Opinion

Hansch analysis 50 years on

Yvonne Connolly Martin*

The invention of quantitative structure–activity relationships (QSAR) launched the use of computers to understand the multiple factors that contribute to the relationship between the chemical and biological properties of molecules. The original Hansch–Fujita QSAR continues to be performed to this day, fifty years since its inception. In addition, it has inspired vigorous research that has inspired many related methods: for example, methods that use descriptors such as substructures or three-dimensional features or methods that are based on mathematical models such as partial least squares or recursive partitioning. The influence of QSAR is felt in the recognition of the value of predictions from such models by regulatory agencies and by the widespread use of octanol–water log *P* in agricultural, environmental, and medicinal chemistry.

How to cite this article:
WIREs Comput Mol Sci 2012, 2: 435–442 doi: 10.1002/wcms.1096

INTRODUCTION

In July 1961, approximately 50 years ago, Corwin Hansch and Toshio Fujita formulated the first QSAR equation.[1] It grew out of their 10-plus years of mutual interest in the structure–activity relationships (SARs) of the plant-growth regulators.[2,3] Frustrated by the inability of the Hammett equation to explain the SAR, Hansch turned to octanol–water partition coefficients.[4] When these also did not explain the SAR, Fujita suggested that both properties must be considered, following the precedent of Taft, who had demonstrated the combined electronic and steric effects on the rate of hydrolysis of esters.[5] Equation (1) shows the relationship that they proposed to fit and Eq. (2) shows the resulting fit of their data:

$$\log(1/C) = k\pi - k'\pi^2 + \rho\sigma + k'' \qquad (1)$$

$$\log(1/C) = 4.08\pi - 2.14\pi^2 + 2.78\sigma + 3.36 \qquad (2)$$

In this equation, π is the difference in the log of the octanol–water partition coefficient of the analogue as compared with the unsubstituted compound and σ is the Hammett constant referred to the ortho position on the aromatic ring. The various *k*s are constants to be fit with the computer. The π^2 term indicates that there is an optimum π value. Note that in this and subsequent early publications,[6,7] there is no term for the steric effect of substituents. Also, no statistics of the fit were presented.

Instead of using a mechanical calculator to fit their multiparameter equation, as Pavelich and Taft[8] and Jaffé[9] had done, Donald McIntyre, their colleague from the geology department helped them fit their data using a primitive computer.[2] Using the computer changed the process from hours of careful work at a calculator using a knowledge of statistics, to minutes, at a keypunch entering just the physical and biological properties of the molecules. Hence, the decision to use a computer to fit the equation opened up the field to include the analysis of larger data sets and more descriptors; it encouraged others to try the method; and it also added the aura that the method is cutting edge.

Even though in this first publication,[1] they made only passing claim of the universality of their approach to investigate the SAR within a series, all the following elements of QSAR were present: (1) using π, the change in octanol–water log *P*, as a descriptor of the lipophilic/hydrophobic effects of substituents; (2) suggesting that here is an optimum π or log *P* within a series; (3) considering that not only log *P* but also electronic substituent effects as described by Hammett σ constants influence relative potency; and (4) fitting the relationship with a computer. Each of these elements has been the focus of many studies in the broader QSAR literature.

*Correspondence to: yvonnecmartin@comcast.net
Martin Consulting, Waukegan, IL USA
DOI: 10.1002/wcms.1096

According to an August 2011 Google Scholar search, this report has been cited 447 times; the 1964 report that generalized their findings,[7] 1357 times; and the book by Hansch and Leo,[10] 1006 times. The corresponding numbers for Web of Science search are 433, 467, and 1790—all impressive figures. It is claimed that the work of Hansch and Leo has been cited more than 100,000 times in the past 10 years.[11] The impact of this work is also reflected in an August 2011 Google search that returned 11,000 hits for the phrase Hansch QSAR.

CURRENT STATE OF TRADITIONAL QSAR

Throughout the remainder of his career, Hansch continued to fit QSARs and published more than 200 articles on the subject. There are now approximately 13,000 biological QSARs and 9000 physicochemical linear free energy relationships (LFERs) in the CQSAR database that he, Al Leo, and David Hoekman use to store equations and data.[11] Many of these equations are unpublished. Fujita also published more than 250 articles on QSAR, with special emphasis on the application to agrochemicals.[12]

Although some might claim otherwise, QSAR is not dead.[13] Evidence for this statement is the observation that in the 19 months from January 1, 2010, to August 1, 2011, MEDLINE identified 131 publications with QSAR as the search term. Although 45 of these emphasized three-dimensional (3D) QSAR, all are QSAR publications nonetheless. These reports appeared in 56 journals. *Bioorganic & Medicinal Chemistry* leads the list with 29 articles, *Chemical Biology & Drug Design* follows with eight, *European Journal of Medicinal Chemistry* and *Journal of Medicinal Chemistry* with seven each, *International Journal of Molecular Sciences* with six, and *Chemometrics and Intelligent Laboratory Systems* and *Journal of Molecular Graphics and Modeling* with five each.

Of these 131 articles, 99 address structures of medicinal chemistry interest; 12 address issues of toxicity; six deal with absorption, distribution, metabolism, and excretion (ADME) properties; four with environmental concerns; and two each with bioinformatics, cheminformatics, chemical properties, food, and QSAR methodology. Thus, it appears that QSAR is widely used to probe the SARs within a series, but that toxicity and ADME properties are also of interest.

The international scope of the use of QSAR is evident in the 665 authors of these publications. It is especially noteworthy that only six scientists who are authors on four or more publications are collaborators based in Beijing at the College of Pharmaceutical Sciences, Capital Medical University, and College of Pharmaceutical Sciences, Peking University. Next on the list with three QSAR publications is Alexander Tropsha from the University of North Carolina at Chapel Hill. Robust contributions from researchers in India are also apparent.

Some recent publications suggest the extent of biological targets considered for QSAR. Vyas and coworkers[14] developed QSAR equations that suggest that hydrophobicity and electronic effects govern the inhibitory potency of 3-aminopyrazolopyridine ureas against KDR in cell-free and whole-cell assays. Complementary models highlight the importance of polar surface area, total potential energy, principal moment of inertia, and bend energy to alternative models; these different results highlight the problem of assigning interpretations to any one QSAR model. Yuan and coworkers[15] showed that hydrophobicity and steric factors are the main determinants of the recognition of phenylurea herbicides by an antibody generated to an immobilized analogue. The analysis by Jain and Chaturvedi[16] suggested that electronic and steric effects dominate the SAR of the angiotensin II receptor antagonist potency of substituted 5-(biphenyl-4-ylmethyl)pyrazoles. Song and coworkers[17] performed QSAR analysis on diarylpyrazole imide analogues as cannabinoid receptor-2 antagonists. Their results emphasize the importance of steric effects on affinity and that no analogues with increased potency could be predicted. Sivaprakasam and coworkers[18] reported the QSAR analysis of the inhibition of glycogen synthetase kinase 3a by 3-anilino-4-phenylmaleimides. They found that substituents on the anilino group exert a positive hydrophobic effect, whereas substituents on the phenyl group exert complex electronic and steric effects. Leonard and Roy[19] analyzed the binding affinity of 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas for the CCR5 receptor. They found that electron-withdrawing substituents at the para positions of the biphenyl portion enhance affinity, that there is an optimum octanol–water log P of 5.58, and that complex steric effects also affect affinity. Thus, QSAR has been recently applied to both receptor affinity and enzyme inhibition potency.

Quantitative structure–activity relationships is widely used in the analysis of toxic endpoints of molecules.[20] A book devoted to the QSAR of mutagens and carcinogens was published in 2003.[21] In addition, recently investigated end points include hERG,[22] the common cold,[23] and melanoma.[24]

Although it is powerful, traditional QSAR has several inherent limitations: Hammett σ constants

may not accurately describe the electronic effect of substituents on the biological interactions of interest. Frequently, the σ values are not available in compilations, but one must guess at a value or perform a quantum chemical calculation. Molecular shape is poorly described with transferable constants. It can be challenging to apply the approach to a set of molecules that do not share a common core. Finally, multiple regression analysis is not appropriate to use if the molecular descriptors are correlated or if the response is categorical such as 'mutagenic' or 'non mutagenic'. Traditional QSAR and any method that examines the quantitative relationship between chemical structure and biological activity can be puzzled by activity cliffs in which a seemingly minor structural change can lead to a dramatic difference in potency.[25]

In addition to the inherent limitations of QSAR, there are also pitfalls to its appropriate use. One problem is the omission of outliers: Such outliers from an equation can provide valuable information if a good reason for the lack of fit can be suggested,[26,27] but thoughtlessly omitting outliers can lead to equations with no predictive value. Another pitfall is to interpret a fit to a QSAR equation as a proof that a particular physical property contributes to potency while ignoring the possibility that the property in question may be correlated with another that equally well fits the data. The report by Vyas and coworkers provides the example of multiple explanations from different QSARs of the same data set.[14] A trivial example is that in a series of alkyl analogues, hydrophobicity is highly correlated with the following two properties that are usually interpreted to indicate steric effects: molar refractivity and the STERIMOL L parameter. If a data set was not designed for QSAR analysis, it may not be possible to overcome such correlations between properties.

APPROACHES THAT OVERCOME THE LIMITATIONS OF TRADITIONAL QSAR

The limitations of traditional QSAR did not stop development in the field, but instead inspired many workers to devise solutions to the problems. These approaches involve inventing more easily calculated descriptors of molecules, specific solutions to describe differences in molecular shape within a set of molecules, and expanding the repertoire of statistical and machine learning methods to analyze the data sets. These various new methods retain the essence of traditional QSAR, that is, the use of a computer to develop a model that relates the chemical and structural features of molecules to their biological properties.

It is now easy to calculate from the structure diagram not just octanol–water log P,[28–30] but thousands of molecular descriptors to use for QSAR.[31] For example, such descriptors may be derived from molecular connectivity analysis,[32] counts or recognition of particular fragments in a molecule,[33,34] or specific fingerprints originally designed to aid substructure searching.[35] Although such descriptors might not provide the clear relationship between physical properties and potency, many have been found to have predictive value.[36]

Today there are many sources of readily available software to develop a model for the relationship between molecular and biological properties. For example, Excel provides for the calculation of a regression equation; the community-supported R statistical package provides access to hundreds of statistical and visualization methods; and Weka provides many machine-learning and data-mining methods.[37,38]

Since the pioneering development of Comparative Molecular Field Analysis,[39] 3D QSAR methods[40–42] have become a valuable approach to the analysis of SARs. Typically, they require choosing a conformation for the calculations. This may rely on explicit or implicit superposition of the molecules over the atoms proposed to be essential for recognition by the target biomolecule,[43–45] or a rule-based method for generating conformations.[46] The resulting shape descriptors might be calculated from aligned molecules, for example aligned in a lattice,[39] or from the conformation itself.[47,48] However, there are also methods that select the conformation as part of the QSAR.[49,50]

The final difference between traditional QSAR and that practiced today is that although the computer is still used to derive a model, the model is not necessarily derived by regression analysis. Instead, one may fit and test the model with various neural networks,[51,52] partial least-squares regression,[53] support vector machines,[54,55] etc. The approach has also been broadened to include classification methods (e.g., active vs inactive) with discriminant analysis,[56] recursive partitioning,[57] Bayesian classifiers,[58] etc.

The pitfalls of QSAR noted in the previous section extend to these expanded methods. In addition, early studies by Topliss and coworkers[59] highlighted the problem of examining too many potential descriptors when using regression analysis. This represents a potential problem with QSAR based on many calculated properties. The pitfall of using many descriptors has led to the use of various validation methods to assess the confidence that one might have in a model.[60,61]

A more serious issue with some of the approaches is that the easy interpretation and use of the models is not forthcoming—hence the interest in 'inverse QSAR' methods that design molecules to fix a complex equation.[62]

Finally, a particular pitfall of 3D QSAR is the tendency to believe that the results shed direct insight into the structure of the macromolecular binding site: Again, artificial correlations may obscure the true basis of the statistical result.

THE USE OF QSAR PREDICTIONS BY REGULATORY AGENCIES

A major validation of the utility of QSAR models and impetus for the improvement in methodology is the recognition that they have enough signal to be useful for setting priorities of regulatory agencies. Both the United States Food and Drug Administration (US FDA) and the European Union promote the use of QSARs to identify chemicals of concern.

Thus, the US FDA implemented CRADA (Confidential Research and Development Agreements) to provide software vendors with validated toxicity data to support the development of QSARs. The resulting QSARs are validated by the US FDA.[20,63–66] Importantly, they focus on the complementarity between models developed from the same data but with different descriptors or statistical methods. They also address the important issue of the applicability domain, which molecules should be predicted, of the models.

The European Union created the European Chemicals Agency to administer REACH (Registration, Evaluation, Authorisation and restriction of CHemicals). It states that for each chemical circulating in the European territory, a complete dossier on physicochemical, biological, and toxicological properties has to be compiled.[67] QSAR results can be used provided that (1) they are derived from a (Q)SAR model whose scientific validity has been established, (2) the substance falls within the applicability domain of the (Q)SAR model, (3) the results are adequate for the purpose of classification and labeling and/or risk assessment, and (4) adequate and reliable documentation of the applied method is provided.[68]

THE SPECIAL ROLE OF HYDROPHOBICITY ON THE BIOLOGICAL PROPERTIES OF MOLECULES

Perhaps the most pervasive influence of QSAR has been the recognition that the octanol–water log *P* of a molecule influences not just its penetration through membranes, but also its affinity for the biological target. At the same time that Hansch decided to use octanol–water log *P* as a property that might be related to potency, biochemists recognized the key role that hydrophobic forces play in protein stability.[69] In fact, hydrophobicity is thought to be the main driving force for moving a molecule from water into a binding site.[70,71]

That the importance of the log *P* of a molecule has penetrated the thinking of medicinal chemists is illustrated by the fact that an August 2011 search of the *Journal of Medicinal Chemistry* yielded 993 hits on the terms 'octanol and water and log'. However, log *P* is not relegated to medicinal chemistry as demonstrated by the numbers in Table 1. In fact, the journal *Environmental Science and Toxicology* contains the most articles (1422) that discuss octanol and water. Note also that, in accord with Fujita's influence on the field, the *Journal of Agricultural and Food Chemistry* contains more than 300 articles that discuss the issue.

Table 1 shows that the importance of octanol–water log *P* continues to be a concept of importance in the articles published from January 2010 to July 2011. At least 22 scientific journals publish more than 10 articles that include these terms. Of those articles from journals published by the American Chemical Society, 113 are considered to be related to environmental concerns, 99 to medicinal chemistry, 55 to biochemistry, 27 to agricultural chemistry, and 25 to toxicology. This again illustrates the reach of one of the central concepts of QSAR into many areas of investigation into the biological effects of chemicals.

The value of log *P* to biological research is underscored by the development of many predictors of this property. Although CLOGP is an expert system,[72] many of the other programs derive the predictions from the same type of descriptors and statistics used by QSAR.[73,74]

Log *P* is a key property for triaging hits from high throughput screening and combinatorial chemistry. Which of the active hits is a better choice for lead optimization? The octanol–water log *P* is frequently one criterion or it might be combined with potency as in the ligand-lipophilic efficiency index, LLE (LLE = pIC50–log *P*).[75] In addition, the popular rule-of-five for predicting permeability includes log *P* as one of its four criteria.[76] The Golden Triangle considers the distribution coefficient *D*, which is correlated with log *P* within a series of constant pK_a.[77] Log *P* is also a prime factor in models for ADME[78–80] and brain penetration.[81,82] In such models, notion of an optimum is often implied, if not made explicitly.

TABLE 1 | Number of Articles that Include the Words 'Octanol' and 'Water' and 'Log'

Journal	Total Number of Articles	Articles Published From January 2010–July 2011
Environmental Science & Technology	1422	145
Journal of Medicinal Chemistry	993	87
Journal of Pharmaceutical Sciences	735	52
European Journal of Medicinal Chemistry	165	51
Journal of Agricultural and Food Chemistry	310	42
Journal of Chemical Information and Modeling	357	32
Bioorganic & Medicinal Chemistry	255	31
Medicinal Chemistry Research 2004–2011	39	31
Journal of Chemistry Engineering Data	114	22
Molecular Informatics (incorporating QSAR & Combinatorial Science and Quantitative Structure–Activity Relationships)	292	20
Journal of Environmental Monitoring	66	20
Langmuir	124	19
Analytical Chemistry	275	18
Journal of Physical Chemistry B	196	18
Journal of Physical Chemistry C	196	18
Molecular Pharmaceutics	58	17
Chemistry Research in Toxicology	89	16
Bioconjugate Chemistry	68	13
Dalton Transactions	42	13
ACS Symposium Series	184	12
Industrial Engineering Chemistry Research	112	11
Chemistry Reviews	84	11

CONCLUSIONS

Although the traditional Hansch–Fujita QSAR approach continues to be practiced today, the seminal papers inspired a plethora of other approaches to understand biological SARs, an appreciation of the power of such approaches by regulatory agencies, and the widespread understanding of the importance of the octanol–water log *P* to the biological potency of compounds.

REFERENCES

1. Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* 1962, 194:178–180.
2. Hansch C. The advent and evolution of QSAR at Pomona college. *J Comput Aided Mol Des* 2011, 25:495–507.
3. Fujita T. In memoriam professor Corwin Hansch: birth pangs of QSAR before 1961. *J Comput Aided Mol Des* 2011, 25:509–517.
4. Hansch C. A quantitative approach to biochemical structure-activity relationships. *Acc Chem Res* 1969, 2:232–241.
5. Taft RW Jr. Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters. 1. *J Am Chem Soc* 1952, 74:3120–3128.
6. Hansch C, Muir RM, Fujita T, Maloney PP, Geiger F, Streich M. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. *J Am Chem Soc* 1963, 85:2817–2824.
7. Hansch C, Fujita T. Rho sigma pi analysis. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc* 1964, 86:1616–1626.
8. Pavelich WA, Taft RW Jr. The evaluation of inductive and steric effects on reactivity. The methoxide ion-catalyzed rates of methanolysis of l-menthyl esters in methanol. *J Am Chem Soc* 1957, 79:4935–4940.

9. Jaffé HH. Theoretical considerations concerning Hammett's equation. IV. Calculation of values. *J Chem Phys* 1953, 21:415.
10. Hansch C, Leo A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. Washington, DC: American Chemical Society; 1995.
11. http://biobyte.com/index.html. Claremont, CA (Accessed January 2, 2012).
12. Hammock BD. Innovation in discovery. The career of Toshio Fujita. In: *Natural and Engineered Pest Management Agents*. Hedin PA, Menn JJ, Hollingworth RM, eds. Washington, DC: American Chemical Society; 1993, xv–xvi.
13. Doweyko A. QSAR: dead or alive? *J Comput Aided Mol Des* 2008, 22:81–89.
14. Vyas VK, Joshi G, Namdeo B, Gupta A. Exploring structure indenture of some aminopyrazolopyridine ureas as potent VEGFR/PDGFR multitargeted kinase inhibitors: a QSAR approach. *Indian J Chem Sect B* 2011, 50:858–867.
15. Yuan M, Liu B, Liu E, Sheng W, Zhang Y, Crossan A, Kennedy I, Wang S. Immunoassay for phenylurea herbicides: application of molecular modeling and quantitative structure-activity relationship analysis on an antigen–antibody interaction study. *Anal Chem* 2011, 83:4767–4774.
16. Jain A, Chaturvedi SC. QSAR modeling of some substituted benzimidazole as angiotensin II AT1 receptor antagonist. *Med Chem Res* 2010, 19:177–185.
17. Song K-S, Kim MJ, Seo HJ, Lee S-H, Jung ME, Kim S-U, Kim J, Lee J. Synthesis and structure–activity relationship of novel diarylpyrazole imide analogues as CB1 cannabinoid receptor ligands. *Bioorg Med Chem* 2009, 17:3080–3092.
18. Sivaprakasam P, Xie AH, Doerksen RJ. Probing the physicochemical and structural requirements for glycogen synthase kinase-3 alpha inhibition: 2D-QSAR for 3-anilino-4- phenylmaleimides. *Bioorg Med Chem* 2006, 14:8210–8218.
19. Leonard JT, Roy K. Comparative QSAR modeling of CCR5 receptor binding affinity of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas. *Bioorg Med Chem Lett* 2006, 16:4467–4474.
20. Matthews EJ, Contrera JF. In silico approaches to explore toxicity endpoints: issues and concerns for estimating human health effects. *Expert Opin Drug Metab Toxicol* 2007, 3:125–134.
21. Benigni R, ed. *Quantitative Structure-Activity Relationship (QSAR) Modles of Mutagens and Carcinogens*. Boca Raton, FL: CRC Press; 2003.
22. Aronov AM. Tuning out of hERG. *Curr Opin Drug Discov Devel* 2008, 11:128–140.
23. Verma RP, Hansch C. Understanding human rhinovirus infections in terms of QSAR. *Virology* 2007, 359:152–161.
24. Verma RP, Mekapati SB, Kurup A, Hansch C. A QSAR review on melanoma toxicity. *Bioorg Med Chem* 2005, 13:5508–5526.
25. Maggiora GM. On outliers and activity cliffs—why QSAR often disappoints. *J Chem Inf Model* 2006, 46:1535–1535.
26. Verma RP, Hansch C. An approach toward the problem of outliers in QSAR. *Bioorg Med Chem* 2005, 13:4597–4621.
27. Kim KH. Outliers in SAR and QSAR: is unusual binding mode a possible source of outliers? *J Comp Aided Mol Design* 2007, 21:63–86.
28. Mannhold R, van de Waterbeemd H. Substructure and whole molecule approaches for calculating log *P*. *J Comput Aided Mol Des* 2001, 15:337–354.
29. Martin YC. In: *Quantitative Drug Design A Critical Introduction*. 2nd ed. Boca Raton, FL: CRC Press; 2010, 58:73–74.
30. Mannhold R, Poda GI, Ostermann C, Tetko IV. Calculation of molecular lipophilicity: state-of-the-art and comparison of log *P* methods on more than 96,000 compounds. *J Pharm Sci* 2009, 98:861–893.
31. Todeschini R, Consonni V. *Handbook of Molecular Descriptors*. Weinheim: Wiley-VCH; 2008.
32. Hall L, Kier L. *Molconn-z. 4.00 Version*. Richmond, VA: Edusoft; 2007.
33. Blower P, Fligner M, Verducci J, Bjoraker J. On combining recursive partitioning and simulated annealing to detect groups of biologically active compounds. *J Chem Inf Comput Sci* 2002, 42:393–404.
34. Klopman G, Chakravarti SK, Zhu H, Ivanov JM, Saiakhov RD. ESP: a method to predict toxicity and pharmacological properties of chemicals using multiple mcase databases. *J Chem Inf Comp Sci* 2004, 44:704–715.
35. Scitegic chemistry components vol. 2008. Accelry, San Diego, CA; 2008.
36. Hall LM, Hall LH, Kier LB. QSAR modeling of beta-lactam binding to human serum proteins. *J Comp Aided Mol Design* 2003, 17:103–118.
37. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. 2nd ed. San Francisco, CA: Morgan Kaufmann; 2005.
38. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explorations* 2009, 11–18.
39. Cramer III RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 1988, 110:5959–5967.
40. Kubinyi H, ed. *3D QSAR in Drug Design: Theory Methods and Applications*. Vol 1. Leiden: ESCOM; 1993.

41. Kubinyi H, Folkers G, Martin YC, eds. *3D QSAR in Drug Design. Ligand-Protein Interactions and Molecular Similarity*. Vol 2. Leiden: ESCOM; 1998.
42. Kubinyi H, Folkers G, Martin YC, eds. *3D QSAR in Drug Design. Recent Advances. Vol 3*. Leiden: ESCOM; 1998.
43. Martin YC. Pharmacophore modeling: 2—applications. In: *Comprehensive Medicinal Chemistry II. Vol 4*. Mason JS, ed. Oxford: Elsevier; 2007, 515–536. Triggle DJ, Taylor JB, Series, eds.
44. Martin YC. Pharmacophore modeling: 1—Methods. In: *Comprehensive Medicinal Chemistry II*. Vol 4. Mason JS, ed. Oxford: Elsevier; 2007, 119–147. Triggle DJ, Taylor JB, Series, eds.
45. Martin YC. Preparation of 3D structures of molecules for 3D QSAR. In: *Quantitative Drug Design a Critical Introduction*. 2nd ed. Boca Raton, FL: CRC Press; 2010, Chap 3:31–47.
46. Cramer III RD, Clark RD, Patterson DE, Ferguson AM. Bioisosterism as a molecular diversity descriptor: steric fields of single topomeric conformers. *J Medicin Chem* 1996, 39:3060–3069.
47. Clementi S, Cruciani G, Riganelli D, Valigi R, Costantino G, Baroni M, Wold S. Autocorrelation as a tool for a congruent description of molecules in 3D-QSAR studies. *Pharm Pharmacol Lett* 1993, 3:5–8.
48. Wagener M, Sadowski J, Gasteiger J. Autocorrelation of molecular surface properties for modeling corticosteroid binding globulin and cytosolic AH receptor activity by neural networks. *J Am Chem Soc* 1995, 117:7769–7775.
49. Jain AN, Dietterich TG, Lathrop RH, Chapman D, Critchlow RE Jr, Bauer BE, Webster TA, Lozano-Perez T. Compass: a shape-based machine learning tool for drug design. *J Comput Aided Mol Des* 1994, 8:635–652.
50. Sprague PW. Automated chemical hypothesis generation and database searching with catalyst. *Perspect Drug Discov* 1995, 3:1–20.
51. Winkler DA, Burden FR. Robust QSAR models from novel descriptors and Bayesian regularised neural networks. *Mol Simulat* 2000, 24:243–258.
52. Livingstone DJ, Manallack DT, Tetko IV. Data modeling with neural networks—advantages and limitations. *J Comp Aided Mol Design* 1997, 11:135–142.
53. Wold S, Ruhe A, Wold H, Dunn WJ. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J Sci Stat Comp* 1984, 5:735–743.
54. Norinder U. Support vector machine models in drug design: applications to drug transport processes and QSAR using simplex optimisations and variable selection. *Neurocomputing* 2003, 55:337–346.
55. Byvatov E, Fechner U, Sadowski J, Schneider G. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J Chem Inf Comput Sci* 2003, 43:1882–1889.
56. Martin YC, Holland JB, Jarboe CH, Plotnikoff N. Discriminant analysis of the relationship between physical properties and inhibition of monoamine oxidase by aminotetralins and aminoindanes. *J Med Chem* 1973, 17:409–413.
57. Hawkins DM, Young SS, Rusinko A. Analysis of a large structure-activity data set using recursive partitioning. *Quant Struct-Act Relationships* 1997, 16:296–302.
58. Watson P. Naïve Bayes classification using 2D pharmacophore feature triplet vectors. *J Chem Inf Model* 2008, 48:166–178.
59. Topliss JG, Edwards RP. Chance factors in studies of quantitative structure-activity relationships. *J Med Chem* 1979, 22:1238–1244.
60. Gramatica P. Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 2007, 26:694–701.
61. Guidance document on the validation of (quantitative)structure–activity relationships (QSAR) models. *Joint meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides, and Biotechnology*. Paris: Organization for Economic Co-operation and Development; 2007, 153.
62. Cho SJ, Zheng WF, Tropsha A. Rational combinatorial library design. 2. rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches. *J Chem Inf Comput Sci* 1998, 38:259–268.
63. Matthews EJ, Kruhlak NL, Daniel Benz R, Sabaté DA, Marchant CA, Contrera JF. Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: part C: Use of QSAR and an expert system for the estimation of the mechanism of action of drug-induced hepatobiliary and urinary tract toxicities. *Regul Toxicol Pharmacol* 2009, 54:43–65.
64. Matthews EJ, Ursem CJ, Kruhlak NL, Benz RD, Sabaté DA, Yang C, Klopman G, Contrera JF. Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: part B. Use of (Q)SAR systems for early detection of drug-induced hepatobiliary and urinary tract toxicities. *Regul Toxicol Pharmacol* 2009, 54:23–42.
65. Ursem CJ, Kruhlak NL, Contrera JF, MacLaughlin PM, Benz RD, Matthews EJ. Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans. Part A: use of FDA post-market reports to create a database of hepatobiliary and urinary tract toxicities. *Regul Toxicol Pharmacol* 2009, 54:1–22.
66. Matthews EJ. *A Pharmaceutical Perspective on QSAR and Expert System Tools*. Crystal City, VA: Pesticide Program Dialogue Committee; 2009. http://epa.gov/pesticides/ppdc/testing/jan09/ejm-presentation.pdf. (Accessed January 2, 2012).

67. The European Parliament and The Council Of The European Union. Regulation (EC) no 1907/2006 of the European parliament and of the council. *OJ* 2007, L136: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:136:0003:0280:en:PDF. (Accessed January 2, 2012).
68. QSARs and grouping of chemicals In: *Guidance On Information Requirements And Chemical Safety Assessment*. Chapter R.6: http://echa.europa.eu/documents/10162/17224/information_requirements_r6_en.pdf. (Accessed January 2, 2012) 134 pages.
69. Kauzmann W. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 1959, 14:1–63.
70. Eisenberg D, McLachian AD. Solvation energy in protein folding and binding. *Nature* 1986, 319:199–203.
71. Miyamoto S, Kollman PA. What determines the strength of noncovalent association of ligands to proteins in aqueous solution? *Proc Natl Acad Sci USA* 1993, 90:8402–8406.
72. Leo A. *CLOGP. 4.3 ed*. Claremont, CA: Biobyte Corporation; 2007.
73. Klopman G, Wang S, Anderson PS, Huff JR. A computer automated structure evaluation (CASE) approach to calculation of partition coefficient. *J Comput Chem* 1991, 12:1025–1032.
74. Viswanadhan VN, Ghose AK, Wendoloski JJ. Estimating aqueous solvation and lipophilicity of small organic molecules: a comparative overview of atom/group contribution methods. *Perspect Drug Discov* 2000, 19:58–98.
75. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* 2007, 6:881–890.
76. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997, 23:3–25.
77. Johnson TW, Dress KR, Edwards M. Using the golden triangle to optimize clearance and oral absorption. *Bioorg Med Chem Lett* 2009, 19:5560–5564.
78. Hansch C, Leo A, Mekapati SB, Kurup A. QSAR and ADME. *Bioorg Med Chem* 2004, 12:3391–3400.
79. van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* 2003, 2:192–204.
80. Norinder U. In silico modelling of ADMET—a minireview of work from 2000 to 2004. *SAR QSAR Environ Res* 2005, 16:1–11.
81. Clark DE. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. *J Pharm Sci* 1999, 88:815–821.
82. Norinder U, Haeberlein M. Computational approaches to the prediction of the blood-brain distribution. *Adv Drug Deliv Rev* 2002, 54:291–313.

