Article

A Fragment-Based Approach for the *in Silico* Prediction of Blood-Brain Barrier Permeation

Tiago L. Moda, Alexandre E. Carrara and Adriano D. Andricopulo*

Instituto de Física de São Carlos, Universidade de São Paulo, 13560-970 São Carlos-SP, Brazil

A permeabilidade da barreira hematoencefálica (BBB, do inglês *blood-brain barrier*) é uma propriedade fundamental no planejamento de fármacos que atuam no sistema nervoso central (CNS) no tratamento de doenças como a epilepsia, depressão, mal de Alzheimer, mal de Parkinson, esquizofrenia, entre outras. No presente trabalho, estudos das relações quantitativas entre a estrutura e propriedade (QSPR) foram conduzidos para o desenvolvimento e validação de modelos *in silico* para a predição da permeabilidade da BBB. O conjunto de dados utilizado possui significativa diversidade química e ampla distribuição dos valores da propriedade alvo. Os modelos de QSPR gerados apresentaram bons parâmetros estatísticos e foram empregados com sucesso na predição de um conjunto teste de 48 compostos. Os modelos desenvolvidos são úteis na identificação, seleção e planejamento de candidatos a novos fármacos com propriedades farmacocinéticas otimizadas.

Blood-brain barrier (BBB) permeation is an essential property for drugs that act in the central nervous system (CNS) for the treatment of human diseases, such as epilepsy, depression, Alzheimer's disease, Parkinson disease, schizophrenia, among others. In the present work, quantitative structure-property relationship (QSPR) studies were conducted for the development and validation of *in silico* models for the prediction of BBB permeation. The data set used has substantial chemical diversity and a relatively wide distribution of property values. The generated QSPR models showed good statistical parameters and were successfully employed for the prediction of a test set containing 48 compounds. The predictive models presented herein are useful in the identification, selection and design of new drug candidates having improved pharmacokinetic properties.

Keywords: ADME, central nervous system, pharmacokinetics, QSPR, blood-brain barrier

Introduction

The challenges facing the pharmaceutical industry are tremendous at every step of the drug discovery and development process. Technology-based discovery certainly is one of the most important elements to increase research and development (R&D) productivity. The new paradigm of drug discovery involves a combination of classical and modern technologies with innovative strategies addressed to the design of new chemical entities (NCEs) with improved properties.[1-3] NCEs expected to advance into clinical trials should have a good balance of pharmacodynamic and pharmacokinetic properties. For the past decades, problems with absorption, distribution, metabolism and excretion (ADME) have been one of the major reasons for the failure of attractive compounds in advanced stages of drug development.[4-6]

Traditionally, *in vivo* and *in vitro* models are employed in the pharmaceutical industry for the evaluation of pharmacokinetic parameters. However, animal models and cell-based assays are typically time consuming and expensive, and thus not applicable to the early screening of large libraries of compounds.[7-9] In recent years, the appearance and consolidation of *in silico* ADME models have provided useful tools for a faster, simpler, and more cost-effective evaluation of pharmacokinetic properties.[10-12] Quantitative structure-activity and structure-property relationships (QSAR/QSPR) are powerful technologies that correlate descriptors based on molecular structures and use computational algorithms to relate the key descriptors to relevant ADME properties.[10,13-15]

Drugs to treat human central nervous system (CNS) diseases and disorders, such as epilepsy, Alzheimer's disease, Parkinson disease, schizophrenia, depression and brain tumors are required to cross the blood-brain barrier (BBB) by passive diffusion or through the help of

*e-mail: aandrico@ifsc.usp.br

transporters. In contrast, drugs that do not target the CNS should present limited capacity to cross the BBB in order to avoid drug-induced side effects in the brain.[16] Delivering drugs into the brain is a complex process that depends on multiple factors, such as log*P*, hydrogen-bond acceptors and donors, molecular weight, polar surface area and other molecular properties.[17] The tight junctions between the endothelial cells of the brain's capillaries make it almost impossible for anything to get into the brain around the cells. In addition, efflux pumps such as P-glycoprotein (P-gp) and the multidrug resistance-associated protein family (MRP) significantly hinder permeation across the BBB turning chemical compounds back to the way of blood.[16,18]

The most commonly used experimental approaches to predict membrane permeation are the octanol/water partition coefficient, high performance liquid chromatography (HPLC)-related techniques, and the *in vitro* approaches, such as parallel artificial membrane permeation (PAMPA) assays and MDCKII-MDR1 cell line.[19-21] Nonetheless, the complexity, costs, resources and time involved in these assays have increased the importance of *in silico* approaches to predict BBB permeability of lead compounds that selectively target the CNS.[7,8,22,23] In the present work, robust QSPR models were developed for the consensus prediction of BBB permeation using the fragment-based hologram QSAR (HQSAR) approach.[10,14] To the best of our knowledge, the majority of the models reported in the literature are associated with qualitative data (cross/not cross BBB) that offers imprecise values of BBB permeability, thus, the quantitative nature of the models generated in this work is of considerable importance in medicinal chemistry and drug design.

Experimental

Data set

The relative affinity for the blood or brain tissue can be expressed in terms of the blood-brain partition coefficient, log(C brain/C blood), where C brain and C blood are the equilibrium concentrations of the drug in the brain and the blood, respectively (also known as logBB). A data set of 255 structurally diverse molecules with known logBB was collected from literature and the PK/DB - database for pharmacokinetic properties (http://www.pkdb.ifsc.usp.br).[24-48] The data consist of *in vivo* measurements in rats of the compound's partition coefficient between the brain and blood. Compounds containing one asymmetric (chiral) center, for which the corresponding BBB permeation was determined for the racemate, were considered as the individual enantiomers and modeled accordingly, as previously described.[10,14] The list of compounds along with the corresponding logBB data is shown in Table S1 in Supplementary Information (SI) section. This structurally diverse (Figure 1) data set consists of several important therapeutic classes, including anxiolytics (e.g., alprazolam), anti-ulcers (e.g., cimetidine), analgesics (e.g., acetylsalicylic acid), sedatives (e.g., diazepam, flunitrazepam), anti-inflammatories (e.g., ibuprofen and indometacin), antivirals (e.g., nevirapine, zidovudine, indinavir), antihypertensives (e.g., verapamil and clonidine), antihistamines (e.g., mepyramine), antidepressants (e.g., mianserin), and so on.

The 3D structures of the molecules employed in this work were constructed using CONCORD and standard
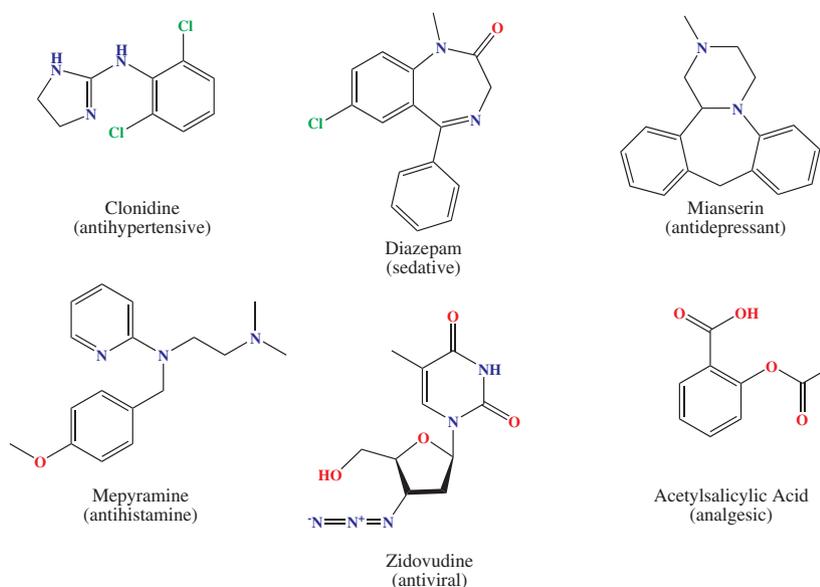
Figure 1. Chemical structures and therapeutic classes of representative drugs included in the data set.

geometric parameters available in the Sybyl 8.0 molecular modeling package (Tripos, St. Louis, USA) and stored as SDF files.[49] The optimization process of the chemical structures was performed by carrying out several standard operations present in ChemAxon Standardizer including 3D depiction layout, hydrogen addition and correction, salt and solvent removal, chirality and bond type normalization and harmonization of the representation of aromatic rings, and others.[50] Each molecule in the set was energetically minimized using the Tripos force field.[48]

In this study, the original data set of 255 compounds was arranged in training (**001**-**207**) and test sets (**208**-**255**) in Table S1 (SI section) to give approximately 80% and 20% of the data set, respectively. The structurally diverse molecules having a significant coverage of property values were included in both sets, as depicted in Figure 2. Thus, the data set is suitable for QSPR model development. The training set was then used to generate the models, while

the test set was hold out for the process of model external validation.

QSPR studies

All 2D QSPR (HQSAR) calculations and analyses were performed using the Sybyl 8.0 package,[48] as previously described. The HQSAR technique employed in this work required the 2D structures and the property value (logBB) as input. Initially, the calculations of the several parameters for the generation of the molecular holograms were performed using the standard parameters implemented in Sybyl 8.0.[10,14,51] All generated models were investigated using full cross-validated r^2 (q^2) partial least squares (PLS) leave-one-out (LOO) and leave-many-out (LMO) methods. The predictive ability of the models was assessed by their q^2 values.

Results and Discussion

HQSAR analyses

The generation of the molecular fragments for the data set compounds was carried out using the following fragment distinctions: atoms (A), bonds (B), connections(C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). In order to assess the process of hologram generation and to seek the best predictive models, several combinations of these parameters were considered using the fragment size default (4-7) (Table 1). The Ch descriptor was considered in all fragment combinations due to the presence of several (R) and (S) enantiomers. The absence of this descriptor could lead to an over-training of the models because two different compounds would be considered as one and treated as such (*i.e.*, calculated twice). The HQSAR analysis
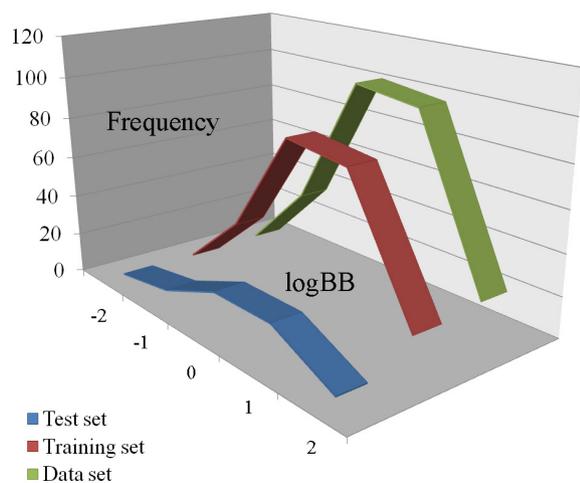
Figure 2. Data set, training set (modeling set) and test set (validation set) distribution.

Table 1. Results of HQSAR analyses for various fragment distinctions on the key statistical parameters using fragment size default (4-7)

Model	Fragment distinction	Statistical parameters				
		q^2	r^2	SEE	N	HL
1	A/B/C/Ch	0.61	0.81	0.33	8	199
2	A/Ch/DA	0.58	0.79	0.35	8	401
3*	A/B/C/H/Ch	0.66	0.87	0.27	8	353
4	A/C/Ch/DA	0.55	0.86	0.29	8	353
5*	A/H/Ch/DA	0.66	0.86	0.29	8	353
6	A/B/C/Ch/DA	0.53	0.86	0.29	7	353
7*	A/B/H/Ch/DA	0.68	0.88	0.26	8	307
8*	A/B/C/H/Ch/DA	0.69	0.91	0.22	8	401

*The four best HQSAR models; q^2, cross-validated correlation coefficient (LOO); r^2, noncross-validated correlation coefficient; SEE, noncross-validated standard error; N, optimal number of components; HL, hologram length.

was performed over the twelve default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins. The patterns of fragment counts from the training set compounds were then related to the measured experimental BBB permeation data.

The best statistical results (predefined accuracy thresholds for training $r^2 \geq 0.80$ and $q^2 \geq 0.60$) were obtained using the fragment distinctions A/B/C/H/Ch (model 3, $q^2 = 0.66$ and $r^2 = 0.87$), A/H/Ch/DA (model 5, $q^2 = 0.66$ and $r^2 = 0.86$), A/B/H/Ch/DA (model 7, $q^2 = 0.68$ and $r^2 = 0.88$) and A/B/C/H/Ch/DA (model 8, $q^2 = 0.69$ and $r^2 = 0.91$).

The influence of different fragment sizes in the statistical parameters was further investigated for the four best HQSAR models marked with asterisk in Table 1 (models 3, 5, 7 and 8), and the results are summarized in Table 2. Fragment size parameters control the minimum and maximum length of fragments to be included in the hologram fingerprint. These parameters represent a fundamental aspect to this fragment-based approach, and should be considered to provide larger or smaller fragments into the molecular holograms.[10,14,51]

The results show that the variation of the fragment size did provide a considerable improvement for the majority of the models (marked with asterisk in Table 2) when compared to the results obtained using the fragment size default (4-7). The exception was model 5, for which no improvement was observed using a set of different fragment sizes. It is worth noting that model 3 exhibited improved cross-validated correlation coefficients (q^2 of 0.68 and 0.71). In the case of model 7, the q^2 value increased from 0.68 to 0.71. For model 8, the q^2 value varied slightly from 0.69 to 0.70, whereas the r^2 remained unchanged.

As the molecular structure encoded within a 2D hologram is directly related to the property value of the training set molecules, the HQSAR model should be able to predict the logBB for new compounds from its fingerprint. The q^2 LOO procedure used may give a suitable representation of the internal consistency and predictive power of the models. However, the real predictive ability of the HQSAR model derived with the 207 training set molecules was assessed by predicting logBB values of an external test set of 48 molecules (compounds **208**-**255**, Table S1 (SI section)). Prior to prediction, the test set compounds were processed identically to the training set compounds as previously indicated. The external validation process can be considered the most valuable validation method as these compounds were completely excluded during the training of the model. The results are listed in Table S1 and show that the test set compounds, which represent the different structural features incorporated with in the training set, are reasonably well predicted by

Table 2. HQSAR analysis for the influence of various fragment sizes on the key statistical parameters using four selected fragment distinctions: A/B/C/H/Ch, A/H/Ch/DA, A/B/H/Ch/DA and A/B/C/H/Ch/DA

Fragment size	Statistical parameters				
	q^2	r^2	SEE	N	HL
Model 3					
2-5	0.68	0.85	0.30	8	307
3-6*	0.71	0.87	0.28	8	401
4-7	0.66	0.87	0.27	8	353
5-8	0.60	0.81	0.33	7	307
6-9	0.53	0.80	0.34	7	307
Model 5					
2-5	0.64	0.78	0.36	8	71
3-6	0.65	0.84	0.30	8	353
4-7*	0.66	0.86	0.29	8	353
5-8	0.64	0.86	0.29	8	353
6-9	0.62	0.86	0.29	8	257
Model 7					
2-5*	0.71	0.85	0.29	8	199
3-6	0.68	0.86	0.28	8	353
4-7	0.68	0.88	0.26	8	307
5-8	0.68	0.87	0.27	8	307
6-9	0.61	0.84	0.31	7	307
Model 8					
2-5	0.68	0.89	0.25	8	307
3-6*	0.70	0.91	0.23	8	401
4-7	0.69	0.91	0.22	8	401
5-8	0.64	0.85	0.29	6	401
6-9	0.60	0.85	0.29	6	353

*The four selected HQSAR models; q^2, cross-validated correlation coefficient (LOO); r^2, noncross-validated correlation coefficient; SEE, noncross-validated standard error; N, optimal number of components; HL, hologram length.

the four selected HQSAR models (marked with asterisk in Table 2). The good agreement between experimental and predicted BBB permeation values indicates the robustness of the HQSAR models.

The predictive power of the models 3, 5, 7, 8 and consensus (r^2_{pred}) are also showed in Table S1. As can be seen, model 7 exhibited higher predictive ability ($r^2_{pred} = 0.79$) than that of models 3, 5 and 8 ($r^2_{pred} = 0.72$, $r^2_{pred} = 0.69$ and $r^2_{pred} = 0.62$, respectively). The consensus approach exhibited an r^2_{pred} of 0.75. Thus, the results indicated that models 3, 7 and consensus could provide better predictions of the property value for new compounds. The graphic representation of the experimental *versus* predicted BBB permeation for both training (model

generation) and test (external evaluation) sets for model 7 is displayed in Figure 3. Similar graphic results were obtained for models 3 and consensus (not shown).
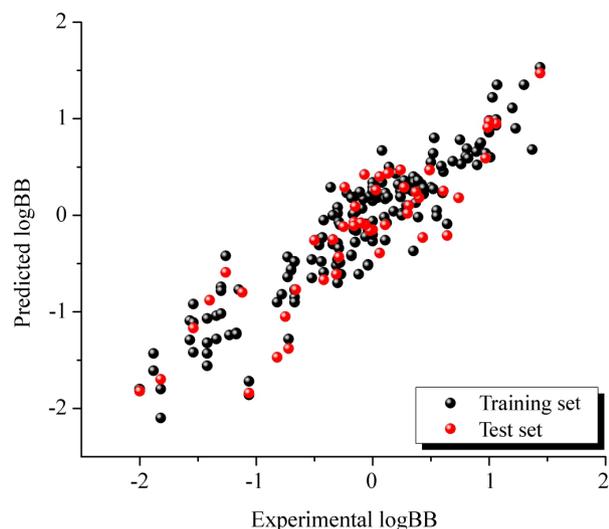
Figure 3. Plot of experimental *versus* predicted BBB permeation for both training and test sets.

The models were successfully validated as shown in Table S1 and Figure 3, especially taking into account the complexity of the BBB biological system. Despite not having the highest r^2_{pred}, the consensus approach is also an attractive tool for the prediction of logBB, considering that the ensemble of models would allow a greater coverage of the chemical space, which, in turn, could be useful for the selection and design of new compounds with improved logBB properties. Additionally, the HQSAR technique can provide predictions for a broad scope of molecules when compared to other methods, considering that the molecular fragmentation offers a much larger range of different scaffold possibilities.

Conclusions

A key challenge in the development of drugs that act in the CNS for the treatment of a variety of human diseases and disorders is their transport across the BBB. The final HQSAR models described here possesses high internal and external consistency. In addition, the quantitative models showed good predictive power and could potentially be used to assist the processes of chemical library design and virtual screening. Compound libraries usually possess a broad chemical diversity, and therefore, *in silico* ADME models that are needed to screen these libraries should inevitably be able to cover a substantial portion of the chemical space. This is hard to be achieved by training the model with few hundreds of compounds. It should be

noted, however, that this limitation may be overcome by the application of similarity analyses in the way of selecting appropriated compounds for screening, thus, avoid making predictions for compounds that differ substantially from the training set molecules.[12] What is clear at this point is that the predictive models generated in this work are useful in the processes of early compound identification and selection, as well as in the design of lead compounds with improved BBB permeability.

Supplementary Information

Supplementary data are available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

References

1. Postigo, M. P.; Guido, R. V. C.; Oliva, G.; Castilho, M. S.; Pitta, I. R.; Albuquerque, J. F. C.; Andricopulo, A. D.; *J. Chem. Inf. Model.* **2010**, *50*, 1693.

2. Gashaw, I.; Ellinghaus, P.; Sommer, A.; Asadullah, K.; *Drug Discovery Today* **2011**, *17*, S24.

3. Chiaradia, L. D.; Martins, P. G. A.; Cordeiro, M. N. S.; Guido, R. V. C.; Ecco, G.; Andricopulo, A. D.; Yunes, R. A.; Vernal, J.; Nunes, R. J.; Terenzi, H.; *J. Med. Chem.* **2012**, *55*, 390.

4. Segall, M.; Champness, E.; Obrezanova, O.; Leeding, C.; *Chem. Biodivers.* **2009**, *6*, 2144.

5. Michielan, L.; Moro, S.; *J. Chem. Inf. Model.* **2010**, *50*, 961.

6. Dragunow, M.; *Nat. Rev. Drug Discovery* 2008, *7*, 659.

7. Willis, C. L.; *Methods Mol. Biol.* **2012**, *814*, 515.

8. Gerebtzoff, G.; Seelig, A.; *J. Chem. Inf. Model.* **2006**, *46*, 2638.

9. Garg, P.; Verma, J.; *J. Chem. Inf. Model.* **2006**, *46*, 289.

10. Moda, T. L.; Montanari, C. A.; Andricopulo, A. D.; *Bioorg. Med. Chem.* **2007**, *15*, 7738.

11. Hansch, C.; Leo, A.; Mekapati, S. B.; Kurup, A.; *Bioorg. Med. Chem.* **2004**, *12*, 3391.

12. Tropsha, A.; *Mol. Inf.* **2010**, *29*, 476.

13. Gleeson, M. P.; Hersey, A.; Hannongbua, S.; *Curr. Top. Med. Chem.* **2011**, *11*, 358

14. Moda, T. L.; Andricopulo, A. D.; *Bioorg. Med. Chem. Lett.* **2012**, *22*, 2889.

15. Hansch, C.; Kurup, A.; Garg, R.; Gao, H.; *Chem. Rev.* **2001**, *101*, 619.

16. Cecchelli, R.; Berezowski, V.; Lundquist, S.; Culot, M.; Renftel, M.; Dehouck, M. P.; Fenart, L.; *Nat. Rev. Drug Discovery* **2007**, *6*, 650.

17. Fischer, H.; Gottschlich, R.; Seelig, A.; *J. Membr. Biol.***1998**, *165*, 201.

18. Nielsen, P. A.; Andersson, O.; Hansen, S. H.; Simonsen, K. B.; Andersson, G.; *Drug Discovery Today* **2011**, *16*, 472.

19. Nicolazzo, J. A.; Charman, S. A.; Charman, W. N.; *J. Pharm. Pharmacol.* **2006**, *58*, 281.

20. Di, L.; Kerns, E. H.; Bezar, I. F.; Petusky, S. L.; Huang, Y.; *J. Pharm. Sci.* **2009**, *98*, 1980.

21. Carrara, S.; Reali, V.; Misiano, P.; Dondio, G.; Bigogno, C.; *Int. J. Pharm.* **2007**, *345*, 125.

22. Fan, Y.; Unwalla, R.; Denny, R. A.; Di, L.; Kerns, E. H.; Diller, D. J.; Humblet, C.; *J. Chem. Inf. Model.***2010**, *50*, 1123.

23. Abbott, N. J.; Dolman, D. E.; Drndarski, S.; Fredriksson, S. M.; *Methods Mol. Biol.***2012**, *814*, 415.

24. Bickel, M. H.; Genny, R.; *J. Pharm. Pharmacol.* **1980**, *32*, 669.

25. Mandema, J. W.; Kuck, M. T.; Danhof, M.; *Br. J. Pharmacol.* **1992**, *105*, 164.

26. Tsuneizumi, T.; Babb, S. M.; Cohen, B. M.; *Biol. Psychiatry* **1992**, *32*, 817.

27. Dambisya, Y. M.; Chan, K.; Wong, C. L.; *J. Pharm. Pharmacol.* **1992**, *44*, 687.

28. Abraham, M. H.; Chadha, H. S.; Mitchell, R. C.; *J. Pharm. Sci.* **2006**, *83*, 1257.

29. de Lange, E. C.; Danhof, M.; de Boer, A. G.; Breimer, D. D.; *Brain Res.* **1994**, *666*, 1.

30. Argenti, D.; D'Mello, A. P.; *J. Pharmacol. Exp. Ther.* **1994**, *270*, 512.

31. Chadha, H. S.; Abraham, M. H.; Mitchell, R. C.; *Bioorg. Med. Chem. Lett.* **1994**, *4*, 2511.

32. Lombardo, F.; Blake, J. F.; Curatolo, W. J.; *J. Med. Chem.* **1996**, *39*, 4750.

33. Salminen, T.; Pulli, A.; Taskinen, J.; *J. Pharm. Biomed. Anal.* **1997**, *15*, 469.

34. Norinder, U.; Sjöberg, P.; Österberg, T.; *J. Pharm. Sci.* **1998**, *87*, 952.

35. Chou, R. C.; Levy, G.; *J. Pharmacol. Exp. Ther.* **1981**, *219*, 42.

36. Luco, J. M.; *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 396.

37. Kelder, J.; Grootenhuis, P. D. J.; Bayada, D. M.; Delbressine, L. P. C.; Ploemen, J. P.; *Pharm. Res.* **1999**, *16*, 1514.

38. Rose, K.; Hall, L. H.; *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 651.

39. Hutter, M. C.; *J. Comput. Aided Mol. Des.* **2003**, *17*, 415.

40. Hou, T. J.; Xu, X. J.; *J. Chem. Inf. Comput. Sci.* **2002**, *43*, 2137.

41. Winkler, D. A.; Burden, F. R.; *J. Mol. Graphics Modell.* **2004**, *22*, 499.

42. Okuyama, S.; Aihara, H.; *Jpn. J. Pharmacol.* **1984**, *35*, 95.

43. Bolander, H. G.; Wahlström, G.; Norberg, L.; *Acta Pharmacol. Toxicol.* **1984**, *54*, 33.

44. Arendt, R. M.; Greenblatt, D. J.; Liebisch, D. C.; Luu, M. D.; Paul, S. M.; *Psychopharmacology* **1987**, *93*, 72.

45. Young, R. C.; Mitchell, R. C.; Brown, T. H.; Ganellin, C. R.; Griffiths, R.; Jones, M.; Rana, K. K.; Saunders, D.; Smith, I. R.; *J. Med. Chem.* **1988***, 31*, 656.

46. Kaplan, G. B.; Greenblatt, D. J.; Leduc, B. W.; Thompson, M. L.; Shader, R. I.; *J. Pharmacol. Exp. Ther.* **1989**, *248*, 1078.

47. Yata, N.; Toyoda, T.; Murakami, T.; Nishiura, A.; Higashi, Y.; *Pharm. Res.* **1990**, *7*, 1019.

48. Moda, T. L.; Torres, L. G.; Carrara, A. E.; Andricopulo, A. D.; *Bioinformatics* **2008**, *24*, 2270.

49. Tripos Inc.; http://tripos.com/ accessed in September 2012.

50. Chemaxon Standardizer; http://www.chemaxon.com/products/standardizer/ accessed in September 2012.

51. HQSAR™; *User's Manual*, St. Louis: Tripos Inc., 2006.