molecular pharmaceutics

articles

Combating the Threat of Anthrax: A Quantitative Structure−Activity Relationship Approach

Rajeshwar P. Verma* and Corwin Hansch

Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, California 91711

Received January 23, 2008; Revised Manuscript Received May 22, 2008; Accepted May 27, 2008

Abstract: Bacterial agents or products more likely to be used as biological weapons of mass destruction are *Bacillus anthracis, Francisella tularensis, Yersinia pestis*, and the neurotoxin of *Clostridium botulinum*. Anthrax is an acute infectious disease with a high mortality rate caused by *Bacillus anthracis*, reinforcing the need for better adjunctive therapy and prevention strategies. In this paper, we developed 7 QSAR models on penicillin-based inhibitors of the class A and B β-lactamases from *B. anthracis* and inhibitors of anthrax lethal factor to understand the chemical−biological interactions. Hydrophobic and steric factors are found to be the most important determinants of the activity. Internal (cross-validation (q^2), quality factor (Q), Fischer statistics (F), and Y-randomization) and external validation tests have validated all the QSAR models.

Keywords: Anthrax; *Bacillus anthracis*; hydrophobicity; molar refractivity; molar volume; QSAR

Introduction

Anthrax is an acute infectious disease caused by *Bacillus anthracis*, a Gram-positive, spore-forming, aerobic, and rod-shaped bacterium. Humans typically acquire the disease through contact with infected livestock or animal products. However, several fatal anthrax cases have been reported as a result of bioterrorism. There were over twenty anthrax cases caused by malicious release of *B. anthracis* spores during the 2001 terrorist attack in the US.[1–4] In recent years, anthrax has received much attention worldwide from both the scientific community and the other sectors of society due to the major threat of *B. anthracis* that is connected to bioterrorism and biological warfare. Spores produced in dry form can be spread by means of letters or aerosols.[5,6]

Three types of anthrax are distinguished, depending on the route of spore entry into the body, i.e., cutaneous, gastrointestinal, and inhalational. In cutaneous anthrax, spores

* Author to whom correspondence should be addressed. Mailing address: Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, CA 91711. Phone: (909) 607-4249. Fax: (909) 607-7726. E-mail: rverma@pomona.edu.

(1) Puhar, A.; Montecucco, C. Where and how do anthrax toxins exit endosomes to intoxicate host cells? *Trends Microbiol.* **2007**, *15*, 477–482.

(2) Jernigan, J. A.; Stephens, D. S.; Ashford, D. A.; Omenaca, C.; Topiel, M. S.; Galbraith, M.; Tapper, M.; Fisk, T. L.; Zaki, S.; Popovic, T.; Meyer, R. F.; Quinn, C. P.; Harper, S. A.; Fridkin, S. K.; Sejvar, J. J.; Shepard, C. W.; McConnell, M.; Guarner, J.; Shieh, W. J.; Malecki, J. M.; Gerberding, J. L.; Hughes, J. M.; Perkins, B. A. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. *Emerging Infect. Dis.* **2001**, *7*, 933–944.

(3) Jernigan, D. B.; Raghunathan, P. L.; Bell, B. P.; Brechner, R.; Bresnitz, E. A.; Butler, J. C.; Cetron, M.; Cohen, M.; Doyle, T.; Fischer, M.; Greene, C.; Griffith, K. S.; Guarner, J.; Hadler, J. L.; Hayslett, J. A.; Meyer, R.; Petersen, L. R.; Phillips, M.; Pinner, R.; Popovic, T.; Quinn, C. P.; Reefhuis, J.; Reissman, D.; Rosenstein, N.; Schuchat, A.; Shieh, W.-J.; Siegal, L.; Swerdlow, D. L.; Tenover, F. C.; Traeger, M.; Ward, J. W.; Weisfuse, I.; Wiersma, S.; Yeskey, K.; Zaki, S.; Ashford, D. A.; Perkins, B. A.; Ostroff, S.; Hughes, J.; Fleming, D.; Koplan, J. P.; Gerberding, J. L. Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. *Emerging Infect. Dis.* **2002**, *8*, 1019–1028.

(4) Griffith, K. S.; Mead, P.; Armstrong, G. L.; Painter, J.; Kelley, K. A.; Hoffmaster, A. R.; Mayo, D.; Barden, D.; Ridzon, R.; Parashar, U.; Teshale, E. H.; Williams, J.; Noviello, S.; Perz, J. F.; Mast, E. E.; Swerdlow, D. L.; Hadler, J. L. Bioterrorism-related inhalational anthrax in an elderly woman, Connecticut, 2001. *Emerging Infect. Dis.* **2003**, *9*, 681–688.

(5) Matsumoto, G. Bioterrorism Anthrax powder: state of the art. *Science* **2003**, *302*, 1492–1497.

(6) Atlas, R. M. The medical threat of biological weapons. *Crit. Rev. Microbiol.* **1998**, *24*, 157–168.

10.1021/mp8000149 CCC: $40.75
Published on Web 07/09/2008

infect via skin abrasions, i.e., bacteria are usually contained at the site of entry, and the disease generally resolves without major consequences. Gastrointestinal anthrax, resulting from ingestion of spore-contaminated meat, is exceedingly rare, but often fatal. The most dangerous form, inhalational anthrax, results from inhalation of spores aerosolized in sufficiently small particle sizes ($\leq 5\ \mu$m) to reach the alveoli. Alveolar macrophages phagocytose the spores and are believed to serve both as the sites of germination and as the vehicles for transporting the bacteria to mediastinal lymph nodes. The bacteria then rapidly disseminate to the blood stream causing fatality approaching 100% if untreated.[7,8]

Based on the scientific calculations, the WHO has given an estimate of the possible casualties following the release of 50 kg of dried purified anthrax spores by aerosolization for 2 h on a city of 500,000 inhabitants. The above estimate of the WHO has indicated the hospitalization of 12,500 individuals (10% of those incapacitated), the need of antibiotics for 125,000 people for 60 days, and the disposal of 95,000 dead. This would certainly lead to a rapid total breakdown in the medical resources and the civilian infrastructures.[9–11] In another cited summary of an anthrax threat, it has been estimated that the aerosolized release of 100 kg of anthrax spores over Washington, DC, would yield between 130,000 and 3 million deaths. The number of fatalities could be reduced by prophylactic treatment of the affected population within 24 to 48 h of exposure with antibiotics such as penicillin, doxycycline, and ciprofloxacin.[12]

Unfortunately, the isolates of *B. anthracis* recovered in France, including 1 isolate from a human source, 28 from animal sources, and 67 from other environmental sources, revealed resistance to penicillin G and amoxicillin in 11.5% of the isolates.[13] On the other hand, the repeated subculturing has demonstrated that *B. anthracis* also acquired resistance to ciprofloxacin and doxycycline. This potential is particularly alarming because a course of the antibiotics up to 60 days was approved as a recommended therapy for the inhalation of anthrax.[14] Bacterial resistance to β-lactam antibiotics has been attributed mostly to the production of β-lactamases, which cleave the amide bond in the β-lactam ring of the antibiotic, rendering it ineffective and harmless to the bacteria.[15,16] β-Lactamases are divided into four classes based on primary sequence homology. Catalysis by classes A, C, and D proceeds via a serine-bound acyl intermediate, whereas class B enzymes require zinc for the activity.[17] It has been suggested that the penicillin-susceptible Sterne strain of *B. anthracis* harbors two chromosomal genes, *bla1* and *bla2*, that encode a class A (Bla1) and class B (Bla2) β-lactamase. Bla1 has shown a preference for the hydrolysis of penicillins over cephalosporins and is most efficient with benzylpenicillin, ampicillin, and amoxicillin. On the other hand, Bla2 is able to hydrolyze both penicillins over cephalosporins with similar catalytic efficiencies.[18,19]

B. anthracis owes its lethality to two major virulence factors: capsule and toxin. The capsule of *B. anthracis*, composed of poly-γ-D-glutamic acid (γ-D-PGA), is an essential virulence factor that inhibits innate host defense through its antiphagocytic action. γ-D-PGA is a poor immunogenic, but when covalently bound to a carrier protein, it elicits serum antibodies.[20,21] The anthrax toxin consists of three proteins: lethal factor (LF), edema factor (EF), and protective antigen (PA). Individually, none of these are toxic, but a mixture of PA and LF (lethal toxin; LeTx) can cause lethal shock in experimental animals, and a mixture of PA and EF (edema toxin; EdTx) induces edema at the site of injection. Lethal factor (LF) is a Zn-dependent metalloprotease that cleaves the N-terminal portion of the most isoforms of the mitogen-activated protein kinase kinases (MAPKKs or MEKs). Edema factor (EF) is a calcium- and calmodulin-dependent adenylyl cyclase that perturbs ion homeostasis and cell signaling by greatly increasing the cytosolic cAMP concentration. The third, protective antigen (PA), named for

(7) Collier, R. J.; Young, J. A. T. Anthrax toxin. *Annu. Rev. Cell Dev. Biol.* **2003**, *19*, 45–70.
(8) Guidi-Rontani, C.; Mock, M. Macrophage interactions. *Curr. Top. Microbiol. Immunol.* **2002**, *271*, 115–141.
(9) WHO. *Health aspect of chemical and biological weapons*; WHO: Geneva, Switzerland, 1970.
(10) Spencer, R. C. Bacillus anthracis. *J. Clin. Pathol.* **2003**, *56*, 182–187.
(11) Bouzianas, D. G. Potential biological targets of bacillus anthracis in anti-infective approaches against the threat of bioterrorism. *Expert Rev. Anti-Infect. Ther.* **2007**, *5*, 665–684.
(12) Stubbs II, M. T. Anthrax X-rayed: new opportunities for biodefence. *Trends Pharmacol. Sci.* **2002**, *23*, 539–541.
(13) Cavallo, J.-D.; Ramisse, F.; Girardet, M.; Vaissaire, J.; Mock, M.; Hernandez, E. Antibiotic susceptibilities of 96 isolates of bacillus anthracis isolated in France between 1994 and 2000. *Antimicrob. Agents Chemother.* **2002**, *46*, 2307–2309.
(14) Brook, I.; Elliott, T. B.; Pryor, H. I. 2nd.; Sautter, T. E.; Gnade, B. T.; Thakar, J. H.; Knudson, G. B. In vitro resistance of Bacillus anthracis Sterne to doxycycline, macrolides and quinolones. *Int. J. Antimicrob. Agents* **2001**, *18*, 559–562.
(15) Petrosino, J.; Cantu, C., III; Palzkill, T. Beta-Lactamases: protein evolution in real time. *Trends Microbiol.* **1998**, *6*, 323–327.
(16) Theuretzbacher, U. Beta lactamases and beta - lactamase inhibitors. *Chemother. J.* **1998**, *7*, 136–142.
(17) Majiduddin, F. K.; Materon, I. C.; Palzkill, T. G. Molecular analysis of beta-lactamase structure and function. *Int. J. Med. Microbiol.* **2002**, *292*, 127–137.
(18) Chen, Y.; Succi, J.; Tenover, F. C.; Koehler, T. M. β-lactamase genes of the penicillin-susceptible bacillus anthracis sterne strain. *J. Bacteriol.* **2003**, *185*, 823–830.
(19) Beharry, Z.; Chen, H.; Gadhachanda, V. R.; Buynak, J. D.; Palzkill, T. Evaluation of penicillin-based inhibitors of the class A and B beta-lactamases from bacillus anthracis. *Biochem. Biophys. Res. Commun.* **2004**, *313*, 541–545.
(20) Kubler-Kielb, J.; Liu, T.-Y.; Mocca, C.; Majadly, F.; Robbins, J. B.; Schneerson, R. Additional conjugation methods and immunogenicity of Bacillus anthracis poly-γ-D-glutamic acid-protein conjugates. *Infect. Immun.* **2006**, *74*, 4744–4749.
(21) Candela, T.; Fouet, A. Poly-gamma-glutamate in bacteria. *Mol. Microbiol.* **2006**, *60*, 1091–1098.

its effectiveness in inducing protective immunity against anthrax, is the vehicle for delivery of LF and EF to the cytosol.[1,7,22,23]

In response to the threat of a deliberate release of lethal anthrax spores, the current treatment includes the administration of appropriate antibiotics often prior to the advent of symptoms. Other anti-infective strategies, such as vaccines and antitoxins, are available, but there are concerns over their effectiveness and toxicity. The current anthrax vaccine approved by the FDA for human use in the United States is AVA (anthrax vaccine adsorbed), which consists of a *B. anthracis* culture supernatant that has been adsorbed onto an aluminum adjuvant. This vaccine is able to stimulate antibodies against PA and can provide protection in animal models of anthrax. However, in humans AVA requires six administrations over 18 months in addition to annual booster immunizations, a dosing regimen that is not ideal when there should be rapid vaccination before, or in response to, bioterrorist events.[24–26] Since the major target of protective immunity is PA, the main emphasis in the development of a second-generation vaccine is on the induction of immunity to purified preparations of PA. There is ongoing work worldwide to improve existing vaccines and to develop new vaccines on the basis of recombinant bacterial or viral live vectors, DNA vectors, and through the addition of new adjuvants. Other improvements include the development of anticapsule conjugate vaccines and supplementing the existing toxoid vaccine with an antibacterial component.[11] Among the three proteins of anthrax toxin, LF has been shown to act as the key virulence factor for *B. anthracis* since mouse death is seen in the absence of edema factor, but not in the absence of lethal factor.[27] Because of its vital role in anthrax pathogenesis, much work has focused on finding potent inhibitors of LF. Currently, there are several potent LF inhibitors are available in the literature. However, there are significantly fewer inhibitors with cell-based activity, an important point since LF functions in the cytosol. Therefore, it is essential that agents not only inhibit LF cleavage of substrate but also are active in a lethal toxin challenge of macrophages. Identifying compounds that inhibit LF, yet are able to enter and remain active in cells, constitutes a major challenge in developing effective anti-lethal factor therapeutics.[28,29]

Current antibiotic treatments are limited due to resistance and patient age restrictions; thus, additional targets for therapeutic intervention are needed. One possible candidate is dihydrofolate reductase (DHFR), a biosynthetic enzyme necessary for anthrax pathogenicity. *B. anthracis* exhibits a natural resistance to trimethoprim (TMP), a clinically used antibacterial DHFR inhibitor, due to a lack of affinity between the enzyme and the inhibitor.[30] The crystal structure of DHFR was obtained from *B. anthracis* (baDHFR) in complex with methotrexate (MTX) at 2.4 Å resolution. The structure reveals the crucial interactions required for MTX binding and a putative molecular basis for how baDHFR has natural resistance to trimethoprim (TMP). This structure also allows insights for designing selective baDHFR inhibitors that will have weak affinities for the human enzyme.[31] In a recent report, the selective inhibition of *B. anthracis* rDHFR and the *in vitro* activity against *B. anthracis* have demonstrated that 2,4-diaminopyrimidine derivatives with dihydrophthalazine side chains have the potential to be developed into clinically important therapeutic choices for the treatment of infections caused by TMP-resistant bacteria, such as *B. anthracis*.[32]

In the present paper, we discuss the QSAR (quantitative structure−activity relationship) studies on penicillin-based inhibitors of the class A and B β-lactamases from *B. anthracis* as well as inhibitors of anthrax lethal factor to understand the chemical−biological interactions in mathematical terms. In the past 45 years, the use of QSAR (one of the well-developed areas in computational chemistry) has become increasingly helpful in understanding many aspects of chemical−biological interactions in the drug-design process and pesticide research as well as in the area of

(22) Duesbery, N. S.; Webb, C. P.; Leppla, S. H.; Gordon, V. M.; Klimpel, K. R.; Copeland, T. D.; Ahn, N. G.; Marianne, M. K.; Fukasawa, K.; Paull, K. D.; Vande Woude, G. F. Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **1998**, *280*, 734–737.

(23) Young, J. A. T.; Collier, R. J. Anthrax toxin: receptor binding, internalization, pore formation, and translocation. *Annu. Rev. Biochem.* **2007**, *76*, 243–265.

(24) Pitt, M. L. M.; Little, S. F.; Ivins, B. E.; Fellows, P.; Barth, J.; Hewetson, J.; Gibbs, P.; Dertzbaugh, M.; Friedlander, A. M. In vitro correlate of immunity in a rabbit model of inhalational anthrax. *Vaccine* **2001**, *19*, 4768–4773.

(25) Xie, H.; Gursel, I.; Ivins, B. E.; Singh, M.; O'Hagan, D. T.; Ulmer, J. B.; Klinman, D. M. CpG oligodeoxynucleotides adsorbed onto polylactide-co-glycolide microparticles improve the immunogenicity and protective activity of the licensed anthrax vaccine. *Infect. Immun.* **2005**, *73*, 828–833.

(26) McConnell, M. J.; Hanna, P. C.; Imperiale, M. J. Adenovirus-based prime-boost immunization for rapid vaccination against anthrax. *Mol. Ther.* **2007**, *15*, 203–210.

(27) Pezard, C.; Berche, P.; Mock, M. Contribution of individual toxin components to virulence of bacillus anthracis. *Infect. Immun.* **1991**, *59*, 3472–3477.

(28) Montecucco, C.; Tonello, F.; Zanotti, G. Stop the Killer: How to inhibit the anthrax lethal factor metalloprotease. *Trends Biochem. Sci.* **2004**, *29*, 282–285.

(29) Gaddis, B. D.; Avramova, L. V.; Chmielewski, J. Inhibitors of anthrax lethal factor. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4575–4578.

(30) Barrow, E. W.; Bourne, P. C.; Barrow, W. W. Functional cloning of bacillus anthracis dihydrofolate reductase and confirmation of natural resistance to trimethoprim. *Antimicrob. Agents Chemother.* **2004**, *48*, 4643–4649.

(31) Bennett, B. C.; Xu, H.; Simmerman, R. F.; Lee, R. E.; Dealwis, C. G. Crystal structure of the anthrax drug target, bacillus anthracis dihydrofolate reductase. *J. Med. Chem.* **2007**, *50*, 4374–4381.

(32) Barrow, E. W.; Dreier, J.; Reinelt, S.; Bourne, P. C.; Barrow, W. W. In vitro efficacy of new antifolates against trimethoprim-resistant bacillus anthracis. *Antimicrob. Agents Chemother.* **2007**, *51*, 4447–4452.

toxicology.[33] The QSAR approach employs extra-thermodynamically derived and computational-based descriptors to correlate biological activity in isolated receptors, in cellular systems, and in vivo. Four standard molecular descriptors are routinely used in QSAR analysis: electronic, hydrophobic, steric, and topological indices. These descriptors are invaluable in helping to delineate a large number of receptor−ligand interactions that are critical to biological processes. This method is useful in elucidating the mechanisms of chemical−biological interaction in various biomolecules, particularly enzymes, membranes, organelles, and cells, as well as in human.[34–36] It has also been utilized for the evaluation of absorption, distribution, metabolism, and excretion (ADME) phenomena in many organisms and whole animal studies.[37,38] The quality of a QSAR model depends strictly on the type and quality of the data, and not on the hypotheses, and is valid only for the compound structure analogues to those used to build the model. QSAR models can stand alone to augment other computational approaches or can be examined in tandem with equations of a similar mechanistic genre to establish their authenticity and reliability.[39]

Materials and Methods

All the data has been collected from the literature (see individual QSAR for respective references). IC_{50} is the molar concentration of a compound cause 50% inhibition of the class A and B β-lactamases from *B. anthracis* or anthrax lethal factor; log $1/IC_{50}$ is the subsequent dependent variable that defines the biological parameter for QSAR development. Physicochemical descriptors are autoloaded, and multiregression analyses (MRA) are used to derive the QSAR by using the C-QSAR program.[40] Selection of descriptors is made on the basis of permutation and correlation matrices among the descriptors in order to avoid collinearity problems. Details about the C-QSAR program, the search engine, the choice of parameters and their use in the development of QSAR models have already been discussed.[41,42]

The descriptors used in this paper have been discussed previously in detail along with their application.[34] Briefly, Clog P is the calculated partition coefficient of a compound in *n*-octanol/water and is a measure of its hydrophobicity, whereas π is the hydrophobic parameter for substituents only. CMR is the calculated molar refractivity for the whole molecule, which is calculated from the Lorentz−Lorenz equation: $[(n^2 - 1)/(n^2 + 2)](MW/\delta)$, where n is the refractive index, MW is the molecular weight, and δ is the density of the substance. Molar refractivity (MR) is dependent on both the volume and polarizability. It can be used for a substituent or for the whole molecule. MR is thus a means of characterizing the bulk and polarizability of a substituent or compound. Although it contains no information about the shape, it has found considerable use in biological QSAR, where intermolecular effects predominate. MR is usually scaled at 0.1 to make it equiscalar with π. MgVol is the molar volume for the whole molecule and calculated by the method of Abraham and McGowan.[43] The indicator variable I is assigned the value of 1 or 0 for special features with special effects that cannot be parametrized, and has been explained wherever used.

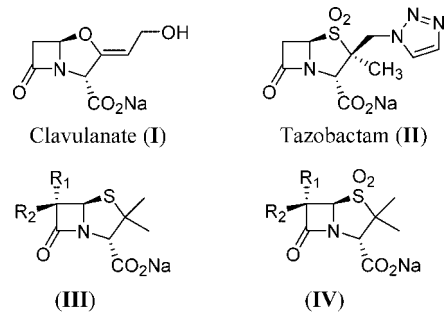
In QSAR equations, n is the number of data points, r is the correlation coefficient between observed values of the dependent and the values calculated from the equation, r^2 is the square of the correlation coefficient and represents the goodness of fit, q^2 is the cross-validated r^2 (a measure of the quality of the QSAR model), and s is the standard deviation. The cross-validated r^2 (q^2) is obtained by using a leave-one-out (LOO) procedure as described by Cramer et al.[44] Q is the quality factor, for which $Q = r/s$. Chance correlation, due to the excessive number of parameters (which also increases the r and s values), can, therefore, be detected by the examination of the Q value. High values of Q indicate the high predictive power of the QSAR models and the lack of "overfitting". F represents the Fischer statistics (Fischer ratio), $F = fr^2/[(1 - r^2)m]$, where f is the number of degrees of freedom $[f = n - (m + 1)]$, n is the number of data points, and m is the number of variables. The F value is actually the ratio between explained and unexplained variance for a given number of degrees of

(33) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* **1962**, *194*, 178–180.
(34) Hansch, C.; Leo, A. In *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
(35) Selassie, C. D.; Garg, R.; Kapur, S.; Kurup, A.; Verma, R. P.; Mekapati, S. B.; Hansch, C. Comparative QSAR and the radical toxicity of various functional groups. *Chem. Rev.* **2002**, *102*, 2585–2605.
(36) Verma, R. P.; Kurup, A.; Mekapati, S. B.; Hansch, C. Chemical-biological interactions in human. *Bioorg. Med. Chem.* **2005**, *13*, 933–948.
(37) Hansch, C.; Leo, A.; Mekapati, S. B.; Kurup, A. QSAR and ADME. *Bioorg. Med. Chem.* **2004**, *12*, 3391–3400.
(38) Verma, R. P.; Hansch, C.; Selassie, C. D. Comparative QSAR studies on PAMPA/modified PAMPA for high throughput profiling of drug absorption potential with respect to Caco-2 cells and human intestinal absorption. *J. Comput. Aided. Mol. Des.* **2007**, *21*, 3–22.
(39) Selassie, C. D.; Mekapati, S. B.; Verma, R. P. QSAR: then and now. *Curr. Top. Med. Chem.* **2002**, *2*, 1357–1379.
(40) C-QSAR Program, BioByte Corp., 201W. 4th St., Suite 204, Claremont, CA 91711, USA, www.biobyte.com.
(41) Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Selassie, C. D. Chem-Bioinformatics: Comparative QSAR at the interface between chemistry and biology. *Chem. Rev.* **2002**, *102*, 783–812.
(42) Verma, R. P.; Hansch, C. Development of QSAR models using C-QSAR program: a regression program that has dual databases of over 21,000 QSAR models. *Nat. Protoc.* (DOI: 10.1038/nprot.2007.125). Freely available at http://www.natureprotocols.com/2007/03/05/development_of_qsar_models_usi_1.php.
(43) Abraham, M. H.; McGowan, J. C. The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography. *Chromatographia* **1987**, *23*, 243–246.
(44) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Cross validation, Bootstrapping and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.

freedom. Thus, it indicates a true relationship, or the significance level for MLR models. The modeling was taken to be optimal when Q reached a maximum together with F, even if slightly nonoptimal F values have normally been accepted. A significant decrease in F with the introduction of one additional variable (with increasing Q and decreasing s) could mean that the new descriptor is not as significant as expected; i.e., its introduction has endangered the statistical quality of the combination. However, the statistical quality could be improved by the introduction of a more convincing descriptor.[45,46] Compounds were deemed to be outliers on the basis of their deviation between observed and predicted activities from the equation (obsd − pred > 2s).[47–52] Outliers are those compounds which have unexpected biological activities and are unable to fit in a QSAR model. The presence of outliers is not only due to the possibility that the molecules may act by different mechanisms or interact with the receptor in different binding modes but also due to the intrinsic noise associated with both the original data and methodological aspects involved in the construction of a QSAR model.[53–55] Each regression equation includes 95% confidence limits for each term in parentheses.

Results and Discussion

1. Inhibitors of Class A (Bla1) and Class B (Bla2) β-Lactamases from *B. anthracis*. In an effort to find new penicillin-based β-lactamase inhibitors, three penicillin derivatives (**III**−**V**) along with the clinical inhibitors clavulanate (**I**) and tazobactam (**II**) have been evaluated by Beharry et al.[19] as the inhibitors of class A serine β-lactamase Bla1 and class B metallo-β-lactamase Bla2 from *B. anthracis* (Figure 1). From the data in Table 1, we derived eqs 1 and 2, respectively.



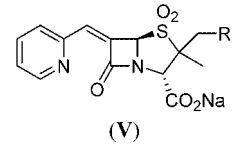


Figure 1. Structure of penicillin derivatives (**I**−**V**) used in the development of QSAR models (eqs 1 and 2).

Table 1. Biological and Physicochemical Parameters Used To Derive QSARs 1 and 2

	log 1/IC_{50}							
	eq *1*			eq *2*				
compd	obsd	pred	Δ	obsd	pred	Δ	Clog *P*	MgVol
clavulanate (**I**)	5.10	4.97	0.13	3.50	3.48	0.02	−1.42	1.20
tazobactam (**II**)	6.54	6.14	0.40	3.33	4.05	−0.72	−1.00	1.79
IIIa	3.09	3.02	0.07	5.22	5.04	0.18	1.52	1.59
IIIb	2.70	3.02	−0.32	4.89	5.04	−0.15	1.52	1.59
IVa	4.00	4.18	−0.18	4.57	4.75	−0.18	0.69	1.71
IVb	3.75	4.18	−0.43	4.85	4.75	0.10	0.69	1.71
Va	7.06	7.22	−0.16	5.15	4.66	0.49	−0.50	2.38
Vb	6.66	6.99	−0.33	5.30	5.02	0.28	0.09	2.52
Vc	7.08	6.60	0.48	5.70	6.03	−0.33	1.70	2.99
Vd	7.10	7.52	−0.42	5.22	5.25	−0.03	0.24	2.78
Ve[a]	6.62	6.42	0.20	5.30	6.18	−0.88	1.98	3.03
Vf	6.96	6.83	0.13	6.00	5.96	0.04	1.52	3.01
Vg	7.70	8.03	−0.33	5.52	5.70	−0.18	0.71	3.15
Vh	6.27	5.50	0.77	5.30	4.83	0.47	0.34	2.07

[a] Not included in the derivation of QSAR 2.

1.1. Inhibition of the Class A (Bla1) β-Lactamase from *B. anthracis* by Penicillin Derivatives (I−V).[19] (Table 1.)

(45) Pogliani, L. Modeling with special descriptors derived from a medium-sized set of connectivity indices. *J. Phys. Chem.* **1996**, *100*, 18065–18077.
(46) Pogliani, L. From molecular connectivity indices to semiempirical connectivity terms: Recent trends in graph theoretical descriptors. *Chem. Rev.* **2000**, *100*, 3827–3858.
(47) Selassie, C. D.; Kapur, S.; Verma, R. P.; Rosario, M. Cellular apoptosis and cytotoxicity of phenolic compounds: a quantitative structure-activity relationship study. *J. Med. Chem.* **2005**, *48*, 7234–7242.
(48) Verma, R. P.; Hansch, C. Cytotoxicity of organic compounds against ovarian cancer cells: A quantitative structure-activity relationship study. *Mol. Pharmaceutics* **2006**, *3*, 441–450.
(49) Verma, R. P.; Hansch, C. Matrix metalloproteinases (MMPs): Chemical-biological functions and QSARs. *Bioorg. Med. Chem.* **2007**, *15*, 2223–2268.
(50) Verma, R. P.; Hansch, C. Understanding human rhinovirus infections in terms of QSAR. *Virology* **2007**, *359*, 152–161.
(51) Hansch, C.; Verma, R. P. 20-(*S*)-Camptothecin analogues as DNA topoisomerase I inhibitors: A QSAR study. *ChemMedChem* **2007**, *2*, 1807–1813.
(52) Verma, R. P.; Hansch, C. Investigation of DNA-binding properties of organic molecules using quantitative structure-activity relationship (QSAR) models. *J. Pharm. Sci.* **2008**, *97*, 88–110.
(53) Verma, R. P.; Hansch, C. An approach toward the problem of outliers in QSAR. *Bioorg. Med. Chem.* **2005**, *13*, 4597–4621.
(54) Polanski, J.; Bak, A.; Gieleciak, R.; Magdziarz, T. Modeling Robust QSAR. *J. Chem. Inf. Model.* **2006**, *46*, 2310–2318.
(55) Ekins, S.; Mestres, J.; Testa, B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* **2007**, *152*, 9–20.

$$\log 1/IC_{50} = -1.03(\pm 0.26)\ \text{Clog}\ P + 2.69(\pm 0.41)\ \text{MgVol} + 0.90(\pm 0.30)$$
$$n = 14,\ r^2 = 0.951,\ s = 0.405,\ q^2 = 0.929,\ Q = 2.407,\ F_{2,11} = 106.745(7.206)$$
$$\text{Clog}\ P\ \text{vs MgVol},\ r = 0.427$$
$$\text{range in log } 1/IC_{50} = 7.70-2.70 \quad (1)$$

In this equation, F is the F-ratio between the variances of predicted and observed activities (within parentheses the figure refers to the F-value at the 99% level[56]). Clog P and MgVol are the calculated hydrophobicity and molar volume of the whole molecule, respectively. The unusual negative Clog P term suggests that the inhibitory activity of these molecules decreases with increasing their hydrophobicity. On the contrary, the increase in MgVol increases the inhibitory activity of these compounds (positive coefficient). The negative impact of Clog P may be offset by the large positive contribution (+2.69) of the steric term as represented by MgVol. In this model, 63% of the variance in the data is explained by the steric descriptor MgVol while the hydrophobic descriptor accounts for only 32% of the variance in the data. There is no high mutual correlation between Clog P and MgVol (r = 0.427). Thus, more hydrophilic and bulkier compounds of penicillin derivatives (**I**−**V**) would present better inhibitory activity.

1.2. Inhibition of the Class B (Bla2) β-Lactamase from *B. anthracis* by Penicillin Derivatives (I−V).[19] (Table 1.)

$$\log 1/IC_{50} = 0.44(\pm 0.25)\ \text{Clog}\ P + 0.65(\pm 0.38)\ \text{MgVol} + 3.33(\pm 0.83)$$
$$n = 13,\ r^2 = 0.821,\ s = 0.362,\ q^2 = 0.706,\ Q = 2.503,\ F_{2,10} = 22.933(7.559)$$
$$\text{Clog}\ P\ \text{vs MgVol},\ r = 0.341;\ \text{outlier: } \mathbf{Ve}$$
$$\text{range in log } 1/IC_{50} = 6.00-3.33 \quad (2)$$

A positive hydrophobic parameter is present in QSAR 2, expressing that highly hydrophobic compounds would be more active. This suggests that the enzyme may have a hydrophobic pocket. MgVol is the calculated molar volume and has sometimes been used lately instead of molar refractivity as an alternative theoretically assessable bulk factor. Thus, the positive MgVol brings out a favorable steric effect. One compound (**Ve**) was deemed to be an outlier on the basis of its deviations (obsd − pred > 2s). Since, both the descriptors in eq 2 account for nearly equal contribution of the variance in the data; the parallel increment in the values of both the descriptors (Clog P and MgVol) will result in the better inhibitory activity of compounds (**I**−**V**) against class B (Bla2) β-lactamases from *B. anthracis*.

Although the same compounds (**I**−**V**) and their same descriptors have been used in the formulation of QSARs (eqs 1 and 2), the resulting two QSARs are not reasonably similar. Equation 1 has large Clog P (−1.03) and MgVol (+2.69) terms. On the other hand, the eq 2 has very low Clog P (+0.44) and MgVol (+0.65) terms as compared to eq 1. Thus, eq 1 suggests more hydrophilic/bulkier compounds, whereas eq 2 suggests more hydrophobic/bulkier compounds for better inhibitory activities. The different results from eqs 1 and 2 further suggest that the two classes of enzymes, i.e., class A serine β-lactamase Bla1 and class B metallo-β-lactamase Bla2 from *B. anthracis*, are different from each other. This is in agreement with the experimental results. It has been suggested that the penicillin-susceptible Sterne strain of *B. anthracis* harbors two chromosomal genes, *bla1* and *bla2*, that encode a class A (Bla1) and class B (Bla2) β-lactamase, which have different substrate specificities.[18] Catalysis by class A proceeds via a serine-bound acyl intermediate, whereas class B enzymes require zinc for the activity.[17]

Table 2. Biological, Physicochemical, and Structural Parameters Used To Derive QSAR 3

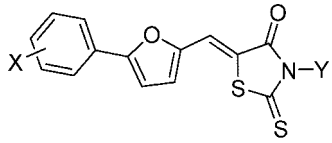
			log 1/IC_{50} (eq *3*)				
compd	X	Y	obsd	pred	Δ	π_Y	*I*
VIa	3-CF_3	C_6H_4(4-OCH_3)	3.52	4.10	−0.58	2.16	0
VIb	4-Cl	CH_2-2-furyl	3.82	4.31	−0.49	1.28	0
VIc	4-F	$CH_2CH{=}CH_2$	4.30	4.40	−0.10	1.11	0
VId	4-NO_2	C_6H_4(4-OH)	4.42	4.19	0.23	1.58	0
VIe	3-NO_2	CH_2-2-furyl	4.44	4.31	0.13	1.28	0
VIf	2-NO_2	$(CH_2)_2OC_6H_5$	4.50	4.10	0.40	2.10	0
VIg	3-NO_2	$(CH_2)_3COOH$	4.70	5.09	−0.39	0.26	0
VIh	2-NO_2	CH_2CH_2COOH	4.89	5.53	−0.64	−0.11	0
VIi	4-NO_2	CH_2-2-furyl	4.90	4.31	0.59	1.28	0
VIj[a]	3-Cl-4-COOH	CH_2-3-pyridyl	5.00	5.93	−0.93	0.61	1
VIk[a]	4-SO_2NH_2	CH_2COOH	5.04	5.88	−0.84	−0.37	0
VIl	3,4-Cl_2	H	5.13	5.39	−0.26	0.00	0
VIm	3-Cl-4-Br	H	5.15	5.39	−0.24	0.00	0
VIn	4-COOH	$CH_2C_6H_5$	5.22	5.27	−0.05	2.10	1
VIo	4-I	CH_2COOH	5.26	5.88	−0.62	−0.37	0
VIp	4-COOH	C_6H_4(4-NO_2)	5.32	5.27	0.05	1.99	1
VIq	3-NO_2	CH_2COOH	5.51	5.88	−0.37	−0.37	0
VIr	4-COOH	C_6H_4(3-CF_3)	5.54	5.54	0.00	3.13	1
VIs	4-NO_2	CH_2CH_2COOH	5.57	5.53	0.04	−0.11	0
VIt	3-Cl-4-COOH	$CH_2CH{=}CH_2$	5.57	5.57	0.00	1.11	1
VIu	4-Br	$(CH_2)_3COOH$	5.64	5.09	0.55	0.26	0
VIv	4-Cl	CH_2COOH	6.05	5.88	0.17	−0.37	0
VIw	4-Br	CH_2COOH	6.07	5.88	0.19	−0.37	0
VIx	2-NO_2-4-Cl	CH_2COOH	6.30	5.88	0.42	−0.37	0
VIy	2-NO_2	CH_2COOH	5.51	5.88	−0.37	−0.37	0
VIz	3-Cl-4-OCH3	CH_2COOH	6.53	5.88	0.65	−0.37	0
VIaa	3,4-Cl_2	CH_2COOH	6.58	5.88	0.70	−0.37	0

[a] Not included in the derivation of QSAR 3.

2. Inhibitors of Anthrax Lethal Factor (LF). 2.1. Inhibition of Anthrax Lethal Factor (LF) by Phenylfuran-2-ylmethylenerhodanine Derivatives (VI)..[57] (Table 2.) Lethal factor (LF), a key protein of anthrax lethal toxins, is a member of the zinc metalloprotease family, which also includes matrix metalloproteinases (MMPs). Matrix metalloproteinases (MMPs) are a large family of calcium-dependent zinc-containing endopeptidases, which are responsible for the tissue remodeling and degradation of the

(56) Bennett, C. A.; Franklin, N. L. In *Statistical Analysis in Chemistry and the Chemical Industry*; John Wiley & Sons: New York, 1967; pp 708−709. (5th printing).

extracellular matrix (ECM), including collagens, elastins, gelatin, matrix glycoproteins, and proteoglycan.[49] The proteolytic activity of LF MMPs is essential for the onset, progression, and lethality of anthrax. Because of the structural similarity of LF with the MMPs, compounds which inhibit both the MMPs and LF protease families could cause undesired side effects. On the basis of a fragment-based approach, a series of phenylfuran-2-ylmethylenerhodanine derivatives (**VI**) was designed, synthesized, and tested for their inhibitory activity against anthrax lethal factor (LF) by Pellecchia and co-workers.[57] On the basis of their observations, the authors suggested that the phenyl with a small X electronegative group significantly increases the inhibitory activity, whereas a small group containing a carboxylic moiety in position Y also seems to improve the potency. On the contrary, a large group, such as a substituted phenyl in Y, causes a dramatic decrease in the activity, especially if not balanced with an effective group in X. For mechanistic interpretation of these compounds, the authors also obtained the X-ray structure for LF in complex with a representative compound [B1-MFM3; X = 4-Cl, Y = CH_2CH_2COOH]. Analysis of the docked structure of B1-MFM3 revealed that the rhodanine ring is capable of interacting with Zn^{2+} metal ion via the thiazolidine sulfur atom. The carboxylic group of Y is pointing toward a hydrophilic region of the protein close to its surface, which explains the variability of the substitutions allowed at this position and the increased affinity of the compounds when Y is a small charged group. In addition, hydrophobic interactions between the phenyl ring and hydrophobic side chains of LF were also observed. However, electron density of the benzene ring is less evident in the X-ray structure of B1-MFM3, indicating a possible conformational mobility around the carbon−carbon bond of the para-substituted benzene ring and the larger available space around this portion of ligand. Equation 3 was developed from the data in Table 2.



VI

$$\log 1/IC_{50} = -1.21(\pm 0.41)\pi_Y + 0.28(\pm 0.19)\pi_Y{}^2 + 1.17(\pm 0.64)I + 5.39(\pm 0.22)$$
$$n = 25,\ r^2 = 0.732,\ s = 0.434,\ q^2 = 0.658,\ Q = 1.972,\ F_{3,21} = 19.119(4.874)$$
$$\text{inversion point for } \pi_Y = 2.14(1.55-4.99)$$
$$\text{outliers: } \mathbf{VIj}, \mathbf{VIk}$$
$$\text{range in } \log 1/IC_{50} = 6.58-3.52 \qquad (3)$$

This is an inverted parabolic correlation in terms of π_Y (calculated hydrophobicity of Y-substituents), which suggests that activity of these compounds first decreases as the hydrophobicity of the Y-substituents increases and that after a certain point (inversion point; $\pi_Y = 2.14$) activity begins to increase. This may correspond to an allosteric reaction.[41,58–61] The concept of allosteric proteins was developed in the early 1960s by Monod[62] and Koshland.[63] Allosteric (derived from Greek origin *allo*) means "the other". The binding of a regulatory molecule to a specific allosteric site of the protein, structurally distinct from the active site, brings about the alteration of the confirmation of the protein that indirectly modifies the properties of the biologically active site. This indirect mode of action via a distinct regulatory site is today the most commonly accepted meaning of the word "allosteric".[64,65] Indicator variable (*I*) is the pinpoint for the unusual activity of X = 4-COOH. Positive coefficient of *I* suggests that the presence of X = 4-COOH will promote the inhibitory activity. Two compounds (**VIj** and **VIk**) were deemed to be outliers on the basis of their high deviations (obsd − pred ≥ 2*s*).

The inverted parabolic correlation in terms of π_Y suggests that the Y group of phenylfuran-2-ylmethylenerhodanine derivatives (**VI**) is pointing toward a hydrophilic region of the protein close to its surface. Thus, Y substituents with negative π_Y may accommodate on the surface of the protein and the compound has acquired the higher activity. As the π_Y value increases, the activity of the compound decreases. It may be due to the repulsion between hydrophilic region of the protein and increased hydrophobic Y substituents, which probably brings about the alteration of the confirmation of the protein that indirectly modifies the properties of the biologically active site at the inversion point ($\pi_Y = 2.14$) and the activity of the compound begins to increase. This might be in agreement with the docked result as mentioned above. In addition, the presence of X = 4-COOH promotes the inhibitory activity.

(57) Forino, M.; Johnson, S.; Wong, T. Y.; Rozanov, D. V.; Savinov, A. Y.; Li, W.; Fattorusso, R.; Becattini, B.; Orry, A. J.; Jung, D.; Abagyan, R. A.; Smith, J. W.; Alibek, K.; Liddington, R. C.; Strongin, A. Y.; Pellecchia, M. Efficient synthetic inhibitors of anthrax lethal factor. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9499–9504.

(58) Verma, R. P. Understanding topoisomerase I and II in terms of QSAR. *Bioorg. Med. Chem.* **2005**, *13*, 1059–1067.

(59) Mekapati, S. B.; Kurup, A.; Verma, R. P.; Hansch, C. The role of hydrophobic properties of chemicals in promoting allosteric reactions. *Bioorg. Med. Chem.* **2005**, *13*, 3737–3762.

(60) Verma, R. P. An approach towards the quantitative structure-activity relationships for sulfamate-based estrone sulfatase inhibitors. *Lett. Drug Des. Discovery* **2005**, *2*, 205–218.

(61) Hansch, C.; Verma, R. P. Understanding tubulin/microtubule-taxane interactions: A quantitative structure-activity relationship study. *Mol. Pharmaceutics* **2008**, *5*, 151–161.

(62) Monod, J.; Changeux, J.-P.; Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **1963**, *6*, 306–329.

(63) Koshland, D. E., Jr. In *The Enzymes*; Boyer, P., Ed.; Academic Press: New York, 1970; Vol. 1, p 341.

(64) Changeux, J.-P.; Edelstein, S. J. Allosteric receptors after 30 years. *Neuron* **1998**, *21*, 959–980.

(65) Kovbasyuk, L.; Krämer, R. Allosteric supramolecular receptors and catalysts. *Chem. Rev.* **2004**, *104*, 3161–3187.

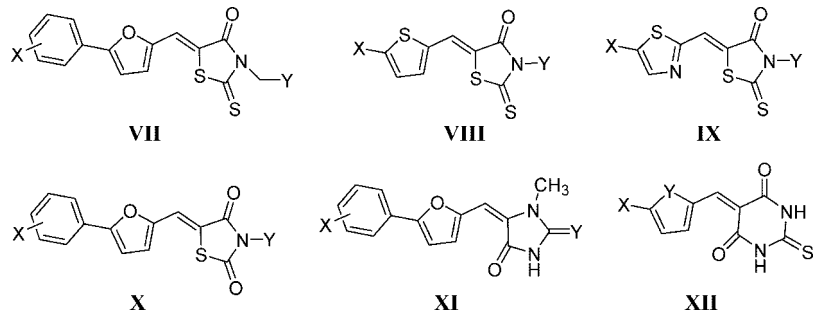
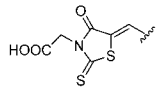


Figure 2. Structure of compounds (**VII**−**XII**) used in the development of QSAR model (eq 4).

Table 3. Biological, Physicochemical, and Structural Parameters Used to Derive QSAR 4[a]

compd	X	Y	log 1/IC_{50} (eq *4*)			CMR	I_{Furan}	$I_{Thiophene}$
			obsd	pred	Δ			
VIIa	4-Cl	CH_2COOH	5.77	6.24	–0.47	10.39	1	0
VIIb[A]	4-I	COOH	5.25	6.46	–1.21	10.74	1	0
VIIc	4-Cl	COOH	6.05	5.83	0.22	9.93	1	0
VIId	4-Br	COOH	6.07	6.09	–0.02	10.21	1	0
VIIe	2-NO_2	COOH	5.51	5.95	–0.44	10.05	1	0
VIIf	3-Cl-4-OCH_3	COOH	6.52	6.34	0.18	10.55	1	0
VIIg	3,4-Cl_2	COOH	6.59	6.26	0.33	10.42	1	0
VIIh	2-Cl-5-CF_3	COOH	6.72	6.27	0.45	10.44	1	0
VIIi	H	COOH	4.78	5.35	–0.57	9.44	1	0
VIIj	2-NO_2-4-Cl	COOH	6.30	6.34	–0.04	10.54	1	0
VIIIa	C_6H_5	CH_2COOH	6.13	6.59	–0.46	10.03	0	1
VIIIb	H	CH_2COOH	4.10	4.02	0.08	7.52	0	1
VIIIc[A]	4-Cl-C_6H_4	CH_2COOH	5.49	6.99	–1.50	10.52	0	1
VIIId		CH_2COOH	5.80	5.79	0.01	12.54	0	1
VIIIe[A]	2-thienyl	CH_2COOH	5.64	6.41	–0.77	9.84	0	1
IXa[A]	H	CH_2COOH	4.14	2.00	2.14	7.31	0	0
IXb	4-Cl-C_6H_4	CH_2COOH	5.00	5.04	–0.04	10.31	0	0
IXc	3,4-$(OCH_3)_2$-C_6H_3	CH_2COOH	5.30	5.41	–0.11	11.06	0	0
IXd	4-Br-C_6H_4	CH_2COOH	5.36	5.24	0.12	10.60	0	0
IXe	4-F-C_6H_4	CH_2COOH	4.65	4.61	0.04	9.84	0	0
Xa	2-Cl-5-CF_3	CH_2COOH	5.23	5.50	–0.27	9.58	1	0
Xb	3,4-$(Cl)_2$	CH_2COOH	5.25	5.48	–0.23	9.56	1	0
XIa	4-Cl	NH	3.70	4.00	–0.30	8.14	1	0
XIb	2-Cl-5-CF_3	NH	4.00	4.53	–0.53	8.65	1	0
XIc	4-Cl	NHCOOH	ND	4.84	ND	8.94	1	0
XIIa	2-Cl-5-CF_3-C_6H_3	O	5.89	5.30	0.59	9.38	1	0
XIIb	3,4-$(Cl)_2$-C_6H_3	O	5.68	5.28	0.40	9.37	1	0
XIIc	C_6H_5	S	5.92	5.54	0.38	8.98	0	1
XIId	4-Cl-C_6H_4	O	5.36	4.77	0.59	8.87	1	0
XIIe	3-Cl-4-OCH_3-C_6H_3	O	5.52	5.40	0.12	9.49	1	0

[a] ND = Not determined. Superscript "A" means not used in the derivation of QSAR 4.

2.2. Inhibition of Anthrax Lethal Factor (LF) by Compounds (VII-XII).[66] (Figure 2, Table 3.) In order to explore the relative importance of various chemical substructures of **VI** in inhibiting the protease activity of LF, Pellecchia and co-workers[66] further designed, synthesized, and tested the inhibitory activities of compounds (**VII**−**XII**) against LF. Authors also presented structure−activity relationship and CoMFA (comparative molecular field analysis). The obtained 3D-QSAR model was subsequently compared with the X-ray structure of the complex between LF and a representative compound. The docked conformations of 17 and 10 compounds were used as training and test sets, respectively, for the CoMFA study. The inhibition constants (log 1/IC_{50}) were correlated with the steric and electrostatic fields, as well as the total molecular surface area (TMSA) of each compound. The cross-validation (leave-one-out) was carried out to obtain the optimal number of components to be used in the final analysis. After the optimal number of components (four) was determined, a non-cross-validated analysis was performed without column filtering. The statistics q^2 (0.51), SPRESS (0.60), r^2 (0.98), and F values (145.94) were obtained. The relative contributions to this

(66) Johnson, S. L.; Jung, D.; Forino, F.; Chen, Y.; Satterthwait, A.; Rozanov, D. V.; Strongin, A. Y.; Pellecchia, M. Anthrax lethal factor protease inhibitors: synthesis, SAR, and structure-based 3D QSAR studies. *J. Med. Chem.* **2006**, *49*, 27–30.

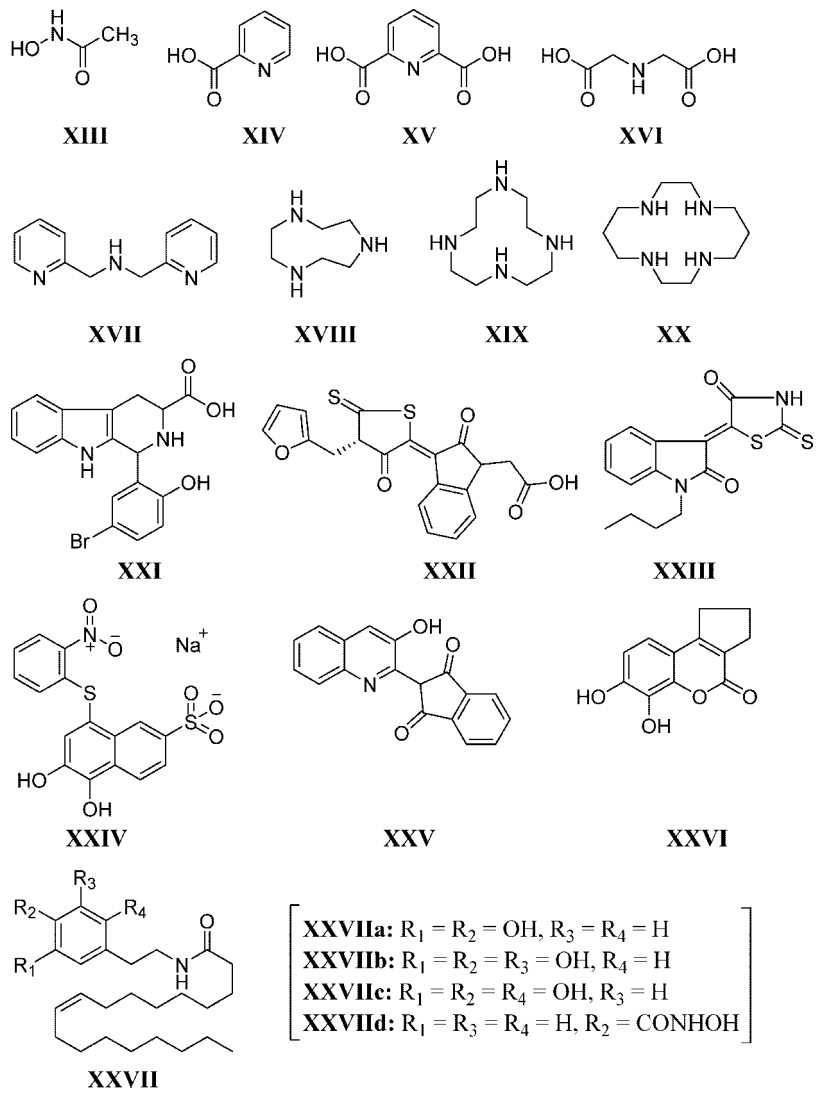


Figure 3. Structure of compounds (**XIII**−**XXVII**) used in the development of QSAR model (eq 5).

CoMFA model were 40.9% for the steric field, 38.5% for electrostatic field, and 20.6% for total molecular surface area. The authors also performed additional CoMFA studies in which 27 compounds were considered. The resulting q^2 value by leaving nine compounds out was 0.54, with similar relative contributions of the steric, electrostatic, and TMSA fields. A comparison of the binding site of LF with the CoMFA contour plots of steric field contribution shows a parallel between favorable steric contours and hydrophobic regions of the protein.

Over the past three decades, a major trend in the evolution of QSAR has been the development of 3D-QSAR, which is viewed as an extension of QSAR and addressed as three-dimensional (3D). The main reason for this trend is an in-depth understanding of protein−ligand interaction or binding to various receptors at the atomic level, supported by a wealth of experimental evidence. The most popular method in 3D-QSAR studies is CoMFA that has been used to develop several QSAR models. However, despite the development of regression equations, the results are best mapped and visualized as 3D color-coded contour plots. Semiquantitative generalizations are drawn from these graphics, but clear-cut quantitative aspects of the models are minimized. Thus, rigorous and direct comparisons or extrapolations cannot be made between CoMFA models for different proteins/receptors, unlike 2D-QSAR. One of the attributes of CoMFA is its ability to combine diverse data sets and examine them as a whole. This operation can also be a weakness since it suggests that all the members of the set bind to a receptor/protein at the "same" binding site and in the "same" mode.[67] The other drawback is the increasing complexity of the models which requires 3D conformations, their alignment and a large number of variables. This can make it more difficult to reproduce a model or at least to apply it to new compounds if the alignment rules are too specific or are not suitable for other chemical classes, limiting the range of chemicals that can be analyzed.[68] In the presence of a sound knowledge of alignment and conformation of the individual

(67) Selassie, C. D.; Garg, R.; Mekapati, S. Mechanism-based QSAR approach to the study of the toxicity of endocrine active substances. *Pure Appl. Chem.* **2003**, *75*, 2363–2373.

molecules, it becomes necessary to examine numerous alignments and conformations to select the best combination that generates a CoMFA model with the highest predictive power (q^2).[69] On the other hand, 2D-QSAR allows for the direct comparison of the descriptors, their contributions, and other statistical terms. It has great utility in mechanistic interpretation as well as predictability within a congener series. It can also pinpoint molecules that behave in an anomalous fashion. A further aspect to be considered is that it can be easily and quickly used by non expert users. Thus, it is justified to develop a 2D-QSAR model and compare them to their 3D-QSAR results. Considering these facts, we derived eq 4 using the data in Table 3.

$$\log 1/IC_{50} = 1.05(\pm 0.26)\,\mathrm{CMR} - 2.83(\pm 1.13)\log(\beta \times 10^{\mathrm{CMR}} + 1) + 1.13(\pm 0.48)I_{\mathrm{furan}} + 1.80(\pm 0.71)I_{\mathrm{thiophene}} - 5.67(\pm 2.67)$$
$$n = 25,\ r^2 = 0.809,\ s = 0.389,\ q^2 = 0.713,\ Q = 2.311,\ F_{5,19} = 16.095(4.171)$$
$$\text{optimum CMR} = 11.10;\ \log\beta = -11.33$$
$$\text{Clog } P \text{ vs CMR};\ r = 0.170$$
$$\text{outliers: } \mathbf{VIIb}, \mathbf{VIIIc}, \mathbf{VIIIe}, \mathbf{IXa}$$
$$\text{range in } \log 1/IC_{50} = 6.72-3.70 \quad (4)$$

This is a bilinear correlation in terms of CMR, which suggests that the activity of these compounds (**VII**−**XII**) first increases linearly with an increase in molar refractivity of the whole molecule up to an optimum CMR of 11.10 and then decreases gradually. Indicator variable (I_{furan}) takes the value of 1 and 0 for the presence and absence of a furan ring, respectively. Similarly, $I_{\mathrm{thiophene}}$ is an indicator variable taking the value of 1 and 0 for the presence and absence of a thiophene ring, respectively. The indicator variables (I_{furan} and $I_{\mathrm{thiophene}}$) with positive coefficients suggest that the presence of either furan or thiophene or both will enhance the activity. In this model, 50.4% of the variance in the data is explained by the steric descriptor (CMR), while the CoMFA model predicted the contribution of 40.9% for the steric field. The difference in the prediction may be due to the number of compounds involved in the development of these two QSAR models. QSAR (eq 4) represents 25 compounds (four outliers; $q^2 = 0.713$), while the CoMFA model represents only 18 compounds (nine outliers and two compounds not considered; $q^2 = 0.540$).

Table 4. Biological and Physicochemical Parameters Used To Derive QSAR 5

	log 1/IC_{50} (eq *5*)					
compd	obsd	pred	Δ	CMR	Clog *P*	ref
XIII	1.94	1.90	0.04	1.66	−1.59	70
XIV	2.46	2.63	−0.17	3.13	1.00	70
XV	2.89	2.95	−0.06	3.78	1.08	70
XVI	2.57	2.45	0.12	2.78	−0.65	70
XVII[a]	2.24	4.08	−1.84	6.07	−0.33	70
XVIII	3.43	3.00	0.43	3.89	−2.23	70
XIX	3.03	3.64	−0.61	5.19	−2.97	70
XX[a]	2.53	4.10	−1.57	6.11	−1.85	70
XXI	4.42	4.92	−0.50	9.48	0.80	71
XXII	4.42	4.78	−0.36	10.78	2.56	71
XXIII	5.00	4.94	0.06	9.14	2.62	71
XXIV	5.23	4.90	0.33	9.64	1.47	71
XXV	5.41	4.90	0.51	8.12	2.93	71
XXVI[a]	5.77	3.89	1.88	5.68	2.23	71
XXVIIa	4.82	4.57	0.25	12.65	7.74	29
XXVIIb	4.89	4.55	0.34	12.80	7.07	29
XXVIIc	4.15	4.55	−0.40	12.80	7.02	29
XXVIId	4.49	4.49	0.00	13.36	7.11	29

[a] Not included in the derivation of QSAR 5.

2.3. Inhibition of Anthrax Lethal Factor (LF) by Compounds XIII−XXVII.[29,70,71] (Figure 3, Table 4). From the anthrax lethal factor (LF) data of compounds (**XIII**−**XXVII**)[29,70,71] in Table 4, we derived a bilinear correlation 5 in terms of CMR. It suggests that the activities of these compounds first increase linearly with an increase in the molar refractivity of the whole molecules up to an optimum CMR value of 8.74 and then decrease gradually. Three compounds [**XVII**, **XX**, and **XXVI**] were deemed to be outliers on the basis of their deviations (obsd − pred > 2*s*). To assess the effects of excluding outliers, QSAR model was examined before and after the removal of compounds.

$$\log 1/IC_{50} = 0.49(\pm 0.13)\,\mathrm{CMR} - 0.61(\pm 0.26)\log(\beta \times 10^{\mathrm{CMR}} + 1) + 1.08(\pm 0.69)$$
$$n = 15,\ r^2 = 0.905,\ s = 0.391,\ q^2 = 0.856,\ Q = 2.432,\ F_{3,11} = 34.930(6.217)$$
$$\text{optimum CMR} = 8.74;\ \log\beta = -8.10;\ \text{outliers: } \mathbf{XVII}, \mathbf{XX}, \text{ and } \mathbf{XXVI}$$
$$\text{range in } \log 1/IC_{50} = 5.77-1.94 \quad (5)$$

There is a high correlation between Clog *P* and CMR ($r = 0.863$). Therefore, Clog *P* can replace CMR, and by doing so in eq 5, we can develop eq 5a. Finally, QSAR (eq 5a) was not considered due to its very low and unacceptable statistics (see below):

$$\log 1/IC_{50} = 0.41(\pm 0.30)\,\text{Clog } P - 0.56(\pm 0.71)\log(\beta \times 10^{\text{Clog}P} + 1) + 3.52(\pm 0.60)$$
$$n = 15,\ r^2 = 0.519,\ s = 0.879,\ q^2 = 0.297,\ Q = 0.819,\ F_{3,11} = 3.956(6.217)$$
$$\text{optimum Clog } P = 4.36;\ \log\beta = -3.94;\ \text{outliers: } \mathbf{XVII}, \mathbf{XX}, \text{ and } \mathbf{XXVI} \quad (5a)$$

Although the data set (Table 4) has some compounds with

(68) Roncaglioni, A.; Benfenati, E. In silico-aided prediction of biological properties of chemicals: oestrogen receptor-mediated effects. *Chem. Soc. Rev.* **2008**, *37*, 441–450.
(69) Hughes, R. A.; Harris, T.; Altmann, E.; McAllister, D.; Vlahos, R.; Robertson, A.; Cushman, M.; Wang, Z.; Stewart, A. G. 2-Methoxyestradiol and analogs as novel antiproliferative agents: Analysis of three-dimensional quantitative structure-activity relationships for DNA synthesis inhibition and estrogen receptor binding. *Mol. Pharmacol.* **2002**, *61*, 1053–1069.
(70) Jacobsen, F. E.; Lewis, J. A.; Cohen, S. M. A new role for old ligands: discerning chelators for zinc metalloproteinases. *J. Am. Chem. Soc.* **2006**, *128*, 3156–3157.
(71) Johnson, S. L.; Chen, L.-H.; Pellecchia, M. A high-throughput screening approach to anthrax lethal factor inhibition. *Bioorg. Chem.* **2007**, *35*, 306–312.

very low activities, the QSAR model (eq 5) is able to explain these activities as follows: (i) Compound **XIII** has the lowest activity (log $1/IC_{50} = 1.94$) due to the lowest CMR value (CMR = 1.66). (ii) According to eq 5, the compound activity increases as the value of CMR increases up to an optimum CMR of 8.74. Thus, compound **XXV** has the highest activity (log $1/IC_{50} = 5.41$) due to the CMR ≈ 8.74 (CMR = 8.12). [Note: compound **XXVI** is an outlier.] (iii) After increasing the CMR value from 8.74 the activity begins to decrease. Thus, compound **XXIV** has lower activity (log $1/IC_{50} = 5.23$) due to the CMR > 8.74 (CMR = 9.64). (iv) Increasing more the value of CMR (CMR ≫ 8.74), the compound activity further decreases. Thus, compound **XXVIIa** has the activity (log $1/IC_{50} = 4.82$) depending on the CMR ≫8.74 (CMR = 12.65). (v) Similarly, compound **XXVIId** with the highest CMR value of 13.36 has the activity (log $1/IC_{50} = 4.49$). The above results are in good agreement with the predicted values obtained from the QSAR model (eq 5).

2.4. Apparent Inhibition Constant (K_i) of Neomycin B and Their Synthetic Analogues (XXVIII) against the Protease Activity of Anthrax Lethal Factor (LF) under the Assay Condition of High Salt Concentration.[72] (Figure 4, Table 5.) Although a number of compounds demonstrated some level of inhibitory activity, neomycin B, a commonly utilized aminoglycoside antibiotic, was found to be the most potent inhibitor of LF with an apparent inhibition constant value in the low nanomolar concentration range ($K_i = 37$ nM) under low-ionic-strength assay conditions. An increase in the ionic strength from 0 to 150 mM KCl drastically shifts the K_i values of all aminoglycosides (**XXVIII**) by a factor of ≈1500−53000 toward higher concentrations (see Table 5 for the data in molar concentration). A possible reason for the observed different sensitivity of different aminoglycosides (**XXVIII**) toward the ionic strength of the solution could be the different number of amino groups and their individual pK_a values. Equation 6 is based on the data (under high salt concentration assay) obtained from Fridman et al.[72] (see Figure 4, Table 5):

$$\log 1/K_i = 0.20(\pm 0.08)\pi_X + 4.71(\pm 0.23)$$
$$n = 11,\ r^2 = 0.795,\ s = 0.288,\ q^2 = 0.673,\ Q = 3.094,\ F_{1,9} = 34.902(10.561)$$
outliers: **XXVIIIj**, **XXVIIIk**; range in $\log 1/K_i = 5.96-3.87$ (6)

π_X is the calculated hydrophobic parameter of X substituents, which is found to be the single most important parameter for this data set. This shows that hydrophobic contacts are made at all positions where X substituents have been entered. The linear π_X model suggests that the molecule with highly hydrophobic X substituent will be more active. The existence of a linear only correlation between activity and hydrophobicity of the X substituents suggests that the π_X values were not great enough to establish the upper limit of the activity, i.e., the balance between π_X values and the compound activity. Thus, more compounds will be needed for this data set to establish the upper limit of the π_X values by the development of either parabolic or bilinear QSAR

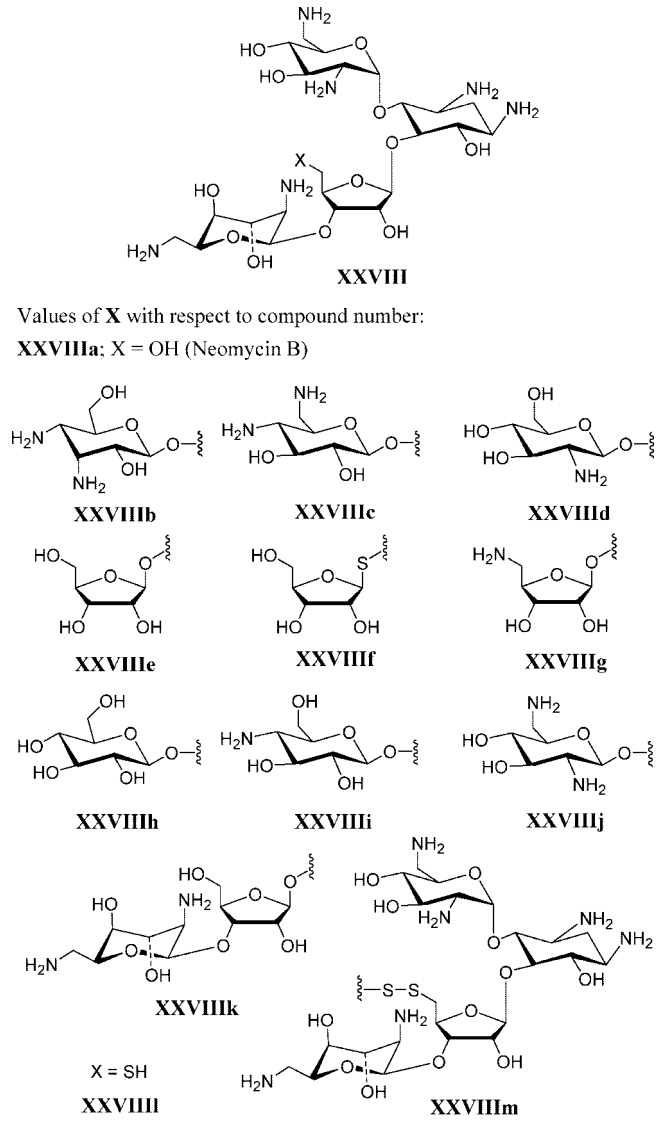


Figure 4. Structure of compounds **XXVIII** used in the development of QSAR model (eq 6).

models. Two compounds (**XXVIIIj** and **XXVIIIk**) were deemed to be outliers on the basis of their deviations (obsd − pred > 2*s*). The result of this QSAR model (eq 6) is not in good agreement with the findings of Fridman et al.,[72] which suggests that the number of amino groups on the ligand is critical for LF-binding affinity, but the structural features of the ligand also play an important role in the proper recognition of LF.

The crystal structures of LF bound to an optimized peptide substrate and to peptide-based inhibitors indicate that the long hydrophobic substrate-binding groove and deep S1′ pocket adjacent to the catalytic Zn^{2+}-binding site seem to be main determinants for strong target affinity.[73] This strong hydro-

(72) Fridman, M.; Belakhov, V.; Lee, L. V.; Liang, F.-S.; Wong, C.-H.; Baasov, T. Dual effect of synthetic aminoglycosides: antibacterial activity against *Bacillus anthracis* and inhibition of anthrax lethal factor. *Angew. Chem., Int. Ed.* **2005**, *44*, 447–452.

Table 5. Biological and Physicochemical Parameters Used To Derive QSAR 6

	log $1/K_i$ (eq *6*)			
compd	obsd	pred	Δ	π_X
XXVIIIa	4.23	4.56	−0.33	−0.75
XXVIIIb	4.30	3.96	0.34	−3.69
XXVIIIc	4.55	4.13	0.42	−2.86
XXVIIId	4.18	4.12	0.06	−2.92
XXVIIIe	3.87	4.27	−0.40	−2.16
XXVIIIf	4.09	4.34	−0.25	−1.83
XXVIIIg	4.41	4.29	0.12	−2.10
XXVIIIh	3.90	4.11	−0.21	−2.95
XXVIIIi	4.07	4.12	−0.05	−2.93
XXVIIIj[a]	4.70	4.13	0.57	−2.86
XXVIIIk[a]	4.68	3.85	0.83	−4.24
XXVIIIl	5.00	4.78	0.22	0.33
XXVIIIm	5.96	5.88	0.08	5.71

[a] Not included in the derivation of QSAR 6.

phobic selectivity has also been supported by the experimental data obtained from non-peptide small molecules.[74] This is further supported by our QSAR (eq 6). Thus, a large hydrophobic X substituent would probably increase the binding affinity of an aminoglycoside (**XXVIII**) for the LF active site.

2.5. Apparent Inhibition Constant (K_i) of Guanidinylated 2,5-Dideoxystreptamine Derivatives (XXX) against Anthrax Lethal Factor (LF)..[75] (Figure 5, Table 6.) It has already been identified that neomycin B (**XXVIIIa**) and other aminoglycosides are able to block LF protease activity. However, the structural complexity of neomycin B renders itself an unattractive lead structure in the drug discovery. An attention was taken to a structurally simpler neamine (**XXIX**) which is the core structure of neomycin B. It was found that neamine acted as a weak inhibitor of LF, but its potency has been enhanced dramatically by the addition of guanidinyl groups. Although guanidinylated neamine analogues showed promising inhibitory activity against LF, the use of these compounds as antianthrax drugs seems to be an unattainable goal due to the poor oral bioavailability and potential toxicity associated with aminoglycosides.[75,76] It has

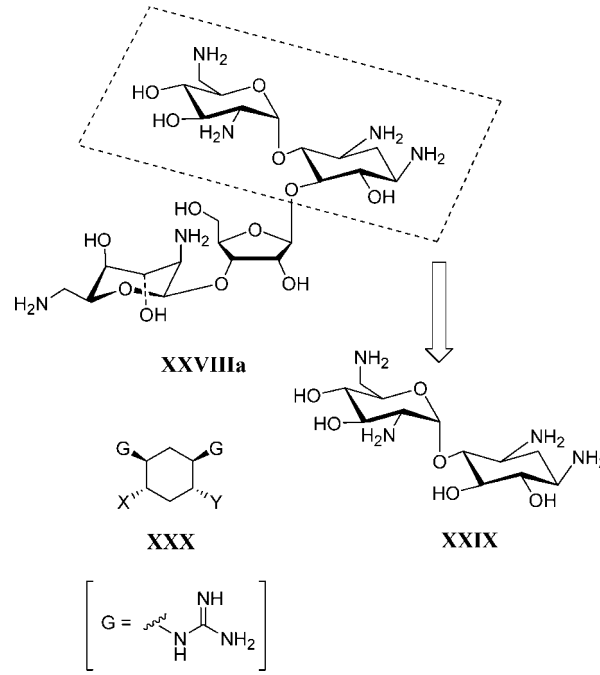


Figure 5. Structure of neomycin B (**XXVIIIa**), neamine (**XXIX**), and guanidinylated 2,5-dideoxystreptamine derivatives (**XXX**). [Compounds **XXX** were used in the development of QSAR model (eq 7).]

been observed that a planar aryl group could function as a surrogate of the pyranose ring in neamine mimetics designed for RNA recognition.[77,78] By considering the above facts, Jiao et al.[75] synthesized a series of guanidinylated 2,5-dideoxystreptamine derivatives (**XXX**) and evaluated their apparent inhibition constant (K_i) values against LF. These compounds have the same unique spatial arrangement of amino groups as well as increased lipophilicity due to the presence of planar non-sugar groups resulting in enhanced bioavailability. From the data in Table 6, we derived eq 7.

$$\log 1/K_i = -0.31(\pm 0.25)\ \text{Clog}\ P + 0.65(\pm 0.26)\ \text{NOG} + 2.03(\pm 0.72)$$
$$n = 9,\ r^2 = 0.953,\ s = 0.256,\ q^2 = 0.915,\ Q = 3.813,\ F_{2,6} = 60.830(10.925)$$
outliers: **XXXg**, **XXXj**
Clog P vs NOG; $r = 0.627$
range in $\log 1/K_i = 7.19-3.81$ (7)

NOG is the number of guanidinyl groups, which is found to be the most important parameter and accounts for 87% of the variance in the data. Its positive coefficient suggests

(73) Turk, B. E.; Wong, T. Y.; Schwarzenbacher, R.; Jarrell, E. T.; Leppla, S. H.; Collier, R. J.; Liddington, R. C.; Cantley, L. C. The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. *Nat. Struct. Mol. Biol.* **2004**, *11*, 60–66.

(74) Panchal, R. G.; Hermone, A. R.; Nguyen, T. L.; Wong, T. Y.; Schwarzenbacher, R.; Schmidt, J.; Lane, D.; McGrath, C.; Turk, B. E.; Burnett, J.; Aman, M. J.; Little, S.; Sausville, E. A.; Zaharevitz, D. W.; Cantley, L. C.; Liddington, R. C.; Gussio, R.; Bavari, S. Identification of small molecule inhibitors of anthrax lethal factor. *Nat. Struct. Mol. Biol.* **2004**, *11*, 67–72.

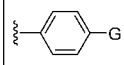
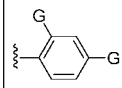
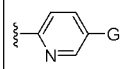
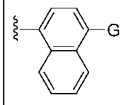
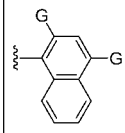
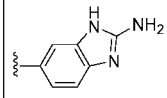
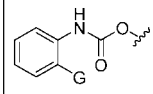
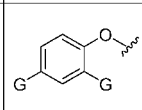
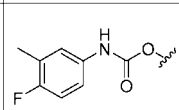
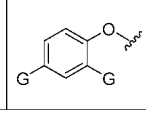
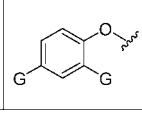
(75) Jiao, G.-S.; Cregar, L.; Goldman, M. E.; Millis, S. Z.; Tang, C. Guanidinylated 2,5-dideoxystreptamine derivatives as anthrax lethal factor inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1527–1531.

(76) Luedtke, N. W.; Carmichael, P.; Tor, Y. Cellular uptake of aminoglycosides, guanidinoglycosides, and poly-arginine. *J. Am. Chem. Soc.* **2003**, *125*, 12374–12375.

(77) Ding, Y.; Hofstadler, S. A.; Swayze, E. E.; Griffey, R. H. An efficient synthesis of mimetics of neamine for RNA recognition. *Org. Lett.* **2001**, *3*, 1621–1623.

(78) Vourloumis, D.; Takahashi, M.; Winters, G. C.; Simonsen, K. B.; Ayida, B. K.; Barluenga, S.; Qamar, S.; Shandrick, S.; Zhao, Q.; Hermann, T. Novel 2,5-dideoxystreptamine derivatives targeting the ribosomal decoding site RNA. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3367–3372.

Table 6. Biological, Physicochemical, and Structural Parameters Used To Derive QSAR 7[a]

compd	X	Y	log $1/K_i$ (eq 7)			Clog *P*	NOG
			obsd	pred	Δ		
XXXa		OH	4.83	4.83	0.00	–2.46	3
XXXb		OH	4.51	5.10	–0.59	–3.27	3
XXXc		OH	6.22	6.32	–0.10	–4.90	4
XXXd		OH	5.18	5.00	0.18	–2.98	3
XXXe		OH	5.39	5.28	0.11	–3.79	3
XXXf		OH	4.50	4.43	0.07	–1.28	3
XXXg[A]		OH	4.97	5.92	–0.95	–3.72	4
XXXh		OH	3.81	3.91	–0.10	–1.74	2
XXXi		OH	4.61	4.77	–0.16	–2.29	3
XXXj[A]			6.30	4.82	1.48	–0.48	4
XXXk			7.19	7.38	–0.19	–4.10	6

[a] NOG = Number of guanidinyl groups. Superscript "A" means not included in the derivation of QSAR 7.

that increasing the number of guanidinyl groups would probably show strong inhibitory activity against LF. There is no positive role for hydrophobic properties (and in fact the negative Clog *P* term has marginal significance). Two compounds (**XXXg** and **XXXj**) were deemed to be outliers on the basis of their deviations (obsd − pred > 2*s*).

The X-ray crystal structure of anthrax lethal factor provides a clue for interpreting this QSAR model (eq 7). The active site of the protease has been found to be a deep, broad, 40 Å groove with a region of highly negative electrostatic potential due to the presence of acidic Asp and Glu residues.[57,73,74,79] Once the 2,5-dideoxystreptamine derivatives (**XXX**) occupy the active site, multiple electrostatic and H-bonding interactions may occur between the positively charged guanidinyl groups of **XXX** and the negatively charged residues of LF. Therefore, the guanidinyl groups, which might mimic the positively charged arginine- and lysine-rich section of LF substrates such as MAPKKs, have

Table 7. Comparison of the Statistics Obtained from the Multiregression Analysis (MRA) Process for QSARs 1−7

			descriptor coefficient							
QSAR no.	compd	*n*	hydrophobic	steric/pol.	indicator variable	r^2	q^2	*s*	*Q*	F^a	intercept
1	**I**−**V**	14	−1.03 Clog *P*	2.69 MgVol		0.951	0.929	0.405	2.407	106.745(7.206)	0.90
2	**I**−**V**	13	0.44 Clog *P*	0.65 MgVol		0.821	0.706	0.362	2.503	22.933(7.559)	3.33
3	**VI**	25	$-1.21\pi_Y + 0.28\pi_Y^2$		$1.17I$	0.732	0.658	0.434	1.972	19.119(4.874)	5.39
4	**VII**−**XII**	25		1.05 CMR − 2.83 $\log(\beta \times 10^{CMR} + 1)$	$1.13I_{furan} + 1.80I_{thiophene}$	0.809	0.713	0.389	2.311	16.095(4.171)	5.67
5	**XIII**−**XXVII**	15		0.49 CMR − 0.61 $\log(\beta \times 10^{CMR} + 1)$		0.905	0.856	0.391	2.432	34.930(6.217)	1.08
6	**XXVIII**	11	$0.20\pi_X$			0.795	0.673	0.288	3.094	34.902(10.561)	4.71
7	**XXX**	9	−0.31 Clog *P*		0.65 NOG	0.953	0.915	0.256	3.813	60.830(10.925)	2.03

[a] The figure within parentheses refers to the *F*-value at the 99% level.

Table 8. Y-Randomization Data for QSARs 1−7

	NOR-1[a]		NOR-2		NOR-3		NOR-4		NOR-5	
QSAR no.	r^2	q^2	r^2	q^2	r^2	q^2	r^2	q^2	r^2	q^2
1	0.484	0.200	0.482	0.174	0.480	0.150	0.484	0.143	0.509	0.091
2	0.218	−0.153	0.400	0.046	0.311	−0.108	0.220	−0.066	0.332	−0.187
3	0.380	−0.130	0.030	−0.359	0.118	−0.106	0.104	−0.434	0.152	−0.464
4	0.190	−0.174	0.182	−0.347	0.138	−0.469	0.456	0.013	0.236	−0.167
5	0.465	0.265	0.448	0.234	0.444	0.194	0.408	0.142	0.422	0.146
6	0.229	−2.434	0.016	−0.372	0.059	−0.294	0.125	−0.104	0.059	−0.826
7	0.192	−0.764	0.014	−1.185	0.031	−3.131	0.120	−0.810	0.014	−1.649

[a] NOR = number of Y-randomization.

a primary effect on the inhibitory activity of these molecules (**XXX**) against LF. Thus, it has been suggested that the higher the number of guanidinyl groups in 2,5-dideoxystreptamine derivatives (**XXX**), the stronger the interactions, leading to the better inhibition of LF protease.[75]

Validation of QSAR Models. QSAR model validation becomes an essential part to understand statistically robust models capable of making accurate and reliable predictions of biological activities of new compounds not present in the data set. A comparison of the statistics of QSARs 1−7 obtained from multiregression analyses (MRA) is shown in Table 7. The following approaches have been used to validate QSAR models (eqs 1–7):

Internal Validation: (1) Fraction of the Variance (r^2). It is believed that the closer the value of r^2 to unity, the better the QSAR model. The values of r^2 for these seven QSAR models (eqs 1–7) range from 0.732 to 0.953, which suggests that these QSAR models explain 73.2−95.3% of the variance in the data. According to the literature, the predictive QSAR model must have $r^2 > 0.6$.[80,81]

(2) Cross-Validation Test. The cross-validated r^2 [q^2 (LOO-q^2)] values for these QSAR models range from 0.658 to 0.929. High values of q^2 validate the QSAR models. According to the literature, the predictive QSAR model must have $q^2 > 0.5$.[80,81]

(3) Standard Deviation (*s*). The smaller the value of *s*, the better the QSAR model. The values of *s* for these QSAR models range from 0.256 to 0.434.

(4) Quality Factor (*Q*). High values of *Q* (1.972−3.813) for these QSAR models suggest their high predictive power.

(5) Fischer Statistics (*F*). The larger the value of *F*, the greater the probability that the QSAR model is significant. The *F*-values for these QSAR models range from 16.095 to 106.745, which are statistically significant at the 99% level.

(6) Y-Randomization Test. In this test, the dependent-variable vector (**Y** vector) is randomly shuffled, and a new QSAR model is developed using the original independent variable matrix. The process is repeated several times. It is expected that the resulting QSAR models should have low r^2 and low q^2 values. This is a widely used technique to ensure the robustness of a QSAR model. The statistical data of r^2 and q^2 for five runs are listed in Table 8 (eqs 1–7). The poor values of r^2 and q^2 in the Y-randomization test ensure the robustness of these QSAR models.[47,51,61,81–83]

(7) Lack of Overfitting. A model overfits if it includes more descriptors than required. The lack of overfitting for

(79) Pannifer, A. D.; Wong, T. Y.; Schwarzenbacher, R.; Renatus, M.; Petosa, C.; Blenkowska, J.; Lacy, D. B.; Collier, R. J.; Park, S.; Leppla, S. H.; Hanna, P.; Liddington, R. C. Crystal structure of the anthrax lethal factor. *Nature* **2001**, *414*, 229–233.

(80) Golbraikh, A.; Tropsha, A. Beware of q^2! *J. Mol. Graphics Modell.* **2002**, *20*, 269–276.

(81) Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.

(82) Wold, S.; Eriksson, L. In *Chemometrics Methods in Molecular Design: Statistical Validation of QSAR Results*; van de Waterbeemd, H., Ed.; VCH: Weinheim, Germany, 1995; pp 309−318.

(83) Melzig, M. F.; Tran, G. D.; Henke, K.; Selassie, C. D.; Verma, R. P. Inhibition of neutrophil elastase and thrombin activity by caffeic acid esters. *Pharmazie* **2005**, *60*, 869–873.

all these QSAR models (eqs 1–7) was confirmed by using the following conditions: (i) Number of data points/number of descriptors ≥4. (ii) High values of Q. (iii) All the QSAR models (eqs 1–7) were checked for their correlation with fewer descriptors than that of the original. (iv) Y-randomization test (Table 8) suggests that the high r^2 values of the QSAR models (eqs 1–7) are not due to chance correlation or overfitting.[61,84]

External Validation: Selection of the Training and Test Sets. The original data set of QSAR model (eq 1) was divided into training [n = 10 (∼75%)] and test [n = 4 (∼25%)] sets in a random manner. The QSAR model for this training set was generated by using the same descriptors as those of eq 1 and validated on the basis of their statistics (acceptance criteria: $r^2 > 0.6$ and $q^2 > 0.5$). The predictive capacity of this model is judged from their predictive R^2 (R^2_{pred}) value, which was calculated by eq 8:

$$R^2_{\text{pred}} = 1 - \frac{\sum (Y_{\text{pred(test)}} - Y_{\text{test}})^2}{\sum (Y_{\text{test}} - \bar{Y}_{\text{training}})^2} \qquad (8)$$

In eq 8, $Y_{\text{pred(test)}}$ and Y_{test} are the respective predicted and observed activities of the test set compounds and $\bar{Y}_{\text{training}}$ is the observed mean activity of the training set compounds. Similarly, the QSAR models (eqs 2–7) were also divided into training and test sets for their external validation. A random selection pattern of the test set and the statistical parameters of the training set for all these QSAR models (eqs 1–7) are given in Table 9.

Overview. An analysis of our QSAR models (eqs 1–7) reveals a number of interesting points. The most important of these is the hydrophobic parameter of the compound/substituent. Out of 7 QSAR, 5 contain a correlation between activity and hydrophobicity of the compound/substituent. A positive linear correlation is found in two equations (eqs 2 and 6). The coefficient associated with these hydrophobic parameters ranges from 0.44 (eq 2) to 0.20 (eq 6). These data suggest that activity might be improved by increasing compound/substituent hydrophobicity. A negative linear correlation is found in two equations (eqs 1 and 7), and the coefficients are −1.03 and −0.31. Less hydrophobic congeners of these compounds family might display enhanced activity. Parabolic correlation with hydrophobic parameter of the substituents is found in one QSAR (eq 3), which reflects the situation where activity declines with increasing hydrophobicity of the substituents and then changes direction and increases. This may correspond to an allosteric reaction.

Table 9. Random Selection Pattern of the Test Set and Statistical Parameters of the QSAR for Their Respective Training Set Obtained from the Division of the Original Set of the QSAR Models (Eqs *1–7*)

		statistical parameters[a]				
QSAR no.	test sets	n	r^2	q^2	s	R^2_{pred}
1	**I**, **IIIb**, **Vb**, **Ve**	10	0.940	0.888	0.465	0.966
2	**I**, **II**, **IVa**	10	0.773	0.532	0.189	0.641
3	**VIa**, **VIb**, **VIf**, **VIp**, **VIx**, **VIz**	19	0.635	0.508	0.396	0.748
4	**VIIa**, **VIIf**, **VIIh**, **VIIj**, **XIa**, **XIId**	19	0.765	0.549	0.390	0.861
5	**XIV**, **XXIII, XXV**, **XXVIId**	11	0.900	0.800	0.409	0.824
6	**XXVIIIf**, **XXVIIIg**, **XXVIIIi**	8	0.803	0.657	0.332	0.802
7	**XXXb**, **XXXk**	7	0.973	0.882	0.154	0.927

[a] Statistical parameters of the QSARs for the training sets (the compounds remaining after removal of the test set compounds from the original data set).

The second important parameter is the steric parameter (molar volume and molar refractivity), which is present in four QSAR. A positive linear correlation with MgVol is found in two equations (eqs 1 and 2) with coefficient 2.69 and 0.65, which suggests that the activity might be improved by increasing molar volume of the molecules. A bilinear correlation with molar refractivity of the compounds is also present in two equations (eqs 4 and 5), which suggests that the activity is optimal for a particular value of CMR. The optimal values of CMR for these two equations are 11.10 and 8.74, respectively.

Conclusion

Anthrax is an acute infectious disease with a high mortality rate caused by *Bacillus anthracis*, reinforcing the need for better adjunctive therapy and prevention strategies. The QSAR paradigm may be helpful in the design and development of novel antianthrax agents. Our QSAR results on penicillin-based inhibitors of the class A and B β-lactamases from *B. anthracis* and inhibitors of anthrax lethal factor suggest that the hydrophobic and steric factors are the two most important determinants of the activity. Further development of QSAR studies should not only enlarge the areas of their application but also increase our understanding toward the mechanisms of chemical−biological interactions.

MP8000149

(84) Zhang, S.; Golbraikh, A.; Oloff, S.; Kohn, H.; Tropsha, A. A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. *J. Chem. Inf. Model.* **2006**, *46*, 1984–1995.