigma–Aldrich. Bovine erythrocyte superoxide dismutase was supplied from the same source. Melting points of the complexes were determined on the Griffin capillary melting point apparatus and are reported without correction. Infrared (IR) spectra were obtained on Perkin Elmer System 2000 FTIR as a potassium bromide (KBr) pellet. Magnetic moment was performed by Magnetic Susceptibility Balance, Mark 1, Serial 15257, Sherwood Scientific, Cambridge, UK.

2.2. Synthesis of copper complexes of pyridine derivatives

2.2.1. Copper complex of nicotinic acid-2-hydroxypyridine (**1**)

Nicotinic acid (NA, 0.123 g, 1 mmol) dissolved in methanol (30 mL) was heated (70 °C) and stirred under reflux until the solution was clear. Cupric chloride dihydrate (0.170 g, 1 mmol) was dissolved in methanol (2 mL) and added dropwise to the prepared NA solution. After heating for 45 min, a solution of 2-hydroxypyridine (2Hy, 1 mmol, 0.138 g) in methanol (2 mL) was added dropwise to the reaction mixture. The mixture was then heated for 1 h at the same condition. The precipitated solid was collected by filtration, washed with cold methanol and dried *in vacuo* over silica gel at room temperature to yield cyan powder of complex **1**, m.p. >300 °C; IR (KBr, cm^{-1}): 1389 (m, C–N), 1375 (m, C–N), 1191 (w, C–O), 1093 (w, C–O), 1053 (w, C–O); μ_{eff}: 1.5545 BM.

2.2.2. Copper complex of nicotinic acid-2-aminopyridine (**2**)

Complex **2** was prepared in the same method as complex **1** using NA (0.123 g, 1 mmol), cupric chloride (0.170 g, 1 mmol) and 1 mmol (0.137 g) of 2-aminopyridine (2Am). As aquamarine powder, product **2** was obtained by filtration, washed with cold methanol and dried over silica gel, m.p. >300 °C; IR (KBr, cm^{-1}): 1386 (s, C–N), 1375 (s, C–N), 1191 (s, C–O), 1155 (w, C–O), 1093 (m, C–O), 1050 (m, C–O); μ_{eff}: 1.9381 BM.

2.2.3. Copper complex of nicotinic acid–picolinic acid (**3**)

Copper complex of the nicotinic–picolinic acids (**3**) was prepared in an analogous fashion using 1 mmol of NA, cupric chloride and picolinic acid (Pi; 1 mmol, 0.166 g). The precipitated complex was obtained as blue powder after filtration, washing with cold methanol, and dried over silica gel. m.p. >300 °C; IR (KBr, cm^{-1}): 1475 (w, O–H), 1348 (vs, C–N), 1286 (s, C–O), 1152 (w, C–O), 1049 (s, C–O); μ_{eff}: 1.7014 BM. Structures of pyridine derivatives are shown in Fig. 1.

2.3. Determination of superoxide dismutase activity

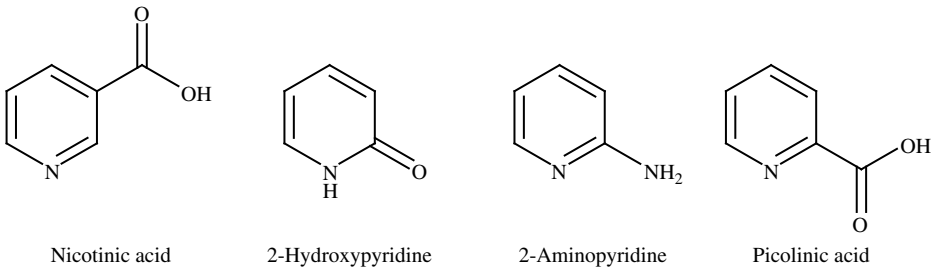
The complexes were tested for SOD activity using the previously described method [11,38]. The SOD activity of copper complexes was assayed in the pH range of 6.9–7.4 by measuring the inhibition of the photoreduction of nitro blue tetrazolium (NBT). This indirect assay is comprised of several reactions: the photochemically excited riboflavin was first reduced by methionine into a semiquinone, which donated an electron to oxygen to form a superoxide source. The superoxide readily converted NBT into a purple formazan product, being spectrophotometrically detected at 550 nm. In this regard, the SOD activity was inversely related to the amount of formazan formed, and expressed in terms of 50% inhibition concentration (IC_{50}) of NBT reduction.

2.4. Antimicrobial activity

The antimicrobial activity of the complexes was investigated against representative microorganisms using the method previously reported [38,39]. Briefly, the tested complexes dissolved in dimethyl sulfoxide (DMSO) were mixed with Müller Hinton (MH) broth to receive a final volume of 1 mL. Two-fold dilution was prepared and the MH broth of the tested sample then was transferred to MH agar solution to yield the final concentrations ranging from 64 to 256 μg/mL. The solutions were placed onto the plates. Twenty-seven strains of microorganisms (Table 1), cultured in the MH broth at 37 °C for 24 h, were diluted with 0.9% normal saline solution to 3×10^8 cells/mL. The organisms were inoculated onto each plate and incubated at 37 °C for 18–48 h. The complexes found to be effective against the tested strains were selected. The inhibition of microbial cell growth was determined.

2.5. Molecular model of superoxide dismutase mimics

The copper complexes were drawn with GaussView 3.09 [40] and calculated by Gaussian 03 [41]. Fullyoptimized geometry without symmetry constraints of the copper complexes was



Nicotinic acid 2-Hydroxypyridine 2-Aminopyridine Picolinic acid

Fig. 1. Chemical structures of the pyridine derivatives.

performed *in vacuo* using density functional theory (DFT) at B3LYP/LANL2DZ level [42].

3. Results and discussions

Copper-based pyridine complexes **1**–**3** were synthesized by the reaction of cupric chloride with nicotinic acid and pyridine derivatives (2-hydroxypyridine, 2-aminopyridine and picolinic acid) in a molar ratio of 1:1:1. The complexes were obtained in good yields (75–87%) as cyan, aquamarine and blue powder, respectively. They are highly polar compounds, insoluble in methanol and water, but soluble in DMSO. Their melting points and magnetic moment are shown in Table 2. IR spectral data are outlined in Table 3.

3.1. Infrared spectra

From the IR spectral data, it can be seen that the C–N stretching vibration of complex **1** showed at 1389 and 1375 cm^{-1} whereas the C–N absorption of free NA and free 2Hy appeared at 1324 and 1243 cm^{-1}, respectively. It was observed that the hydroxyl absorption of NA at 1418 cm^{-1} δ(O–H) disappeared. In addition, strong absorptions at 1575 and 781 cm^{-1} δ(N–H) for 2Hy were not observed. The C–O stretching vibration of complex **1** displayed weak bands at 1191, 1093 and 1053 cm^{-1}, whereas the C–O absorptions of free NA appeared at 1299 cm^{-1} and for 2Hy at 1156 and 1098 cm^{-1}. Based on the IR spectra, the carbonyl absorption of complex **1** could not be observed. It is suggested that complex **1** was formed by coordination of copper atom with the two carbonyl groups (carboxylic of NA and 2-pyridone) and the two nitrogen atoms of pyridine ring of both NA and 2Hy.

Table 1
Microorganisms for antimicrobial assays.

Microorganism	Reference strain	Clinical isolates
Gram-positive bacteria	*Staphylococcus aureus* ATCC 29213	*Streptococcus pyrogenes* II
	Staphylococcus aureus ATCC 25923	*Bacillus cereus*
	Staphylococcus epidermidis ATCC 12228	*Listeria monocytogenes*
	Enterococcus faecalis ATCC 29212	
	Enterococcus faecalis ATCC 33186	
	Micrococcus lutens ATCC 10240	
	Bacillus subtilis ATCC 6633	
	Corynebacterium diphtheriae NCTC 10356	
Gram-negative bacteria	*Escherichia coli* ATCC 25922	*Shigella dysenteriae*
	Klebsiella pneumoniae ATCC 700603	*Salmonella enteritidis* type C
	Serratia marcescens ATCC 8100	*Morganella morganii*
	Salmonella typhimurium ATCC 13311	*Aeromonas hydrophila*
	Shewanella putrefaciens ATCC 8671	*Citrobacter freundii*
	Achromobacter xylosoxidans ATCC 2706	*Plesiomonas shigelloides*
	Pseudomonas aeruginosa ATCC 15442	
	Pseudomonas stutzeri ATCC 17587	
Yeasts	*Saccharomyces cerevisiae* ATCC 2601	
	Candida albicans ATCC 90028	

The IR spectra of complex **2** pointed out the disappearance of carbonyl absorption, while the free NA exhibited absorption at 1718 and 1700 cm^{-1}. The amino (NH_2) absorption of 2Am appeared at 3448 and 3315 cm^{-1}, while the NH absorption of **2** was shown at 3116 cm^{-1}. Strong C–N stretching vibration of complex **2** was observed at 1386 and 1375 cm^{-1}. Such C–N absorption of NA appeared at 1324 cm^{-1} and at 1492, 1442 and 1326 cm^{-1} for 2Am. The strong C–O stretching of free NA was observed at 1299 cm^{-1}, weak hydroxyl (O–H) stretching was noted at 3072–2449 cm^{-1} and strong OH bending displayed at 1418 cm^{-1}. The C–O absorption of complex **2** was observed at 1191, 1155, 1093 and 1050 cm^{-1}. As evidence by the IR spectra, complex **2** was formed by coordination of the carbonyl group and the pyridine ring nitrogen atom of NA with the amino group and nitrogen ring atom of 2Am.

Similarly, the IR spectra of complex **3** were determined by comparing with the free ligands. It can be seen that the frequency of C–O stretching vibration was presented at 1286 cm^{-1} υ(C–O), while the absorption of free NA and Pi appeared at 1299 and 1294 cm^{-1}, respectively. Furthermore, disappearances of the stretching vibration at 1718 and 1700 cm^{-1} of the carbonyl groups υ(C=O) of both NA and Pi were observed. The weak band at 1475 cm^{-1} could be attributed to the O–H bending vibration of complex **3** when compared with the spectra of both NA and Pi, which showed strong bands at 1418 and 1454 cm^{-1}, respectively. The C–N stretching vibration of complex **3** exhibited very strong band at 1348 cm^{-1}, while the absorption of free NA and Pi showed υ(C–N) at 1324 and 1342 cm^{-1}, respectively. Hence, complex **3** was formed using carbonyl groups and nitrogen atoms of pyridine ring of NA and Pi as bidentate ligands to coordinate with copper atom.

It is observed that the copper complexes were formed by the coordination of NA as a primary bidentate ligand with 2-substituted pyridines (2Hy, 2Am and Pi). The bidentate NA used carbonyl moiety and ring nitrogen atom as electron donating group, which caused the C–N shift from 1324 cm^{-1} to higher frequency range of 1348–1389 cm^{-1}. Apparently, the carbonyl absorptions of complexes **1**–**3** were not observed. Likewise, the 2-substituted pyridine was coordinated to the central copper atom as bidentate ligands. In this regard N–H bending of **1** at 781 and 1575 cm^{-1} disappeared with a shift of C–O stretching from 1156 cm^{-1} to higher frequency. This suggested that the copper

Table 2
Physicochemical parameters of pyridine–copper complexes and its ligands.

Compound	Chemical formulae	Formula weight (g/mol)	Color	Melting point (°C)	Yield (%)	μ_{eff} (BM)
NA	$C_6H_5NO_2$	123.11	White	236–239	–	–
2Hy	C_5H_5NO	95.10	White	105–107	–	–
1	$C_{11}H_{10}CuN_2O_4$	297.75	Cyan	>300	84	1.5545
2Am	$C_5H_6N_2$	94.11	White	54–58	–	–
2	$C_{11}H_{11}CuN_3O_3$	279.76	Aquamarine	>300	87	1.9381
Pi	$C_6H_5NO_2$	123.11	White	139–142	–	–
3	$C_{12}H_{10}CuN_2O_4$	309.76	Blue	>300	75	1.7014

μ_{eff}, Magnetic moment; B.M., Bohr magnetron.

Table 3
IR spectra of pyridine-copper complexes and its ligands.

Compound	υC=O	υC–O	υC–N	υO–H	δO–H	υNH	δNH
NA	1718(s), 1700 (s)	1299(s)	1324(s)	3072–2449(w)	1418(s)	–	–
2Hy	1690(s), 1652(vs)	1156(s), 1098(m)	1243(s)	–	–	–	1575(br, s), 781(vs)
1	–	1191(w), 1093(w), 1053(w)	1389(m), 1375(m)	3400(br, w)	1458(m)	–	–
2Am	–	–	1492(s), 1442(s), 1326(m)	–	–	3448(sh, s), 3315(sh, w)	772.84(s)
2	–	1191(s), 1155(w), 1093(m), 1050(m)	1386(s), 1375(s)	3400(br, w)	1426(m)	3116(sh, w)	760.72(s)
Pi	1718(w), 1700(w)	1294(s), 1165(w), 1086(w)	1342(m)	2609(br, w), 2362(m), 2341(m)	1454(s)	–	–
3	–	1286(s), 1152(w), 1049(s)	1348(vs)	3400(br, w)	1475(w)	–	–

Note: vs = Very strong, s = strong, m = medium, w = weak, br = broad, sh = shape.

atom of complex **1** formed a complex by means of the bidentate amino ketone (2-pyridone). Similarly, the coordination of the copper atom from complex **2** is made possible via the amino-pyridine (2Am) ligand, which is evident by the disappearance of NH_2 stretching absorptions and the shift of C–N stretching. As for complex **3**, Pi is also a bidentate ligand which uses the ring nitrogen atom and the carbonyl group to coordinate with the copper atom.

Magnetic susceptibility determination showed that complexes **1**–**3** were paramagnetic with a magnetic moment (μ_{eff}) of 1.55, 1.94 and 1.70 BM, respectively. The μ_{eff} value indicates that the complexes were of tetrahedral geometry with a copper center. Based on the IR spectra and μ_{eff}, we can confirm that complexes **1**–**3** were of copper tetrahedral geometry.

3.2. Superoxide scavenging activity

Complexes **1**–**3** were assayed for SOD activity using the modified approach [11,38] which involved measuring the inhibition of NBT reduction. The results (Table 4) revealed that all of the tested complexes exhibited SOD activity with IC_{50} in the range of 49.07–130.23 μM. Complex **1** was the most powerful superoxide scavenger with an IC_{50} value of 49.07 μM. In addition, the SOD activity of complex **2** was higher than that of complex **3** with IC_{50} values of 50.32 and 130.23 μM, respectively. The SOD activity of these complexes are shown to be in the order of **1** > **2** > **3**. It should be noted that the activity of complex **2** was comparable to that of **1** as observed from the IC_{50} values of 49.07 and 50.32, respectively. Furthermore, the SOD activity of the free ligands NA, 2Hy, 2Am and Pi were elucidated and was found to be very low to the point that the IC_{50} value could not be obtained. The result strongly suggests that the copper atom is of prime importance for the SOD activity as its coordination to the ligands profoundly increased the SOD activity. Such phenomenon can be attributed to the change in oxidation state of the copper atom upon coordination with the ligands. Fascinatingly, these copper complexes were significantly small in molecular weight than the native enzyme. Moreover, these copper coordination compounds are easily formed from bioactive molecules such as nicotinic acid.

Table 4
Superoxide dismutase activity of copper complexes.

Compound[c]	IC_{50} (μM)[a]
1	49.07
2	50.32
3	130.23
SOD[b]	0.0026

[a] IC_{50} was defined as fifty percent inhibition concentration of NBT reduction.
[b] Superoxide dismutase from bovine erythrocytes was a homodimeric protein.
[c] Substituted pyridines (ligands) mimic very low activity that the IC_{50} cannot be obtained.

3.3. Antimicrobial activities

Nicotinic acid has been widely used for hyperlipidemic treatment. The biological activities of 2-hydroxypyridine, 2-amino-pyridine and picolinic acid have previously been reported [19,27,30]. The exploration of antimicrobial activities for the complexes of nicotinic acid and its derivatives is a relatively new endeavor. Results (Table 5) demonstrated that complexes **1**–**3** could completely inhibit the growth of *Bacillus subtilis* ATCC 6633 with MIC of 256 μg/mL. In addition, the tested complexes **1**–**3** could partially inhibit the growth of *Candida albicans* ATCC 90028 with MIC in the range of 128–256 μg/mL. Such augmentation in the observed antimicrobial activity of the metal complexes over those of the free ligands could be attributed to the chelation effect [43,44]. As a result, this alters the lipophilic property of the central metal ion thereby allowing its permeation through the cell membrane to exert its antimicrobial activity [45]. In parallel with

Table 5
Antimicrobial activity of pyridine–copper complexes.

Compound	Concentration (μg/mL)	Activity
Ampicillin	25	Active[a]
NA	256	Not active
$CuCl_2$	64	Not active[h]
	32	Not active[h]
2Hy	256	Not active
1	256	Active[b,e]
	128	Active[c, f]
	64	Not active
2Am	256	Not active
2	256	Active[b, f]
	128	Active[d, g]
	64	Not active
Pi	256	Not active
3	256	Active[b, f]
	128	Active[d, g]
	64	Not active

[a] Active (100% antigrowth) against *P. aeruginosa* ATCC 15442, *S. putrefaciens* ATCC 8671, *A. xylosoxidans* ATCC 2706, *S. aureus* ATCC 25923, *S. epidermidis* ATCC 12228, *E. faecalis* ATCC 29212, *B. subtilis* ATCC 6633, *S. cerevisiae* ATCC 2601, *S. pyrogenes* II, *S. enteridis* type C, *P. shigelloides*, *L. monocytogenes*.
[b] Active against *B. subtilis* ATCC 6633 with 100% antimicrobial activity.
[c] Active against *B. subtilis* ATCC 6633 with 75% antimicrobial activity.
[d] Active against *B. subtilis* ATCC 6633 with 50% antimicrobial activity.
[e] Active against *C. albicans* ATCC 90028 with 75% antimicrobial activity.
[f] Active against *C. albicans* ATCC 90028 with 50% antimicrobial activity.
[g] Active against *C. albicans* ATCC 90028 with 25% antimicrobial activity.
[h] $CuCl_2$ was tested against *B. subtilis* ATCC 6633 and *C. albicans* ATCC 90028.

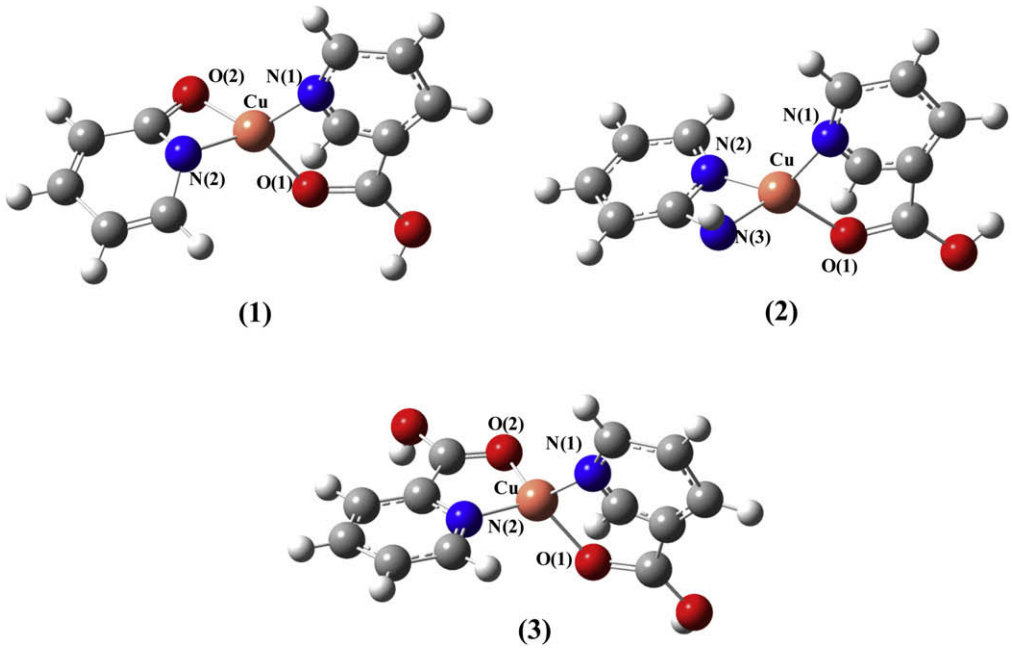


Fig. 2. Geometrically optimized structures of complexes **1**–**3** calculated at B3LYP/LANL2DZ level.

our results, previous studies have found that the free ligands used in this study, comprising of 2-hydroxypyridine [46], 2-aminopyridine [47], and picolinic acid [24], possessed antibacterial, antifungal, and antiviral activities. To discern whether the observed antimicrobial activities arose from the ligand themselves or as a result of coordinating with the metal ions, the biological activities of the uncomplexed ligands were examined. Antimicrobial activities were observable only in the metal complexes and not in the uncomplexed ligands, therefore confirming the participating role of the metal ion in the biological activities.

3.4. Theoretical studies of SOD mimics by computational chemistry

Quantum chemical parameters of the metal complexes were calculated using the minimum energy conformers in order to deduce quantitative relationships of the theoretical descriptors with the superoxide scavenging activity by performing Density Functional Theory calculations at B3LYP/LANL2DZ level. The fully optimized structures of complexes **1**–**3** were of tetragonally distorted geometry as shown in Fig. 2. The calculated theoretical parameters, e.g. binding energy (BE) and electron affinity (EA), has previously been reported to be good descriptors for characterizing superoxide radical scavenging activity [42,48–50]. It is well established that EA is a suitable parameter in accounting for the rate of electron transfer from superoxide anion to copper ion. Ji and Zhang previously reported that compounds with the lowest EA possessed the highest electron transfer rate, thereby conferring the highest SOD activity [50]. The calculated results for EA as shown in Table 5 were in the following order **1** > **2** > **3** and their respective values are −100.40, −109.39, and −162.71 kcal/mol, respectively. The observed trend is positively correlated with the SOD activity where complex **1** had the highest SOD activity as well as the highest calculated EA. Such findings contradict with those previously reported by Branco et al. [51,52] where an inverse relationship was observed for the calculated EA and the SOD activity. A plausible explanation for such discrepancy can be made on the basis of differences in the coordination geometry of the copper coordination complexes used in this study (tetrahedral) and those previously reported by Ji et al. (square planar and square pyramidal) [49,50]. Such observation is further corroborated by previous studies on active site distortion of CuZnSOD [51,52].

Examination of the energy level of HOMO revealed that complex **3** possessed the highest HOMO energy with a value of −3.32 eV while coordination compounds **2** and **1** gave HOMO energies of −4.65 and −5.09 eV, respectively (Table 6). Higher HOMO energy as deduced from the frontier molecular orbital indicates that the molecule can readily transfer electron to the unoccupied orbital of the metal ion thereby indicating strong binding capacity [53]. Such molecules displaying high binding capacity was demonstrated to also exhibit lower SOD activity [54,55]. This coincides with our findings that complex **1** possessed the highest SOD activity as well as having the lowest HOMO energy and vice versa for complex **3** which exhibited the lowest SOD activity while having the highest HOMO energy. An analysis of the calculated energy level for the LUMO indicates the order of **3** > **2** > **1** for the coordination compounds bearing values of −2.61, −2.96, and −2.99, respectively. It was previously reported by Schepetkin et al. [55] that low LUMO energy was important for superoxide dismutation. Moreover, the calculated energies of HOMO and LUMO were significantly correlated with the SOD activity which was represented as $-\log(1/IC_{50})$, yielding r of 0.97 and 0.99, respectively. The results clearly suggest that the energies of HOMO and LUMO are potentially useful theoretical parameters for elucidating superoxide scavenging activity, a phenomenon involving charge- or electron-transfer.

In characterizing the physicochemical properties of the metal complexes, it is worthy to consider the tendency for nicotinic acids to dissociate a proton from the carboxylic acid moiety to form a carboxylate anion. Such notion is valid when nicotinic

Table 6
Quantum chemical parameters of copper complexes calculated at B3LYP/LANL2DZ level.

Complex	TE_{CuII} (hartree)	TE_{CuI} (hartree)	EA[a] (kcal/mol)	HOMO (eV)	LUMO (eV)
1	−955.7254	−955.8854	−100.40	−5.09	−2.99
2	−935.8071	−935.9814	−109.39	−4.65	−2.96
3	−1069.5318	−1069.7911	−162.71	−3.32	−2.61

[a] $EA = TE_{Cu(I)} - TE_{Cu(II)}$.

Table 7
Selected bond lengths and angles of complexes **1**–**3**.

Complex	Bond length (Å)		Angle (°)	
1	Cu–N(1)	1.8013	N(1)–Cu–N(2)	124.420
	Cu–O(1)	2.0100	N(1)–Cu–O(1)	90.624
	Cu–N(2)	1.9076	N(1)–Cu–O(2)	124.119
	Cu–O(2)	1.8561	N(2)–Cu–O(1)	123.331
			N(2)–Cu–O(2)	74.389
			O(1)–Cu–O(2)	124.407
2	Cu–N(1)	1.8014	N(1)–Cu–N(2)	124.390
	Cu–O(1)	2.0105	N(1)–Cu–N(3)	126.729
	Cu–N(2)	1.8336	N(1)–Cu–O(1)	89.214
	Cu–N(3)	1.8542	N(2)–Cu–N(3)	73.359
			N(2)–Cu–O(1)	125.380
			N(3)–Cu–O(1)	122.785
3	Cu–N(1)	1.7990	N(1)–Cu–N(2)	116.156
	Cu–O(1)	2.0031	N(1)–Cu–O(1)	92.119
	Cu–N(2)	1.8255	N(1)–Cu–O(2)	122.541
	Cu–O(2)	1.8221	N(2)–Cu–O(1)	122.623
			N(2)–Cu–O(2)	91.883
			O(1)–Cu–O(2)	114.330

acids are present as free ligands. However such situation may not apply in our case since the ligand is coordinated to a metal ion. Such finding coincides with previous observation that metal coordination exerts direct effect on the ligand's acidity [56]. Correspondingly, Bóka et al. had observed in their efforts to mimic the SOD activity that strong electron donating ability of ligands is detrimental towards SOD activity [57]. The coordination complexes was shown to possess good SOD activity which implied that the coordinated nicotinic acids should be in the neutral form whereas the anionic form are expected to possess greater electron donating properties, which would consequently hinder its reactivity with superoxide anions. Similar findings were also observed in our previous investigation on mixed ligand complexes of copper with nicotinic and other selective carboxylic acids (phthalic, salicylic and anthranilic acids) [12]. It was shown that the complex constituting stronger electron withdrawing group exhibited better SOD activity.

The contributions of metal–ligand binding affinity on the observed SOD activity was investigated using information such as bond lengths and angles as derived from geometrically optimized structures at B3LYP/LANL2DZ level (Table 7). It was observed that the mean bond distances at the axial position for Cu–N(1) and Cu–O(1) of complexes **1**–**3** were 1.801 and 2.008 Å, respectively. The longest axial distance for Cu–N(2) and Cu–O(2) bonds corresponds to that of complex **1** with 1.908 and 1.856 Å, respectively. The bond lengths for Cu–N(2) and Cu–N(3) at axial position for complex 2 were observed to be 1.834 and 1.854 Å, respectively, which was shorter than that of complex **1**. This is attributed to the fact that the amino group of 2-aminopyridine from complex **2** possessed greater electron donating capability than the hydroxyl group of 2-hydroxypyridine from complex **1** as indicated by the higher HOMO energy of **2** (Table 6). The shortest axial bond lengths for Cu–N(2) and Cu–O(2) were observed in complex **3** which gave values of 1.826 and 1.822 Å, respectively. This result coincides with the fact that complex **3** possessed the lowest SOD activity which could be attributed to the higher metal binding affinity. Previous study by Li et al. has reported that long axial bond lengths were critical for the dismutation of superoxide anion [58]. Interestingly, complex **1** which had the longest axial bond length was also shown to possess the highest SOD activity. The calculated bond angles for the three coordination complexes further corroborate a distorted tetragonal geometry with bond angles in the range of 73.35°–126.73°.

4. Conclusion

We report the synthesis of novel copper complexes **1**–**3** with nicotinic acid as the primary ligand along with 2-substituted pyridines: 2-hydroxypyridine, 2-aminopyridine and picolinic acid. The experimentally derived magnetic moments and the calculated bond angle suggest a tetragonally distorted geometry for the coordination complexes. Nicotinic acid is a bidentate ligand, which utilizes the ring N-atom of pyridine and the carbonyl group of carboxylic acid. Likewise, the 2-substituted pyridines are also bidentate ligands making use of the following functional groups: amino ketone of 2-pyridone (2Hy), aminopyridine (2Am) and carbonyl pyridine (Pi), respectively. All of the coordination compounds used in this study displayed promising superoxide radical scavenging properties as well as antimicrobial activities against *B. subtilis* ATCC 6633 and *C. albicans* ATCC 90028. Of these compounds, complex **1** displayed the greatest superoxide scavenging activity with IC_{50} of 49.07 μM, while complex **1** possessed the highest antimicrobial activity with IC_{50} of 256 μg/mL. The physicochemical parameters as derived from low energy conformer calculated at B3LYP/LANL2DZ level were well correlated with the observed superoxide scavenging properties. In particular, it has been shown that HOMO and LUMO energies are useful theoretical parameters for elucidating the superoxide scavenging activity as observed from the correlation coefficient of 0.97 and 0.99, respectively. Such coordination complexes described herein demonstrate great potential for the development of value-added metallovitamins for therapeutic applications.

Acknowledgements

We would like to acknowledge the Faculty of Medical Technology, Mahidol University and the Faculty of Science, Srinakharinwirot University for facilities and supports. T.S. is grateful for the graduate fellowship from The Ministry Staff Development Project by the Ministry of Education, Thailand. This project was also supported in part by the Young Scholars Research Fellowship from the Thailand Research Fund to C.N. (Grant No. MRG5080450) and the annual research grant of Mahidol University (B.E. 2551-2555).

References

[1] A. Banerjee, N. Dasgupta, B. De, Food Chem. 90 (2005) 727–733.
[2] M. Valko, D. Leibfritz, J. Moncol, M.T. Cronin, M. Mazur, J. Telser, Int. J. Biochem. Cell Biol. 39 (2007) 44–84.
[3] M. Valko, C.J. Rhodes, J. Moncol, M. Izakovic, M. Mazur, Chem. Biol. Interact. 160 (2006) 1–40.
[4] T. Sakurai, G. He, A. Matsuzawa, G.Y. Yu, S. Maeda, G. Hardiman, M. Karin, Cancer Cell 14 (2008) 156–165.
[5] M.K. Cathcart, Arterioscler. Thromb. Vasc. Biol. 24 (2004) 23–28.
[6] A.A. Andreadis, S.L. Hazen, S.A. Comhair, S.C. Erzurum, Free Radic. Biol. Med. 35 (2003) 213–225.
[7] J. Emerit, M. Edeas, F. Bricaire, Biomed. Pharmacother. 58 (2004) 39–46.
[8] A.S. Hearn, C. Tu, H.S. Nick, D.N. Silverman, J. Biol. Chem. 274 (1999) 24457–24460.
[9] Y. Wang, Z.Y. Yang, B.D. Wang, Transition Met. Chem. 30 (2005) 879–883.
[10] B. Gauter-Fleckenstein, K. Fleckenstein, K. Owzar, C. Jiang, I. Batinic-Haberle, Z. Vujaskovic, Free Radic. Biol. Med. 44 (2008) 982–989.
[11] T. Piacham, C. Isarankura-Na-Ayudhya, C. Nantasenamat, S. Yainoy, L. Ye, L. Bülow, V. Prachayasittikul, Biochem. Biophys. Res. Commun. 341 (2006) 925–930.
[12] T. Suksrichavalit, S. Prachayasittikul, T. Piacham, C. Isarankura-Na-Ayudhya, C. Nantasenamat, V. Prachayasittikul, Molecules 13 (2008) 3040–3056.
[13] E.T. Bodor, S. Offermanns, Br. J. Pharmacol. 153 (2007) S68–S75.
[14] B.G. Brown, X.Q. Zhao, Am. J. Cardiol. 101 (2008) B58–B62.
[15] R. Singh, R.S.S. Rao, J. Mol. Struct. 71 (1981) 23–30.
[16] M.A. El-Saadani, A.Y. Nassar, S.H.A. El-Ela, T.H. Metwally, A.M. Nafady, Biochem. Pharmacol. 46 (1993) 1011–1018.
[17] R.H. Salama, A.Y. Nassar, A.A. Nafady, H.H. Mohamed, Liver Int. 27 (2007) 454–464.
[18] P.J. Aggett, P.K. Fenwick, H. Kirk, J. Nutr. 119 (1989) 1432–1437.
[19] M.P. Heffernan, M.M. Nelson, M.J. Anadkat, Br. J. Dermatol. 156 (2007) 548–552.

[20] R. Shrivastava, R. Nagar, G.A. Ravishankar, R.K. Upreti, U.C. Chaturvedi, Indian J. Med. Res. 126 (2007) 440–446.
[21] S. Abe, W. Hu, H. Ishibashi, K. Hasumi, H. Yamaguchi, J. Infect. Chemother. 10 (2004) 181–184.
[22] H. Tomioka, T. Shimizu, Y. Tatano, Int. J. Antimicrob. Agents 29 (2007) 460–464.
[23] J.A. Fernandez-Pol, D.J. Klos, P.D. Hamilton, Anticancer Res. 21 (2001) 3773–3776.
[24] A. Mucci, L. Varesio, R. Neglia, B. Colombari, S. Pastorino, E. Blasi, Med. Microbiol. Immunol. 192 (2003) 71–78.
[25] J.M. Rawson, R.E.P. Winpenny, Coord. Chem. Rev. 139 (1995) 313–374.
[26] H. Cheng, F. Huq, P. Beale, K. Fisher, Eur. J. Med. Chem. 40 (2006) 772–781.
[27] F. Huq, H. Tayyem, P. Beale, J.Q. Yu, J. Inorg. Biochem. 101 (2007) 30–35.
[28] G. Zhao, H. Lin, P. Yu, H. Sun, S. Zhu, X. Su, Y. Chen, J. Inorg. Biochem. 73 (1999) 145–149.
[29] J.M. Hyvelin, M. Gautier, M.C. Lemaire, P. Bonnet, V. Eder, Pflügers. Arch. Eur. J. Physiol. 457. (2009) 721–730.
[30] J.L.R. Williams, G.K. Fyfe, C.P. Sibley, P.N. Baker, S.L. Greenwood, Am. J. Physiol. Regul. Integr. Comp. Physiol. 295 (2008) 1204–1213.
[31] J. Molgó, M. Lemeignan, F. Peradejordi, P. Lechat, J. Pharmacol. 16 (1985) 109–144.
[32] V.N. Pashkov, H.C. Hemmings Jr., Anesth. Analg. 95 (2002) 1274–1281.
[33] G. Mohan, R. Nagar, S.C. Agarwal, K.A. Mehta, C.S. Rao, J. Enzyme Inhib. Med. Chem. 20 (2005) 55–60.
[34] S. Bo, M. Durazzo, R. Gambino, C. Berutti, N. Milanesio, A. Caropreso, L. Gentile, M. Cassader, P. Cavallo-Perin, G. Pagano, J. Nutr. 138 (2008) 305–310.
[35] M. Olivares, R. Uauy, Am. J. Clin. Nutr. 63 (1996) 791S–796S.
[36] E. Koca, Y. Buyukasik, D. Cetiner, R. Yilmaz, N. Sayinalp, U. Yasavul, A. Uner, Leuk. Res. 32 (2008) 495–499.
[37] P.J. Twomey, T.M. Reynolds, A.S. Wierzbicki, A. Viljoen, Int. J. Clin. Pract. 62 (2008) 485–487.
[38] S. Prachayasittikul, P. Buraparuangsang, A. Worachartcheewan, C. Isarankura-Na-Ayudhya, S. Ruchirawat, V. Prachayasittikul, Molecules 13 (2008) 904–921.
[39] S. Prachayasittikul, T. Suksrichavalit, C. Isarankura-Na-Ayudhya, S. Ruchirawat, V. Prachayasittikul, EXCLI J. 7 (2008) 63–70.
[40] R. Dennington II, T. Keith, J. Millam, K. Eppinnett, W.L. Hovell, R. Gilliland, GaussView: Version 3.09, Semichem, Inc., Shawnee Mission, KS, 2003.
[41] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J. Montgomery, T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.
[42] L. Shen, H.Y. Zhang, H.F. Ji, J. Mol. Struct. Theochem. 817 (2007) 161–162.
[43] T.J. Franklin, G.A. Snow, Biochemistry of Antimicrobial Action, second ed. Chapman and Hall, London, 1971.
[44] K.N. Thimmaiah, W.D. Lloyd, G.T. Chandrappa, Inorg. Chim. Acta 106 (1985) 81–83.
[45] S.C.S. Jadon, N. Gupta, R.V. Singh, Indian J. Chem. 34 (1995) 733–736.
[46] S. Ahmed, D. Ismail, J. Surfactants Detergent 11 (2008) 231–235.
[47] N.A. Lagutkin, N.I. Mitin, M.M. Zubairov, V.A. Dorokhov, B.M. Mikhailov, Pharm. Chem. J. 16 (1982) 464–467.
[48] G. de Oliveira, J.M. Martin, F. de Proft, P. Geerlings, Phys. Rev. A 60 (1999) 1034–1045.
[49] H.F. Ji, H.Y. Zhang, Bioorg. Med. Chem. Lett. 15 (2005) 21–24.
[50] H.F. Ji, H.Y. Zhang, Chem. Res. Toxicol. 17 (2004) 471–475.
[51] R.J.F. Branco, P.A. Fernandes, M.J. Ramos, Theor. Chem. Acc. 115 (2006) 27–31.
[52] R.J.F. Branco, P.A. Fernandes, M.J. Ramos, J. Mol. Struct. Theochem. 729 (2005) 141–146.
[53] R. Sun, J. Ge, J. Yao, S. Li, H. Shen, R. Gua Spectrochim. Acta Part A 71 (2008) 1535–1539.
[54] J.T. Garbutt, A.L. Morehouse, A.M. Hanson, J. Agric. Food Chem. 9 (1961) 285–289.
[55] I. Schepetkin, A. Potapov, A. Khlebnikov, E. Korotkova, A. Lukina, G. Malovichko, L. Kirpotina, M.T. Quinn, J. Biol. Inorg. Chem. 11 (2006) 499–513.
[56] E.P. Cappellani, S.D. Drouin, G. Jia, P.A. Maltby, R.H. Morris, C.T. Schweitzer, J. Am. Chem. Soc. 116 (1994) 3375–3388.
[57] B. Bóka, A. Myari, I. Sóvágó, N. Hadjiliadis, J. Inorg. Biochem. 98 (2004) 113–122.
[58] Q.X. Li, Q.H. Luo, Y.Z. Li, M.C. Shen, Dalton Trans. 15 (2004) 2329–2335.

Molecules **2009**, *14*, 1869-1888; doi:10.3390/molecules14051869

OPEN ACCESS

molecules

ISSN 1420-3049

www.mdpi.com/journal/molecules

Article

Modeling the LPS Neutralization Activity of Anti-Endotoxins

Chadinee Thippakorn [1], Thummaruk Suksrichavalit [1], Chanin Nantasenamat [1], Tanawut Tantimongcolwat [1], Chartchalerm Isarankura-Na-Ayudhya [1], Thanakorn Naenna [2] and Virapong Prachayasittikul [1,*]

[1] Department of Clinical Microbiology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand

[2] Department of Industrial Engineering, Faculty of Engineering, Mahidol University, Nakhon Pathom 73170, Thailand

* Author to whom correspondence should be addressed; E-mail: mtvpr@mahidol.ac.th; Tel.: +662-441-4376; Fax: +662-441-4380

Received: 6 January 2009; in revised form: 15 May 2009 / Accepted: 19 May 2009 / Published: 20 May 2009

Abstract: Bacterial lipopolysaccharides (LPS), also known as endotoxins, are major structural components of the outer membrane of Gram-negative bacteria that serve as a barrier and protective shield between them and their surrounding environment. LPS is considered to be a major virulence factor as it strongly stimulates the secretion of pro-inflammatory cytokines which mediate the host immune response and culminating in septic shock. Quantitative structure-activity relationship studies of the LPS neutralization activities of anti-endotoxins were performed using charge and quantum chemical descriptors. Artificial neural network implementing the back-propagation algorithm was selected for the multivariate analysis. The predicted activities from leave-one-out cross-validation were well correlated with the experimental values as observed from the correlation coefficient and root mean square error of 0.930 and 0.162, respectively. Similarly, the external testing set also yielded good predictivity with correlation coefficient and root mean square error of 0.983 and 0.130. The model holds great potential for the rational design of novel and robust compounds with enhanced neutralization activity.

Keywords: lipopolysaccharide; endotoxin; anti-endotoxin; artificial neural network; QSAR

Introduction

Lipopolysaccharides (LPS) are major structural components of the outer membrane of Gram-negative bacteria. LPS confers a net negative charge to the membrane and thereby serves as a protective shield against antibacterial agents, as well as playing a crucial role in maintaining the integrity of the overall membrane structure. LPS are endotoxins which when present in the systemic circulation lead to the development of septic shock. The endotoxins initiate such events by first binding to CD14/TLR4/MD2 receptor complex on phagocytic cells causing the release of pro-inflammatory cytokines which consequently increase the releasing of other inflammatory mediators in many cell types (e.g. neutrophils, monocytes, vascular endothelial cell) as well as initiate the neuroendocrine response. Aside from this, LPS can also mediate the activation of plasma protein cascades (e.g. complement system) [1-4]. Among the mediators produced by activated macrophage, tumor necrosis factor (TNF) is the most important and the earliest to be released. TNF can amplify the response to endotoxins by further activating inflammatory cytokines (e.g. IL-1, IL-6, and IL-8) which consequently leads to the synthesis of other mediators such as arachidonic acid, platelet-activating factor (PAF), nitric oxide and reactive oxygen species [5, 6]. Such elevated levels of mediator decrease blood flow to vital organs (e.g. kidney, heart, and brain) by increasing vascular permeability as well as causing microvacular occlusion and damage thereby leading to multiple organ dysfunctions and eventually culminating in death [7].

Among those susceptible to the development of sepsis are children, immunocompromised individuals, and the elderly. In fact, global incidences of septic shock have increased over the past decade as a result of the growing number of immunologically compromised patients. Furthermore, LPS is considered to be one of the leading causes of mortality in intensive care units worldwide [8-10]. Therapeutic strategies for increasing the chances of survival include: (i) modulating inflammatory mediator release via the use of anti-cytokine or anti-inflammatory agents, (ii) supporting major organ dysfunction by increasing blood flow and (iii) administrating early antibiotic or anti-endotoxin usage. Therefore, drugs or compounds able to combat LPS toxicity have been of great interest [11]. Potential therapeutic agents such as anti-endotoxin antibodies, short peptides and lipopolyamines, as well as other small molecules, have been found to display promising sequestering effect towards lipid A but the efficiency of those compounds needs to be improved [12-18].

The toxic compartment of LPS is their structurally conserved glycolipid component called lipid A. Lipid A is a fatty acid chain bound to two phosphorylated glucosamine residues that is linked to the oligosaccharide core and the distal O-antigen polysaccharide chain. The O-polysaccharide chain is made of oligosaccharide repetitive units (O-units) while the oligosaccharide core is made of an inner core (ketodeoxyoctonic acid and heptose) and an outer core. The diversity in the composition and length of the O-antigen vary among different bacterial species.

In order to investigate the molecular parameters of interaction between lipid A and anti-endotoxin agents, Guo *et al*. utilized molecular modeling analysis for elucidating the underlying mechanism governing such high binding affinity. Their results indicated that a correlation exists between the binding affinity and the electrostatic interaction of ligands with lipid A [19]. In parallel, our previous investigations have revealed that quantitative structure-activity/property relationship (QSAR/QSPR) are useful tools for establishing relationships between molecular structures with the respective

activities or properties of a wide array of biological and chemical systems [20-26]. Elucidations on the feasibility of potential ligands prior to performing actual experimentations are useful on the economic and time-saving view points.

To the best of our knowledge, we present the first development of a quantitative structure-activity relationship model of LPS neutralization activity by anti-endotoxins as modeled by multivariate analysis. Molecular descriptors accounting for charge and electronic properties of anti-endotoxins were used as input variables for calculating the half-maximal effective displacement (ED_{50}).

Results and Discussion

Structural considerations

It has been reported that electrostatic and hydrophobic forces are necessary for LPS neutralization activity [27-30]. To eliminate the endotoxin, an agent would need to bind to lipid A. This is achieved by small molecules that are capable of engaging in strong hydrogen bond formation with the cationic and phosphate groups of lipid A. Furthermore, hydrophobic moieties are also necessary in stabilizing and enhancing the affinity of the agent in binding to the endotoxin [31]. Therefore to account for the electrostatic and hydrophobic forces, descriptors derived from RECON and Spartan'04 software packages were used for QSAR model development. Descriptors from RECON, which is based on the TAE methodology, are suitable for this study as it directly accounts for the molecular recognition in terms of the electronic charge properties. Additional descriptors from Spartan'04 were selected to account for the aforementioned interaction forces. Particularly, the hydrophobic interaction was approximated by the water accessible hydrophobic surface area of the molecule (CPK_{Area}). Molecular descriptors such as total energy (E_{Total}), atomic charge (Q_A) and total hydrogen atomic charge (Q_{TH+}) are important auxiliary variables that could also account for the electrostatic properties of the molecules. It was observed that the combination of both sets of descriptors exerted a dramatic enhancement of predictive power than solely relying on TAE descriptors (data not shown).

Variable reduction

Prior to performing the calculation of LPS neutralization activity, the redundant and multi-collinear descriptors present in the dataset were initially reduced by UFS in order to achieve a better efficiency of prediction. The subset of descriptors left after variable reduction was in the range of 3 and 20. The optimal number of descriptor to use was determined by making a plot of the number of selected descriptors as a function of RMS (Figure 1). It was observed that the optimal number was 10, which was selected for further investigations.

Parameter optimization

The network architecture was determined by a trial-and-error adjustment of various parameters for the purpose of obtaining an optimal configuration. The empirically determined parameters included the number of nodes in the hidden layer, the learning epoch size, and the learning rate and momentum. Parameter that exhibited the lowest RMS was chosen as optimum. The optimal number of hidden nodes was determined by varying the number of nodes from 1 to 25.

As represented in the plot of RMS versus hidden nodes (Figure 2), the optimal number of nodes was found to be seven. In order to ovoid overtraining of the network, the learning epoch size was subsequently optimized from 1 to 700 in increments of 50 and learning was stopped once a detectable rise in RMS for the leave-one-out cross-validated testing set was observed. The best learning time could be observed by making a plot of the RMS as a function of the learning epoch size (Figure 3). The optimal value was found to be 300.

Figure 1. Plot of RMS as a function of the number of TAE molecular descriptors.

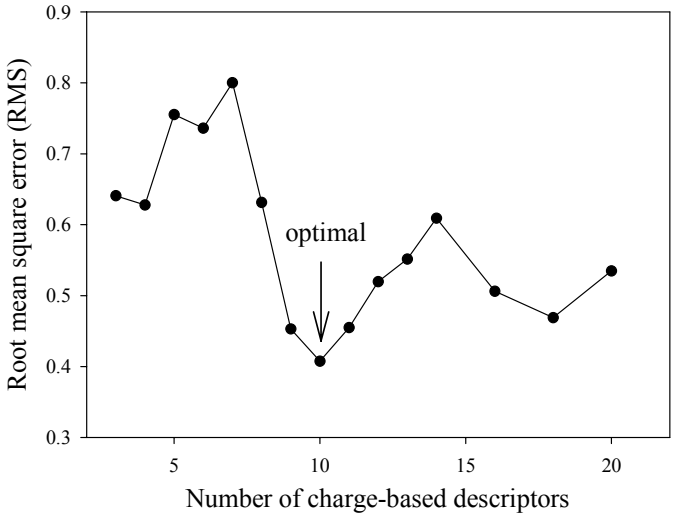
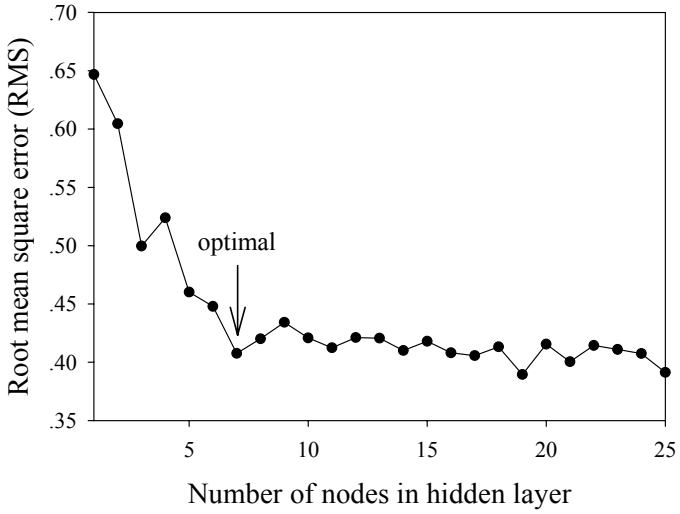


Figure 2. Plot of RMS as a function of the number of nodes in hidden layer.



Subsequently, the optimal learning rate and momentum was selected by making a contour plot of the RMS as function of the learning rate and momentum (Figure 4). The lines in the contour plot represented constant values of RMS, while shaded boxes designated RMS values that were obtained from the learning procedures and fitted onto the same surface model [32]. As shown on the contour plot, the best learning rate and momentum lies in the middle left region of the graph and was found to be 0.1 and 0.4, respectively.

Figure 3. Plot of RMS as a function of the number of learning epochs. The cross-validated test set is represented as a solid line while the training set is represented as a dotted line.

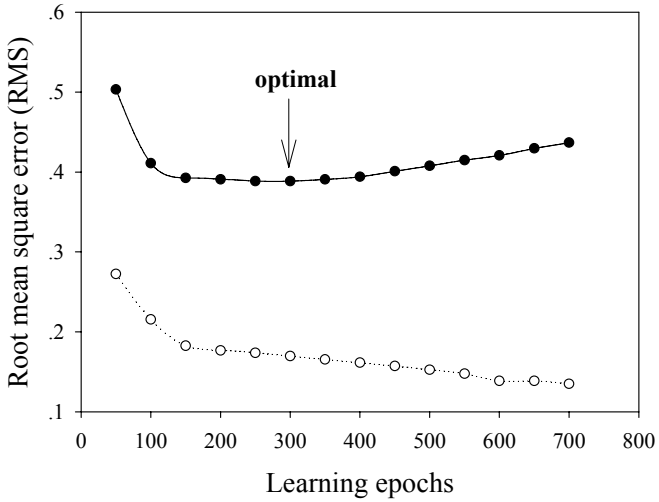
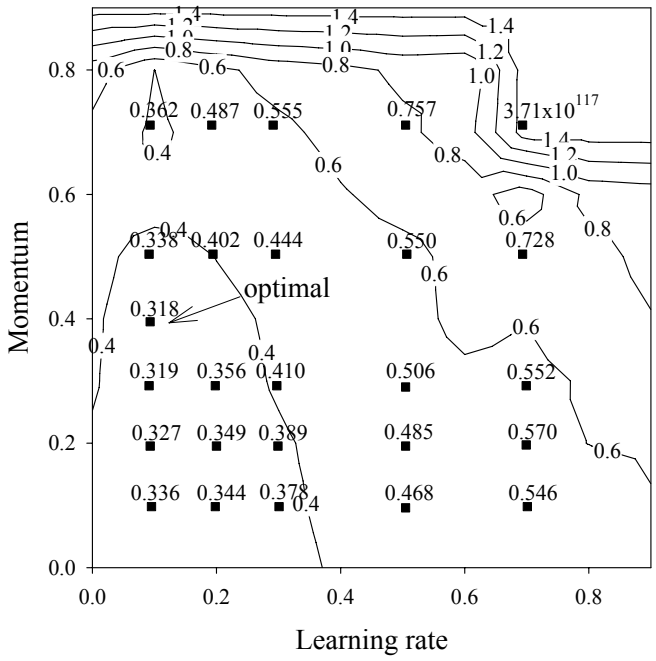


Figure 4. Contour plot of RMS as function of learning rate and momentum for the cross-validated testing set.



Prediction of LPS neutralization activity using artificial neural network

The optimal configuration of the predictive model of LPS neutralization activity was identified to be 7, 300, 0.1 and 0.4 for the number of hidden nodes, learning epoch size, learning rate and momentum, respectively. The network was created by means of a leave-one-out cross validation [33-35], by which one sample of the dataset was withdrawn for use as the test set while the rest served as the training set. This process was repeated iteratively until all samples of the dataset were used as the test set [32, 36-37]. It was observed that the predicted and experimental neutralization activities were moderately correlated as can be observed from the correlation coefficient of 0.781. To further refine

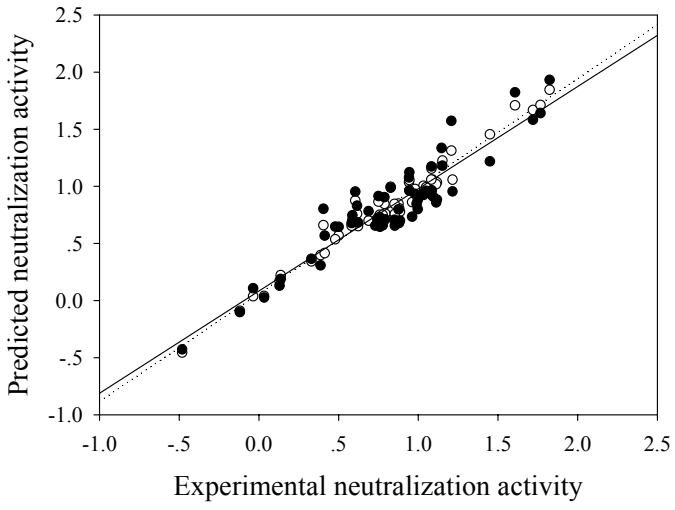
the model, identification of potential outliers present inherently in the predictive model was performed using standard statistical analysis where compounds with absolute standardized residuals exceeding the cut-off value of 2 were marked as outliers. This statistical analysis was reiterated to yield a final predictive model which was constructed using previously determined optimal set of network parameters. As indicated in Table 1, results yielded correlation coefficient and root mean square error of 0.980 and 0.124 for the training set, respectively, while 0.930 and 0.162 was observed for the testing set. A plot showing the experimental versus predicted LPS neutralization activity for model 6 is shown in Figure 5.

Table 1. Summary of the predictive performance.

Model	*N*	r_{Tr}	RMS_{Tr}	r_{CV}	RMS_{CV}	R^2	R^2_{adj}	*F* ratio	Critical *F* value
1	73	0.938	0.167	0.781	0.287	0.610	0.524	6.480	1.866[a]
2	69	0.942	0.158	0.728	0.341	0.530	0.419	4.350	1.879[b]
3	64	0.947	0.141	0.857	0.221	0.734	0.665	9.658	1.899[c]
4	63	0.952	0.146	0.823	0.250	0.677	0.591	7.186	1.904[d]
5	60	0.972	0.124	0.907	0.186	0.823	0.773	14.946	1.918[e]
6	58	0.980	0.124	0.930	0.162	0.865	0.825	19.680	1.929[f]

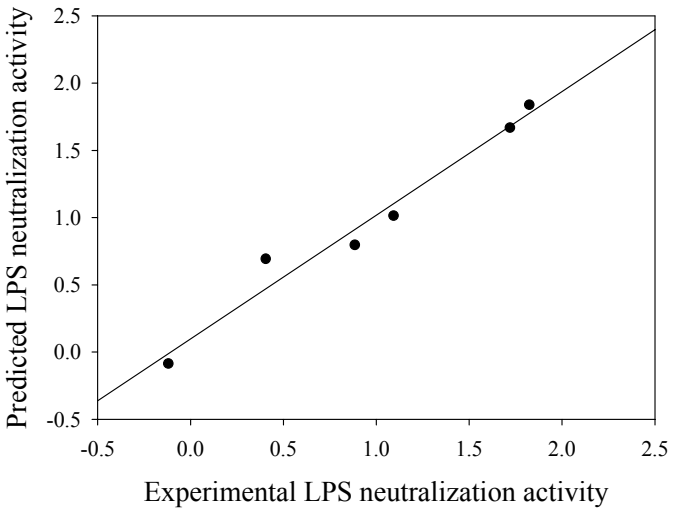
N, sample size of the data set; r_{Tr}, correlation coefficient of training set; RMS_{Tr}, root mean square error of training set; r_{CV}, correlation coefficient of leave-one-out cross validation testing set; RMS_{CV}, root mean square error of testing set; R^2, squared correlation coefficient of leave-one-out cross validation testing set, R^2_{adj}, adjusted squared correlation coefficient of leave-one-out cross validation testing set; *F* ratio, calculated *F* ratio of cross validation testing set. Critical *F* values at the 95% confidence level with *m* and $n - m - 1$ degrees of freedom ($F_{(m,n-m-1)}$) as follows [a] $F_{(14,58)}$, [b] $F_{(14,54)}$, [c] $F_{(14,49)}$, [d] $F_{(14,48)}$, [e] $F_{(14,45)}$, and [f] $F_{(14,43)}$ [38].

Figure 5. Plot of predicted versus experimental ED_{50} for training set (○; regression line is represented as a dotted line) and leave-one-out cross-validation testing set (●; regression line is represented as a solid line) of anti-endotoxins.



In order to evaluate the predictive power of the QSAR model discussed herein, an external testing set would be necessary. To achieve this, the data set from model 6 was further divided into two sets of data: (i) one portion for deriving the optimal network parameters by leave-one-out cross-validation and (ii) an external testing set for evaluating the extrapolation capability of the QSAR model. This was carried out by randomly selecting 10% of the data set as an external testing set which equates to 6 data samples while the remaining 52 data samples were used for performing leave-one-out cross-validation. Optimization of the network parameter was performed as previously discussed to give the optimal parameters as follows: 15 nodes in the hidden layer, 150 learning epochs, learning rate of 0.2 and momentum of 0.4. This set of parameter was used for deriving the predictive performance of the external test set. Results indicated that the proposed QSAR model could accurately predict the LPS neutralization activity as observed by the correlation coefficient of 0.983 and root mean square error of 0.130 as shown in Figure 6.

Figure 6. Plot of predicted versus experimental ED_{50} for the leave-one-out cross-validation testing set of anti-endotoxins.



Conclusions

The LPS neutralization activities of anti-endotoxin agents were modeled using back-propagation neural network. The predicted neutralization activities were found to be in good agreement with the experimental values. Analysis of these results suggests that the use of charge-based descriptors and quantum chemical descriptors were useful and necessary in obtaining good prediction that is representative of the experimental values. Therefore, such methodology proposed in this study demonstrates a facile approach for the design of novel anti-endotoxin agents with robust properties.

Experimental

Data collection

The half-maximal effective displacement (ED_{50}) of 80 anti-endotoxins were collected from the literature [27, 31, 40] (Table 1). Seven compounds were identified as inactive and were removed due to their high level of ED_{50} >200.

Descriptor generation

The two-dimensional structure of each compound was drawn in ChemAxon's MarvinSketch [41] and exported as SMILES notation. The three-dimensional molecular structures were constructed using the molecular building module of Spartan'04 [42] and subsequently submitted for calculation of quantum chemical descriptors. Likewise, the SMILES notation served as input for the calculation of charge-based descriptors by RECON.

RECON (version 5.5) was used for the generation of 248 transferable atom equivalent (TAE) molecular descriptors. The TAE methodology, based on Bader's quantum theory of atoms in molecule, was developed by Brenemen and co-workers for the rapid reconstruction of molecular charge density and molecular electronic property via pre-computed *ab initio* atomic charge density fragment [42]. TAE descriptors were selected for the prediction of LPS neutralization activity as they could account for the electronic properties of molecules, which is crucial for modeling molecular interaction [25] of the anti-endotoxins with their respective LPS target.

Aside from electronic properties, the involvement of hydrophobic and electrophilic forces has also been demonstrated to play a crucial role in the biological activity of the anti-endotoxins [11, 12]. Hence, Spartan'04 was employed for generation of the following quantum chemical descriptors: molecular weight (MW), water accessible hydrophobic surface area of the molecule (CPK_{Area}), total energy (E_{Total}), and atomic charge (Q_A). The three-dimensional structures of each anti-endotoxin compound were initially optimized using the Merck Molecular Force Field (MMFF) in conjunction with Monte Carlo simulation or systematic conformational search to identify the lowest energy geometry. These pre-optimized structures were then calculated at the semi-empirical level using the Parameterization Method 3 (PM3) [42].

An additional descriptor, the total hydrogen atomic charge (Q_{TH+}) was calculated according to the following equation:

$$Q_{TH^+} = \sum_{i=1}^{n} (Z_i - q_i) \quad (1)$$

where Z_i, q_i and n represent the atomic numbers, the atomic electron populations and the number of hydrogen atoms, respectively.

Descriptor reduction

The molecular descriptors were subjected to variable reduction in order to reduce computational time, minimize multi-collinearity and eliminate redundancy of the descriptors. This was performed using UFS, version 1.8, which is a computer program based on the Unsupervised Forward Selection (UFS) algorithm [43]. Briefly, UFS removes variables when the standard deviation is less than the predefined sdevmin and terminates when the squared multiple correlation coefficients of the remaining variables exhibit values greater than the r-squared-max ($R^2_{\max}$). The standard deviation was left as default at 0.0005 while the $R^2_{\max}$ was varied between 0 and 0.99. Thus, 248 descriptors were reduced to a range of three to 20 descriptors.

Table 2. Data set of the anti-endotoxins[a].

No.	Compound Name	Exp. ED_{50}	Exp. $logED_{50}$	Pred. $logED_{50}$	Residual
1[c]	N^1-Acetyl-1,16-diamino-4,8,13triazahexadecane tetrakis(trifluoroacetic acid)	107.450	2.031	1.395	0.636
2	N^1-Nonanoyl-1,16-diamino-4,8,13-triazahexadecane tetrakis(trifluoroacetic acid)	0.920	-0.036	0.254	-0.290
3	N^1-Decanoyl-1,16-diamino-4,8,13-triazahexadecane tetrakis(trifluoroacetic acid)	1.080	0.033	-0.210	0.243
4	N^1-Pentadecanoyl-1,16-diamino-4,8,13-triazahexadecane tetrakis(trifluoroacetic acid)	1.350	0.130	0.118	0.012
5	N^1-Heptadecanoyl-1,16-diamino-4,8,13-triazahexadecane tetrakis(trifluoroacetic acid)	1.370	0.137	0.335	-0.198
6	N^1-Nonadecanoyl-1,16-diamino-4,8,13-triazahexadecane tetrakis(trifluoroacetic acid)	2.440	0.387	0.306	0.081
7	N^1,N^{20}-Dinonanoyl-1,20-diamino-4,8,13,17-tetrazaicosane tetrakis(trifluoroacetic acid)	0.330	-0.481	-0.203	-0.278
8	N^1,N^{20}-Didecanoyl-1,20-diamino-4,8,13,17-tetrazaicosane tetrakis(trifluoroacetic acid)	0.760	-0.119	0.058	-0.177
9[f]	N^1,N^{20}-Didodecanoyl-1,20-diamino-4,8,13,17-tetrazaicosane tetrakis(trifluoroacetic acid)	6.870	0.837	0.132	0.705
10[f]	N^1,N^{20}-Dipentdecanoyl-1,20-diamino-4,8,13,17-tetrazaicosane tetrakis(trifluoroacetic acid)	8.670	0.938	1.518	-0.580
11	N^1,N^{20}-Diheptadecanoyl-1,20-diamino-4,8,13,17-tetrazaicosane tetrakis (trifluoroacetic acid)	52.530	1.72	1.806	-0.086
12	N^1,N^{20}-Dinonadecanoyl-1,20-diamino-4,8,13,17-tetrazaicosane tetrakis(trifluoroacetic acid)	66.730	1.824	1.658	0.166
13	(S)-1-(1-(2-(2-aminoethoxy)ethylamino)-1-oxo-3-phenylpropan-2-yl)-3-octadecylurea	12.400	1.093	0.871	0.222
14	(S)-N-(2-(2-aminoethoxy)ethyl)-2-(2 (octadecylamino)acetamido)-3-phenylpropanamide	2.540	0.405	0.706	-0.301
15	(S)-N-(2-(2-aminoethoxy)ethyl)-2-(3-(octadecylamino)propanamido)-3-phenylpropanamide	7.680	0.885	0.591	0.294
16	(S)-1-(1-(2-(2-aminoethoxy)ethylamino)-3-(1H-imidazol-4-yl)-1-oxopropan-2-yl)-3-octadecylurea	13.100	1.117	0.881	0.236
17	(S)-N-(2-(2-aminoethoxy)ethyl)-3-(1H-imidazol-4-yl)-2-(2-(octadecylamino)acetamido)propanamide	3.170	0.501	0.557	-0.056
18	(S)-N-(2-(2-aminoethoxy)ethyl)-3-(1H-imidazol-4-yl)-2-(3-(octadecylamino)propanamido)propanamide	5.380	0.731	0.869	-0.138
19	1-(2-(2-(2-aminoethoxy)ethylamino)-2-oxoethyl)-3-octadecylurea	14.000	1.146	0.980	0.166
20	N-(2-(2-aminoethoxy)ethyl)-2-(2 (octadecylamino)acetamido)acetamide	14.200	1.152	1.187	-0.035
21	N-(2-(2-(2-aminoethoxy)ethylamino)-2-oxoethyl)-3-(nonadecylamino)propanamide	10.800	1.033	1.025	0.009

Table 2. Cont.

No.	Compound Name	Exp. ED50	Exp. logED50	Pred. logED50	Residual
22	(S)-1-(1-(3-aminopropylamino)-1-oxo-3-phenylpropan-2-yl)-3-octadecylurea	8.800	0.944	0.990	-0.046
23	(S)-N-(3-aminopropyl)-2-(2-(octadecylamino)acetamido)-3-phenylpropanamide	4.130	0.616	0.707	-0.091
24	(S)-N-(3-aminopropyl)-2-(3-(octadecylamino)propanamido)-3-phenylpropanamide	5.750	0.760	0.666	0.094
25	(S)-1-(1-(3-aminopropylamino)-3-(1H-imidazol-4-yl)-1-oxopropan-2-yl)-3-octadecylurea	4.870	0.688	0.814	-0.126
26[f]	(S)-N-(3-aminopropyl)-3-(1H-imidazol-4-yl)-2-(2-(octadecylamino)acetamido)propanamide	6.860	0.836	0.414	0.422
27	(S)-N-(3-aminopropyl)-3-(1H-imidazol-4-yl)-2-(3-(octadecylamino)propanamido)propanamide	3.010	0.479	0.994	-0.515
28[c]	1-(2-(3-aminopropylamino)-2-oxoethyl)-3-octadecylurea	6.610	0.82	1.493	-0.673
29[b]	N-(3-aminopropyl)-2-(2-(octadecylamino)acetamido)acetamide	2420	3.384	-	-
30	N-(2-(3-aminopropylamino)-2-oxoethyl)-3-(octadecylamino)propanamide	6.140	0.788	0.961	-0.173
31[b]	(S)-N-(5-aminopentyl)-2-(2-(octadecylamino)acetamido)-3-phenylpropanamide	3850	3.585	-	-
32	(S)-N-(5-aminopentyl)-2-(3-(octadecylamino)propanamido)-3-phenylpropanamide	7.510	0.876	0.861	0.015
33	(S)-1-(1-(5-aminopentylamino)-3-(1H-imidazol-4-yl)-1-oxopropan-2-yl)-3-octadecylurea	12.100	1.083	1.343	-0.260
34[g]	(S)-N-(5-aminopentyl)-3-(1H-imidazol-4-yl)-2-(3-(octadecylamino)propanamido)propanamide	18.700	1.272	0.702	0.570
35	1-(2-(5-aminopentylamino)-2-oxoethyl)-3-octadecylurea	28.200	1.450	0.944	0.506
36	N-(5-aminopentyl)-2-(2-(octadecylamino)acetamido)acetamide	11.200	1.049	0.967	0.082
37	N-(2-(5-aminopentylamino)-2-oxoethyl)-3-(octadecylamino)propanamide	9.770	0.990	0.900	0.090
38[g]	1-(3-aminopropyl)-3-octadecylurea	3.800	0.580	0.839	-0.259
39	N-(3-aminopropyl)-2-(octadecylamino)acetamide	9.920	0.997	0.652	0.346
40	N-(3-aminopropyl)-3-(octadecylamino)propanamide	6.210	0.793	0.794	-0.001
41	1-(5-aminopentyl)-3-octadecylurea	8.740	0.942	0.841	0.101
42	1-(2-(2-aminoethoxy)ethyl)-3-octadecylurea	12.150	1.085	1.032	0.053
43	N-(5-aminopentyl)-2-(octadecylamino)acetamide	4.030	0.605	0.908	-0.303
44	N-(2-(2-aminoethoxy)ethyl)-2-(octadecylamino)acetamide	9.160	0.962	0.797	0.165
45	N-(5-aminopentyl)-3-(octadecylamino)propanamide	7.610	0.881	0.895	-0.014
46	N-(2-(2-aminoethoxy)ethyl)-3-(octadecylamino)propanamide	5.730	0.758	0.883	-0.125
47	L-Lys-N1-spermine	40.420	1.607	1.725	-0.118
48	D-Lys-N1-spermine	58.420	1.767	1.611	0.157

Table 2. Cont.

No.	Compound Name	Exp. ED_{50}	Exp. $logED_{50}$	Pred. $logED_{50}$	Residual
49[e]	L-Lys-ε-(eicosanoyl)-N1-spermine	6.460	0.810	1.052	-0.242
50	D-Lys-ε-(stearoyl)-N1-spermine	8.800	0.944	0.988	-0.044
51	L-Lys-ε-(stearoyl)-N1-spermine	16.390	1.215	0.888	0.327
52	L-Lys(ene-Δ11-stearoyl)-N1-spermine	4.200	0.623	0.812	-0.189
53	L-Lys-ε-(heptadecanoyl)-N1-spermine	6.710	0.827	0.990	-0.163
54[c]	L-Lys-ε-(hexadecanesulfonamide)-N1-spermine	5.930	0.773	1.277	-0.504
55	D-Lys-ε-(palmitoyl)-N1-spermine	9.940	0.997	0.926	0.071
56	L-Lys(palmitoyl)-N1-spermine	10.740	1.031	0.962	0.069
57	L-Lys(ene-Δ9-palmitoyl)-N1-spermine	3.820	0.582	0.494	0.088
58	L-Lys-ε-(myristoyl)-N1-spermine	5.630	0.751	0.986	-0.235
59	L-Lys-ε-(octanoyl)-N1-spermine	12.970	1.113	1.112	0.002
60[b]	D-Lys-ε-(isopropyl)-N1-spermine	298.850	2.475	-	-
61[b]	D-Lys-ε-(dimethylpropyl)-N1-spermine	327.040	2.515	-	-
62	D-Lys-ε-(2-norbornaneacetoyl)-N1-spermine	16.160	1.208	0.915	0.293
63[c]	D-Lys-ε-(4-biphenycarboxamide)-N1-spermine	7.860	0.895	1.554	-0.659
64	L-Lys-ε-(4-(1-pyrene)-butanoyl)-N1-spermine	7.090	0.851	0.296	0.555
65[b]	L-Lys-ε-(methylpolyethyleneglycolpropionyl)-N1-spermine	310.950	2.493	-	-
66[b]	L-Lys-ε-(2-[2-(2-methoxyethoxy)ethoxy]acetoyl)-N1-spermine	572.500	2.758	-	-
67[b]	L-Lys-ε-(2-(2-methoxyethoxy)acetoyl)-N1-spermine	495.190	2.695	-	-
68	L-Lys-ε-(hexadecyl)-N1-spermine	5.560	0.745	0.619	0.126
69	L-Lys-ε-(ene -Δ11-hexadecyl)-N1-spermine	2.590	0.413	-0.049	0.462
70	D-Lys-ε-(n-heptyl)-N1-spermine	3.860	0.587	0.751	-0.164
71	L-Lys-ε-(n-heptyl)-N1-spermine	5.990	0.777	0.696	0.081
72	L-Lys-ε-(bis-(n-heptyl))-N1-spermine	2.140	0.330	0.355	-0.025
73	L-Lys-ε-(n-hexyl)-N1-spermine	7.130	0.853	0.877	-0.024
74	D-(S)-Lys-ε-(ene-Δ6(3,7-dimethyl-1-octyl))-N1-spermine	9.550	0.980	0.951	0.029
75	L-Lys-ε-(3,3-dimethyl-1-butyl)-N1-spermine	12.070	1.082	1.174	-0.092
76[d]	D-Lys-ε-(3,3-dimethyl-1-butyl)-N1-spermine	10.930	1.039	1.244	-0.205
77[d]	D-Lys-ε-(3-methylpropyl)-N1-spermine	100.580	2.003	1.385	0.618
78[d]	L-(R)-Lys-ε-((2-isoproply-5-methyl)cyclohexyl)-N1-spermine	16.080	1.206	1.280	-0.074
79[d]	L-Lys-ε-(bis-(cyclohexyl))-N1-spermine	4.040	0.606	0.099	0.507
80[d]	D-Lys-ε-(4-phenylbenzyl)-N1-spermine	3.710	0.569	0.056	0.513

[a] Compounds no. 1-12, 13-46 and 47-80 were derived from [39], [31] and [27], respectively; [b] Compounds identified as inactive and removed from data set; Compounds identified as outliers in Models [c] 1, [d] 2, [e] 3, [f] 4, and [g] 5 according to standardized residual cut-off of 2.

Figure 7. Chemical structures of the anti-endotoxin dataset.

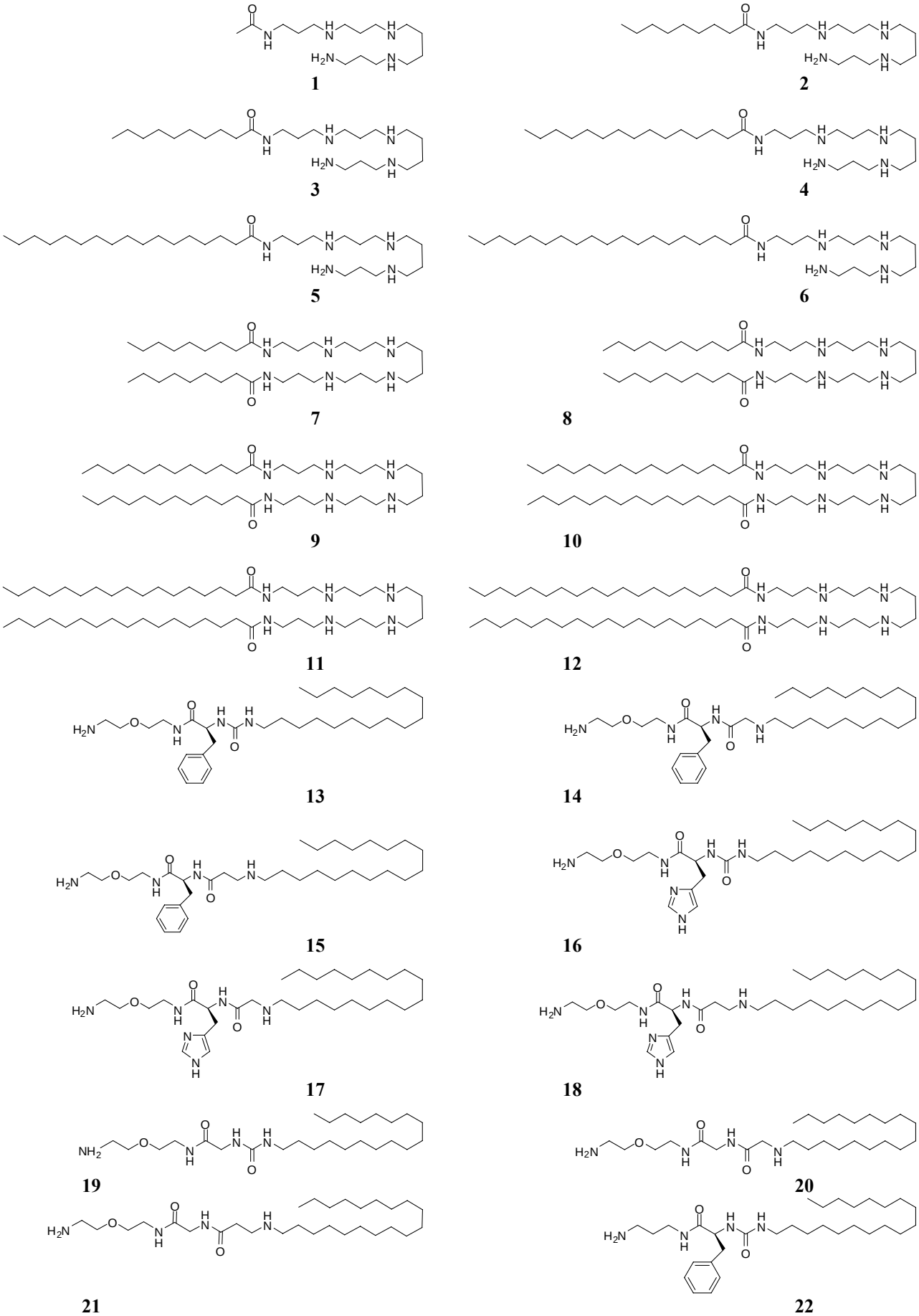
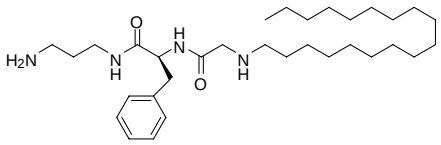
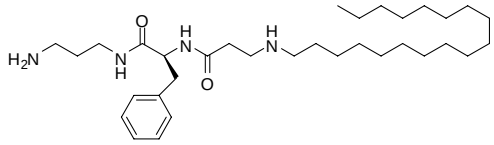


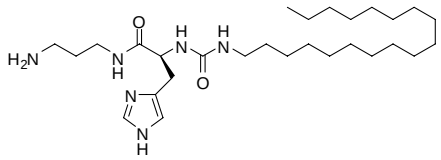
Figure 7. Cont.



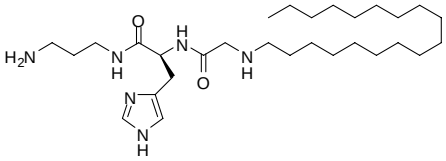
23



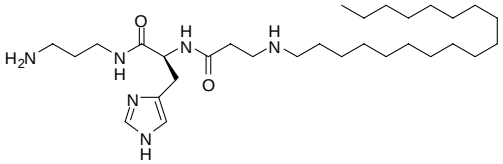
24



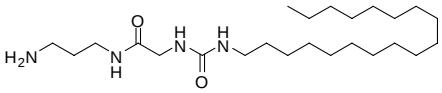
25



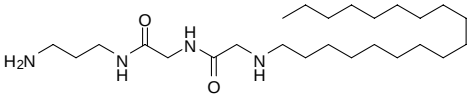
26



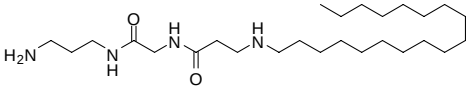
27



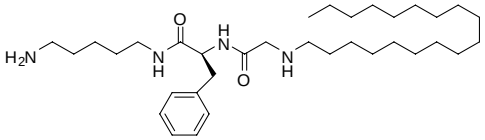
28



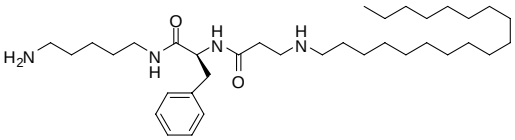
29



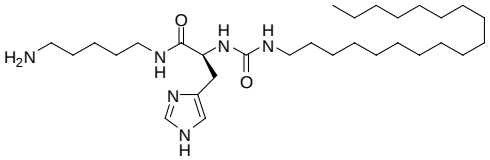
30



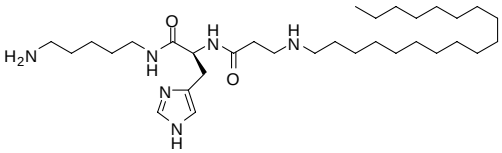
31



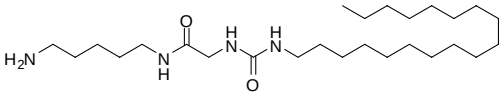
32



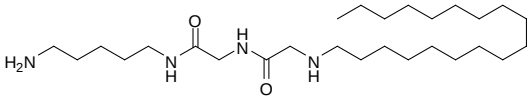
33



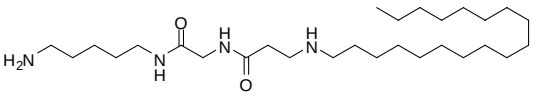
34



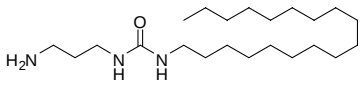
35



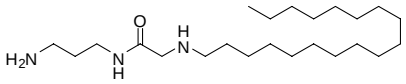
36



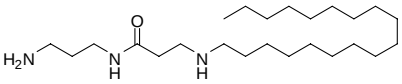
37



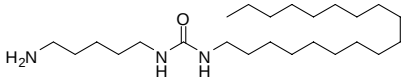
38



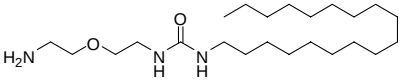
39



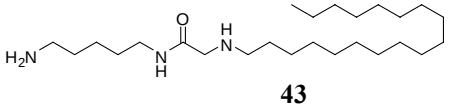
40



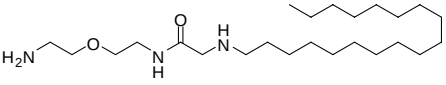
41



42

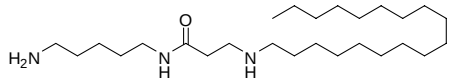


43

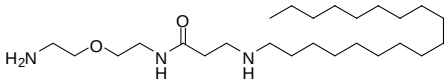


44

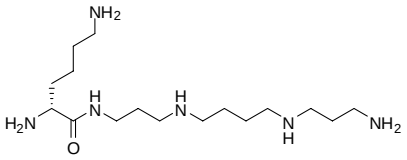
Figure 7. Cont.



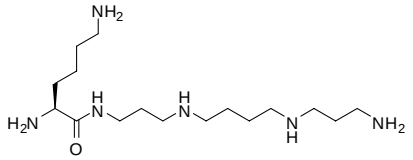
45



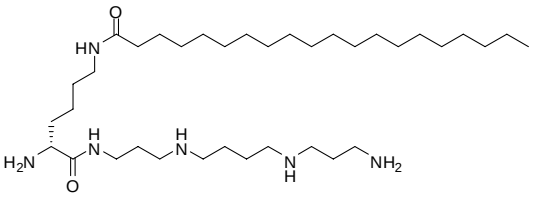
46



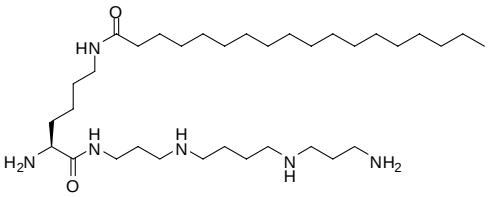
47



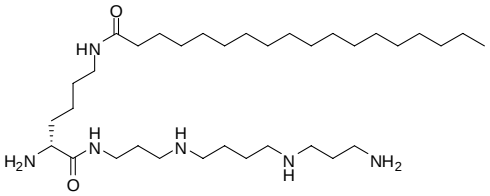
48



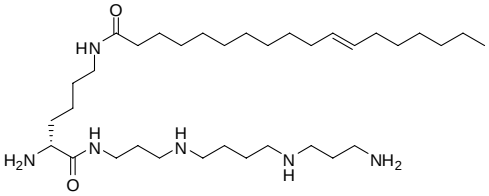
49



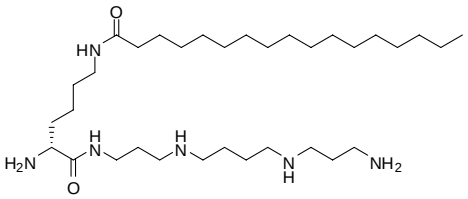
50



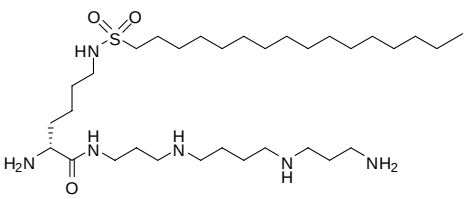
51



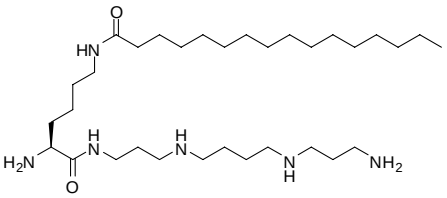
52



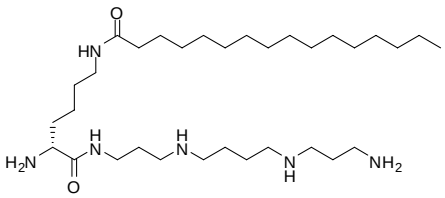
53



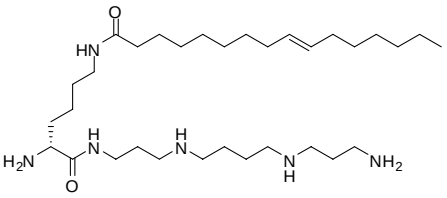
54



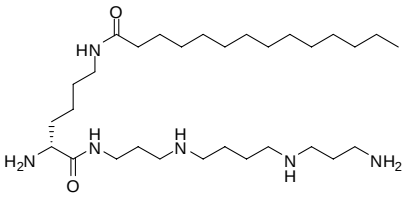
55



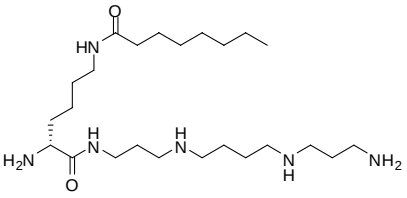
56



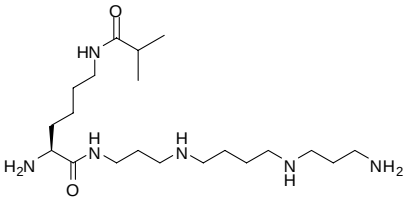
57



58

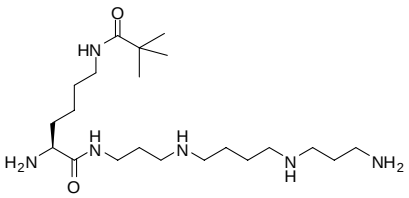


59

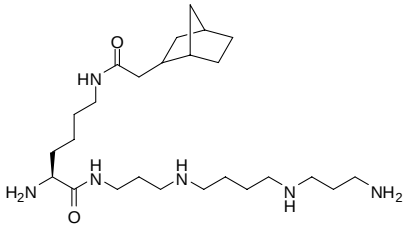


60

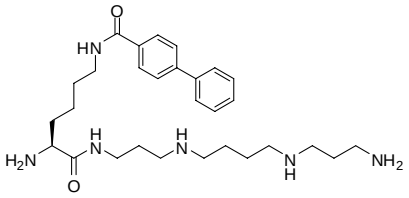
Figure 7. Cont.



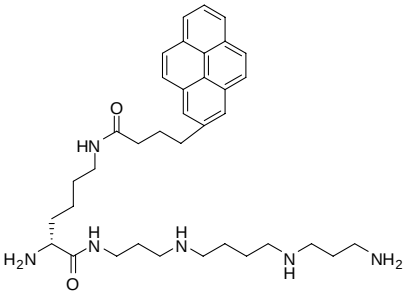
61



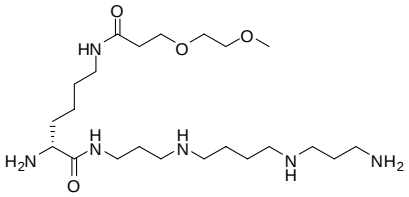
62



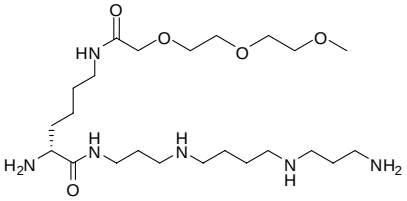
63



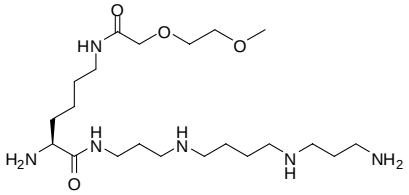
64



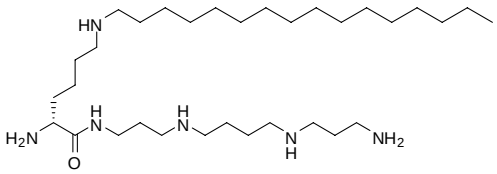
65



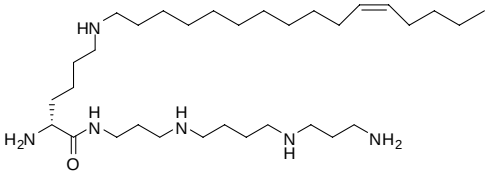
66



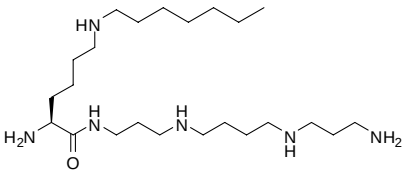
67



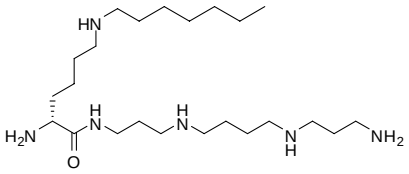
68



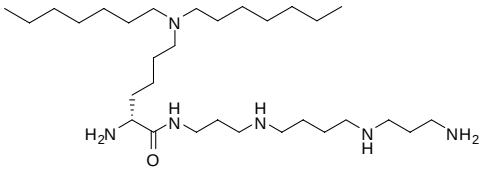
69



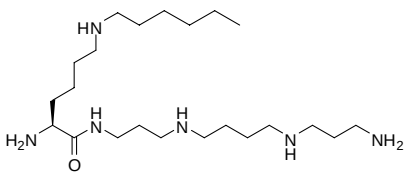
70



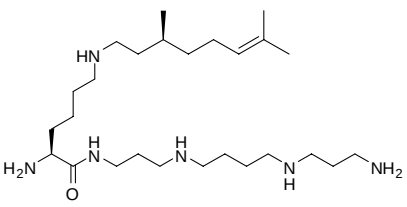
71



72

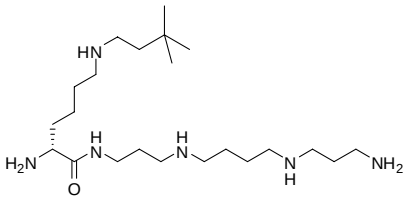


73

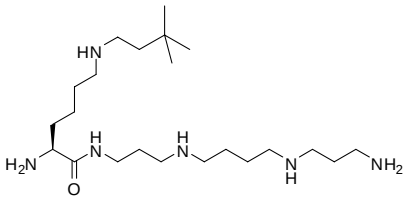


74

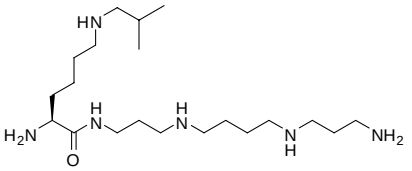
Figure 7. Cont.



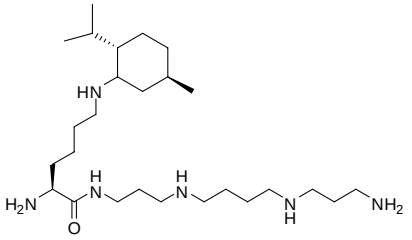
75



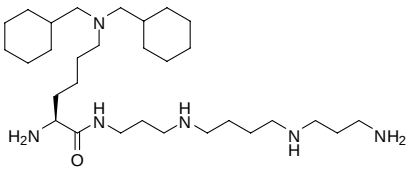
76



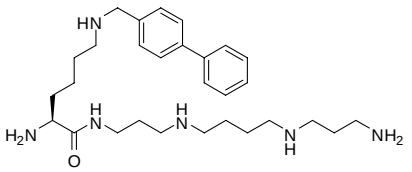
77



78



79



80

Overview of artificial neural network

Artificial neural network (ANN) is a computational model that mimics the learning process of the human brain. This is performed in a supervised manner where the model attempts to find a relationship between the independent and dependent variables. ANN is comprised of multiple layers of artificial neurons that are interconnected in a feed-forward manner where signals are relayed from the input layer through the hidden layer and finally to the output layer. The connections among the network of neurons are assigned numerical values known as weights which alter the signal flow through the neural network and governing its predictive performance. Particularly, the weights are adjusted adaptively to reduce the error by back-propagating signals from the output layer back to the input layer through the hidden layer. Readers are referred to the excellent book by Zupan and Gasteiger [32] for further methodological information.

Prediction of LPS neutralization activity

The predictive model of LPS neutralization activity was developed using back-propagation neural network as calculated by Weka, version 3.4.3 [44]. The independent variables, comprising of TAE descriptors and quantum chemical descriptors, of each compound were normalized to a range of 0 and 1 according to the following equation:

$$x_{norm} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

where x_{norm}, x_i, $x_{\min}$, $x_{\max}$ represent the normalized data, the value of each instance, the minimum value, and the maximum value of the dataset, respectively.

The optimal neural network parameters were obtained by an empirical trial-and-error search of the number of nodes in the hidden layer, the learning epoch size, the learning rate (η) and momentum (μ)

constant. A learning epoch refers to a complete cycle of data propagated through the layers of the network in a feed-forward manner. η controls the speed of weight adjustment, while μ prevents sudden changes in attaining the solution. Under an incremental tuning of such parameter, root mean square error (RMS) was simultaneously measured as an indicator of predictive error, which is calculated according to the following equation:

$$RMS = \frac{\sum_{i=1}^{n}(p_i - a_i)}{n} \tag{3}$$

where p_i, a_i, and n represent the predicted output, the actual output and the number of compounds in the dataset, respectively. As each neural network calculation operates via random initialization of the weights, the averaged RMS of ten runs was used as a measure of predictive error by adjusting the random seed from 0 to 9.

Internal validation procedure

Generation of training and testing sets was made using leave-one-out cross-validation (LOO-CV). Briefly, LOO-CV involved the leaving out of one molecule as the testing set and using the remaining as training set. This is performed iteratively until all samples were given the chance to be left out as testing sets.

Statistical analysis

The adjusted R^2 takes into consideration such information as the number of independent variables that are present in the predictive model. This is calculated according to the following equation:

$$R^2_{adj} = 1 - (1 - R^2) \cdot \left(\frac{n-1}{n-p} \right) \tag{4}$$

where *n* is the sample size and *p* is the number of independent variables.

The *F* ratio measures the explained variance (R^2) in relation to the unexplained variance ($1 - R^2$) with *m* and $n - m - 1$ degrees of freedom:

$$F_{(m,n-m-1)} = \frac{R^2 / m}{(1 - R^2)/(n - m - 1)} \tag{5}$$

where *m* is the number of independent variables and *n* is the number of compounds presented in the data set.

Outlying molecules were identified by standardization of the residuals according to the following equation:

$$x_{ij}^{\sin} = \frac{x_{ij} - \bar{x}_j}{\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2 / N} \tag{6}$$

where $x_{ij}^{\sin}$ represents the standardized residual, x_{ij} represents the residual of each sample, $\bar{x}_j$ represents the mean of the residual, and N represents the sample size of the data set. The cut-off value for the absolute standardized residual was set to 2.

Acknowledgements

C.T. and T.S. are Ph.D. students financially supported by the Royal Golden Jubilee Ph.D. scholarship from The Thailand Research Fund and the Ph.D. scholarship of the Ministry Staff Development Project from the Commission on Higher Education, Ministry of Education, respectively, under the supervision of V.P. This research is partially supported in part by the Young Scholars research fellowship from The Thailand Research Fund to C.N. (Grant No. MRG5080450) and T.T. (Grant No. MRG5080158). Partial support from the annual governmental budget of Mahidol University (B.E. 2551-2555) is also gratefully acknowledged.

References

1. Hurley, J.C. Antibiotic-induced release of endotoxin: a reappraisal. *Clin. Infect. Dis.* **1992**, *15*, 840-854.
2. Masoud, H.; Altman, E.; Richards, J.C.; Lam, J.S. General strategy for structural analysis of the oligosaccharide region of lipooligosaccharides. Structure of the oligosaccharide component of *Pseudomonas aeruginosa* IATS serotype 06 mutant R5 rough-type lipopolysaccharide. *Biochemistry* **1994**, *33*, 10568-10578.
3. Munford, R.S. Severe sepsis and septic shock: the role of gram-negative bacteremia. *Annu. Rev. Pathol.* **2006**, *1*, 467-496.
4. Ulevitch, R.J.; Tobias, P.S. Recognition of gram-negative bacteria and endotoxin by the innate immune system. *Curr. Opin. Immunol.* **1999**, *11*, 19-22.
5. Zhang, F.X.; Kirschning, C.J.; Mancinelli, R.; Xu, X.P.; Jin, Y.; Faure, E.; Mantovani, A.; Rothe, M.; Muzio, M.; Arditi, M. Bacterial lipopolysaccharide activates nuclear factor-kappaB through interleukin-1 signaling mediators in cultured human dermal endothelial cells and mononuclear phagocytes. *J. Biol. Chem.* **1999**, *274*, 7611-7614.
6. Tsiotou, A.G.; Sakorafas, G.H.; Anagnostopoulos, G.; Bramis, J. Septic shock; current pathogenic concepts from a clinical perspective. *Med. Sci. Monit.* **2005**, *11*, RA76-RA85.
7. Muhle, S.A.; Tam, J.P. Design of Gram-negative selective antimicrobial peptides. *Biochemistry* **2001**, *40*, 5777-5785.
8. Khownium, K.; Wood, S.J.; Miller, K.A.; Balakrishna, R.; Nguyen, T.B.; Kimbrell, M.R.; Georg, G.I.; David, S.A. Novel endotoxin-sequestering compounds with terephthalaldehyde-bis-guanylhydrazone scaffolds. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1305-1308.
9. Martin, G.S.; Mannino, D.M.; Eaton, S.; Moss, M. The epidemiology of sepsis in the United States from 1979 through 2000. *New Engl. J. Med.* **2003**, *348*, 1546-1554.
10. Moss, M.; Martin, G.S. A global perspective on the epidemiology of sepsis. *Intens. Care Med.* **2004**, *30*, 527-529.
11. Siegel, J.P. Antiendotoxin antibodies. *Ann. Intern. Med.* **1995**, *122*, 315-316.
12. David, S.A. Towards a rational development of anti-endotoxin agents: novel approaches to sequestration of bacterial endotoxins with small molecules. *J. Mol. Recognit.* **2001**, *14*, 370-387.
13. David, S.A.; Silverstein, R.; Amura, C.R.; Kielian, T.; Morrison, D.C. Lipopolyamines: novel antiendotoxin compounds that reduce mortality in experimental sepsis caused by gram-negative bacteria. *Antimicrob. Agents Chemother.* **1999**, *43*, 912-919.

14. Hase, S.; Rietschel, E.T. Isolation and analysis of the lipid A backbone. Lipid A structure of lipopolysaccharides from various bacterial groups. *Eur. J. Biochem.* **1976**, *63*, 101-107.
15. Kim, K.S.; Kang, J.H.; Cross, A.S.; Kaufman, B.; Zollinger, W.; Sadoff, J. Functional activities of monoclonal antibodies to the O side chain of *Escherichia coli* lipopolysaccharides in vitro and in vivo. *J. Infect. Dis.* **1988**, *157*, 47-53.
16. Raetz, C.R.; Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **2002**, *71*, 635-700.
17. Vaarala, O.; Vaara, M.; Palosuo, T. Effective inhibition of cardiolipin-binding antibodies in gram-negative infections by bacterial lipopolysaccharide. *Scand. J. Immunol.* **1988**, *28*, 607-612.
18. Zahringer, U.; Lindner, B.; Rietschel, E.T. Molecular structure of lipid A, the endotoxic center of bacterial lipopolysaccharides. *Adv. Carbohydr. Chem. Biochem.* **1994**, *50*, 211-276.
19. Gozalbes, R.; Brun-Pascaud, M.; Garcia-Domenech, R.; Galvez, J.; Girard, P.M.; Doucet, J.P.; Derouin, F. Prediction of quinolone activity against *Mycobacterium avium* by molecular topology and virtual computational screening. *Antimicrob. Agents Chemother.* **2000**, *44*, 2764-2770.
20. Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Naenna, T.; Prachayasittikul, V. Quantitative structure-imprinting factor relationship of molecularly imprinted polymers. *Biosens. Bioelectron.* **2007**, *22*, 3309-3317.
21. Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Naenna, T.; Prachayasittikul, V. Prediction of bond dissociation enthalpy of antioxidant phenols by support vector machine. *J. Mol. Graph. Model.* **2008**, *27*, 188-196.
22. Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Tansila, N.; Naenna, T.; Prachayasittikul, V. Prediction of GFP spectral properties using artificial neural network. *J. Comput. Chem.* **2007**, *28*, 1275-1289.
23. Nantasenamat, C.; Naenna, T.; Isarankura-Na-Ayudhya, C.; Prachayasittikul, V. Quantitative prediction of imprinting factor of molecularly imprinted polymers by artificial neural network. *J. Comput. Aid. Mol. Des.* **2005**, *19*, 509-524.
24. Nantasenamat, C.; Naenna, T.; Isarankura-Na-Ayudhya, C.; Prachayasittikul, V. Recognition of DNA Splice Junction via Machine Learning Approaches. *Excli J.* **2005**, *4*, 114-129.
25. Nantasenamat, C.; Piacham, T.; Tantimongcolwat, T.; Naenna, T.; Isarankura-Na-Ayudhya, C.; Prachayasittikul, V. QSAR model of the quorum-quenching *N*-acyl-homoserine lactone lactonase activity. *J. Biol. Syst.* **2008**, *16*, 279-293.
26. Worachartcheewan, A.; Nantasenamat, C.; Naenna, T.; Isarankura-Na-Ayudhya, C.; Prachayasittikul, V. Modeling the activity of furin inhibitors using artificial neural network. *Eur. J. Med. Chem.* **2008**, 44, 1664-1673.
27. Burns, M.R.; Wood, S.J.; Miller, K.A.; Nguyen, T.; Cromer, J.R.; David, S.A. Lysine-spermine conjugates: hydrophobic polyamine amides as potent lipopolysaccharide sequestrants. *Bioorg. Med. Chem.* **2005**, *13*, 2523-2536.
28. Frecer, V.; Ho, B.; Ding, J.L. De Novo Design of Potent Antimicrobial Peptides *Antimicrob. Agents Chemother.* **2004**, *48*, 3349-3357.
29. Liu, D.; DeGrado, W.F. De novo design, synthesis, and characterization of antimicrobial beta-peptides. *J. Am. Chem. Soc.* **2001**, *123*, 7553-7559.

30. Wakefield, D.H.; Klein, J.J.; Wolff, J.A.; Rozema, D.B. Membrane Activity and Transfection Ability of Amphipathic Polycations as a Function of Alkyl Group Size. *Bioconjugate Chem.* **2005**, *16*, 1204-1208.
31. Burns, M.R.; Jenkins, S.A.; Wood, S.J.; Miller, K.; David, S.A. Structure-activity relationships in lipopolysaccharide neutralizers: design, synthesis, and biological evaluation of a 540-membered amphipathic bisamide library. *J. Comb. Chem.* **2006**, *8*, 32-43.
32. Zupan, J.; Gasteiger, J., *Neural Networks in Chemistry and Drug Design*. 2nd Ed.; Wiley-VCH: Weinheim, Germany, 1999.
33. Bak, A.; Polanski, J. A 4D-QSAR study on anti-HIV HEPT analogues. *Bioorg. Med. Chem.* **2006**, *14*, 273-279.
34. Loukas, Y.L. Artificial neural networks in liquid chromatography: efficient and improved quantitative structure-retention relationship models. *J. Chromatogr. A* **2000**, *904*, 119-129.
35. Zhang, R.; Yan, A.; Liu, M.; Liu, H.; Hu, Z. Application of artificial neural networks for prediction of the retention indices of alkylbenzenes. *Chemometr. Intell. Lab. Syst.* **1999**, *45*, 113-120.
36. Agatonovic-Kustrin, S.; Zecevic, M.; Zivanovic, L. Use of ANN modelling in structure--retention relationships of diuretics in RP-HPLC. *J. Pharm. Biomed. Anal.* **1999**, *21*, 95-103.
37. Niculescu, S.P. Artificial neural networks and genetic algorithms in QSAR. *J. Mol. Struct. (Theochem)* **2003**, *622*, 71-83.
38. NIST/SEMATECH. *E-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/.
39. Miller, K.A.; Suresh Kumar, E.V.; Wood, S.J.; Cromer, J.R.; Datta, A.; David, S.A. Lipopolysaccharide sequestrants: structural correlates of activity and toxicity in novel acylhomospermines. *J. Med. Chem.* **2005**, *48*, 2589-2599.
40. MarvinSketch, Version 5.2. ChemAxon: Budapest, Hungary; http://www.chemaxon.com/product/msketch.html.
41. RECON, Version 5.5. Rensselaer Polytechnic Institute: Troy, New York, U.S.A.; http://www.drugmining.com/files/RECON/recondoc/WinRecon.html.
42. Krasowski, M.D.; Hong, X.; Hopfinger, A.J.; Harrison, N.L. 4D-QSAR analysis of a set of propofol analogues: mapping binding sites for an anesthetic phenol on the GABA(A) receptor. *J. Med. Chem.* **2002**, *45*, 3210-3221.
43. Whitley, D.C.; Ford, M.G.; Livingstone, D.J. Unsupervised forward selection: a method for eliminating redundant variables. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1160-1168.
44. Witten, I.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann: San Francisco, CA, USA, 2000.

Sample Availability: Not available.

© 2009 by the authors; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).