# The History of Drug Research: From Hansch to the Present

**Han van de Waterbeemd**

F. Hoffmann-La Roche Ltd., Pharma Research New Technologies, CH-4002 Basel, Switzerland

**Abstract**

The introduction by Hansch of the QSAR concept in the design of bioactive molecules opened new perspectives in organic chemistry. In the present contribution we will briefly review a number of factors contributing to the development of modern medicinal chemistry, particularly focussing on the increasing role of computers.

**Key words:** Structure-property correlations (SPC), rational structure design, biostructure research, QSAR parameters, lipophilicity.

## From Rational to More Rational Structure Design

In the previous paper it was described how structure-activity relationships can be given a quantitative aspect [1]. The pioneering work of Hansch led to a great optimism or even euphoria in the fields of drug, pesticide and agrochemicals design [2]. QSAR practioners however soon had to return to realism by recognizing that many more tools were needed besides the Hansch analysis approach. Indeed, these new methods have their merits and open more possibilities for rational lead optimization. However, an important amount of experience is required to make the best out of it [3].

In this paper we will describe some important new developments in the field of medicinal chemistry and particularly pay attention to the ever increasing role of computers in the drug discovery process. In fact a complete revolution took place in the last three decades. The traditional pharmaceutical chemistry has evolved to become a new discipline named medicinal chemistry or pharmacochemistry. This progress was made due to emerging important new concepts and possibilities. SAR was extended to QSAR, laboratory equipment improved dramatically, many new synthons, e.g. for asymmetric synthesis, were created, Dreiding and CPK models can be handled on a computer screen and large data bases became available on-line. Computer-assisted molecular modeling (CAMM) and multivariate statistical approaches have become well-accepted by most bench organic chemists. Rational molecular design involves the use of all relevant information, such as our understanding of structural, biophysical and physicochemical information, as well as stereoselective drug action [4], biotransformation [5], pharmacokinetics [6] and toxicity [7].

## Statistical Approaches

The first QSAR studies based on Hansch analysis use multiple linear regression for the calculation of QSAR equations. It was indeed his great merit to apply these well-known mathematical methods in pharmaceutical research. However, more complex problems have to be treated differently. The analysis of complex data sets is a problem common to various branches in science. Many other multivariante statistical tools, such as principal component and cluster analysis, appear of benefit to the medicinal chemist. Applications of multivariate statistical strategies in chemistry, including analytical and medicinal chemistry, are referred to as chemometrics [8]. During the last decade new statistical tools, such as partial least squares (PLS) and cross-validation, have been developed and applied to drug research. Some of the important techniques are listed in Table 1. The scope of these multivariate data analysis approaches is summarized under the terms correlation and pattern recognition. Investigation of clustering behaviour in data often gives clues on similarity principles among molecules or uncovers areas which have been unsufficiently explored.

Another concept, which has so far widely been ignored by synthetic organic chemists, is experimental design [10–12]. This approach aims at a proper sampling of the nearly infinite number of possible combinations when e.g. the amino acids in a peptide or substituents on a lead structure are varied.

## Molecular Modelling

The development and implementation of molecular graphics in most pharmaceutical research centers in the 1980s and the integration with computer chemistry introduced an important new tool to the medicinal chemist. For the first he could see the 3D structures of large proteins, allowing the creative mind to design active site-tailored compounds. No doubt, that this was

**Table 1.** Important statistical tools in SPC.

| | |
|---|---|
| Correlation | |
| MLR | Multiple linear regression |
| LDA | Linear discriminant analysis |
| ALS | Adaptive least squares |
| PLS | Partial least squares |
| Pattern Recognition [9] | |
| PCA | Principal component analysis |
| SIMCA | Soft independent modeling of class analogy |
| CA | Cluster analysis |
| NLM | Nonlinear mapping |
| SP | Spectral mapping |
| CSA | Cluster significance analysis |
| Experimental design [10-12] | |
| CP | Craig plot |
| FFD | Fractional factorial design |

0931-8771/92/0206-0200 $ 3.50+.25/0

an important progress. Many reviews on molecular modeling applications have been reported and will not be covered here [13].

Combined QSAR-molecular graphics studies [14] have given us comfort that what can be deduced from the terms in a QSAR study might be consistent with what is seen from the docking of a ligand to the receptor. However, it must be born in mind that chemically related analogs may bind in a slightly or considerably different manner [15]. Considering all molecules in a project in the same QSAR may therefore not be relevant. Similarly, modeling applications of the active analog approach [16] must be done with care.

In the remainder of this paper we will return to the QSAR paradigm and discuss some recent developments.

## QSAR Parameters

The contribution of Hansch and Fujita was extremely important. They showed that physicochemical and other structural properties may be relevant to the pharmacokinetic or pharmacodynamic properties of a bioactive compound. These ideas have been further explored and many molecular or fragmental descriptors or parameters were tested for their potential use in QSAR studies [17]. The intercorrelation among descriptors has been studied [18], showing that broad families of lipophilic, steric and electronic parameters can be defined. Some classification schemes of QSAR parameters are given in Table 2. The meaning of a property descriptor is not always simple, since it has been shown that in fact several parameters are composite ones. A good example are log P values (see below).

## 3D-QSAR

Since only a few of the above mentioned QSAR descriptors encode 3D information, ways have been investigated to include steric and electronic 3D information in a QSAR analysis. The controversial CoMFA (comparative molecular field analysis) method is the result of these efforts [19]. Similar approaches we find in the HASL [20] and GREEN [21] program.

## Understanding Lipophilicity

Ever since the pioneering work of Hansch and Fujita on the lipophilicity of bioactive compounds, this property has fascinated many medicinal chemists [22]. The relationship between activity and lipophilicity has been revisited by Kubinyi, who developed a bilinear model, which in many cases gives indeed a better fit of the data.

Hansch: $\text{activity} = -a(\log P)^2 + b \log P + c$

Kubinyi: $\text{activity} = d \log P - e \log(\beta P + 1) + f$

An explanation for both models was offered by Van de Waterbeemd and coworkers, who could demonstrate that both models can be derived by considering drug transport through a series of water/lipid phase interphases [23]. Further investigations on the role of lipophilicity in drug transport have been reported by several groups [24].

Octanol/water is no longer considered as the only important model system for describing lipophilicity and the interaction with various membrane types. Leahy et al. [25] have proposed to evaluate log P in four systems, namely 1-octanol/water (amphiprotic H-donor and acceptor), chloroform/water (H-donor), PDGP (propylene glycol dipelargonate, H-acceptor), and alkane/water (inert). In several papers it has been demonstrated that the differences in log P values from two systems can give highly valuable information; indeed $\Delta$log P values correlate with e.g. brain [26] and skin penetration [27].

Much information is hidden in log P, probably explaning its success in many QSAR studies. By several lines of evidence, including the solvatochromic equations [28], it has been shown that log P can be decomposed into a bulk or volume term (V) and an electronic term ($\Lambda$).

$$\log P = a.V + \Lambda$$

The $\Lambda$ values can be calculated from the difference with the line defined by nonpolar alkanes [29].

1-Octanol/water: $\log P_{oct} = 0.059\ V + 0.107$

Alkane/water: $\log P_{alk} = 0.069\ V + 1.05$

The $\Lambda$ term includes H-bonding of the solute to the solvent. $\Lambda_{oct}$ correlates with H-bond acceptor basicity ($\beta$) of the solute, while $\Lambda_{alk}$ (calculated from partition coefficients in an alkane/water system) correlates to both H-bond donor acidity ($\alpha$) and acceptor basicity ($\beta$), i.e. the total H-bonding capacity [29]. Thus $\alpha$ and $\beta$, two potentially interesting molecular properties [30], can be obtained when their log $P_{oct}$ and log $P_{alk}$ values have been determined and their molar volumes calculated.

The classical shake flask method developed by Hansch and Fujita has been proven extremely helpful in characterizing the lipophilic properties of molecules. However, a number of disadvantages encouraged the development of new experimental methods. Very successful is the lipophilicity measurement using chromatographic techniques: RP-HPLC and CPC (centrifugal partition chromatography) [31]. Other possibilities to measure lipophilicity values have been explored as well, including the slow-stirring technique [32] and flow-injection extraction [33].

**Table 2.** Classification schemes of SPC parameters.

| Structural | Composition | Origin |
|---|---|---|
| Lipophilic | Pure | Experimental |
| Steric | Composite | Calculated |
| Electronic | | |
| Topologic | | |
| Quantum mechanic | | |

**Table 3.** Programs for log P calculation and hydrophobicity mapping.

| | |
|---|---|
| Fragmental methods | |
| CLOGP | Leo and Hansch (1979) [34] |
| PrologP | Darvas |
| ADAPT | Jurs |
| Atomic contributions | |
| | Broto and Moreau (1984) [35] |
| | Ghose and Crippen (1987, 1989) [36] |
| HINT | Wireko et al. (1991) [37] |
| CHEMICALC2 | Suzuki (1991) [38] |
| MO calculations | |
| CASE | Klopman and Iroff (1981) [39] |
| | Bodor (1989) [40] |
| SELECT | Pearlman |
| Lipophilicity mapping | |
| | Fauchère et al. (1988) [41] |
| | Furet et al. (1988) [42] |
| HINT | Wireko et al. (1991) [37] |

In many applications a quick estimate is required of the lipophilicity of a compound. Computer programs for the calculation of log P values are therefore of use (Table 3). In contrast to fragmental methods, a number of programs emerged based on the summation of atomic contributions. The program CLOGP remains the most elaborate and reliable one [34]. Computer graphics systems allow us to visualize, in analogy to electrostatic potentials, so-called lipophilicity potentials, which discriminate polar and nonpolar surface domains on a molecule [41, 42]. Such pictures may be useful in drug-receptor docking studies.

## Aspects of Modern Drug Research

Two aspects of drug research and development must be distinguished: drug discovery and drug optimization. In drug discovery key roles are played by biochemistry, pharmacology and molecular biology. Gen- and biotechnology are increasingly important for the expression and cloning of receptors and enzymes. A prerequisite for rational and therefore economic pharmaceutical research is the isolation of biological targets and their structure eluicidation by protein crystallography and bio-NMR. However, blind screening also remains a means of lead discovery.

Drug optimization involves several steps. Molecular design can be based on mechanism of action, enzyme structure or known analogs. In the first round affinity or binding to drug targets such as enzymes, receptors or nucleic acids, is optimized. Impressive improvements in synthetical methods and the availability of many ingenious synthons permit to make rather complex and innovative molecules. Intensive interplay between biology and chemistry, supported by e.g. molecular modeling and crystallography are required. Biostructure research has given much insight in drug action at the molecular level. The question which conformation is the active conformation when the molecule is bound to its target is not always easy to answer. In some cases an X-ray structure is nearly identical to a structure in solution determined by NMR and therefore seems a reliable one to build a hypothesis upon. Sometimes the situation is more complex. Indeed, it was recently shown that the NMR structure of cyclosporin A bound to cyclophilin was significantly different from the structures of single crystals and in chloroform solution [43] and yet different from molecular mechanics calculations.

A second and more empirical phase involves the optimization of pharmacokinetic properties, including toxicologic aspects. In this phase physicochemical and biophysical properties come into consideration. Chemistry may then concentrate on e.g. the synthesis of prodrugs.

Among the various computer-assisted methods to design and develop new biologically active entities, databases also have found their own place, including databases for synthesis planning, biotransformation [44, 45] and physicochemical properties. 2D and 3D structure databases are actively being used in the design and discovery process. Examples of 3D databases include the Brookhaven Protein Database and the Cambridge Crystallographic Database. Of several wellknown 2D databases now also a 3D version has been generated, including e.g. the Fine Chemicals Directory and the database of Chemical Abstracts Services. Many companies also have their own in-house structures stored in a 2D and 3D database and several vendors offer the software tools for intelligent searching and model building in these 3D databases.

Quantitative molecular design is often misunderstood and confused with classical Hansch analysis QSAR studies based on linear free energy relationships. The rapid increase in various computer-assisted approaches has contributed to this confusion. First of all very often binding design rather than drug design is done. The actual drug derived from a potent receptor or enzyme binder needs another type of "design", mainly empirical, where pharmacokinetics, including metabolism, are considered.

Taken literally, the acronym QSAR suggests that always some activity is involved in the analysis of molecular properties. In a stimulating paper, Testa and Kier discuss the relation between structure, property and activity [46]. Indeed, there is some semantic confusion about the terms SAR (structure-activity relationship), PAR (property-activity relationship) and SPR (structure-property relationship). We suggest to call all studies aiming at broadening the understanding of putative relationships between molecular intrinsic, chemical and biological properties, SPC (structure property correlation) studies (Figure 1). Thus, SPC studies involve either correlation studies among various types of properties or aim at classification of the molecules by selecting subsets of properties. QSAR studies form a subset of SPC. The scope of SPC studies in the broadest sense is not only the prediction of biological activity, but also to improve our understanding of molecular properties such as aqueous solubility, lipophilicity and even pharmacokinetic parameters. Also molecular similarity studies [47, 48] fall under the scope of SPC. The calculation of the similarity between two molecules has been used as a basis for designing drug screening programs replacing random selection from the available substance database [49].

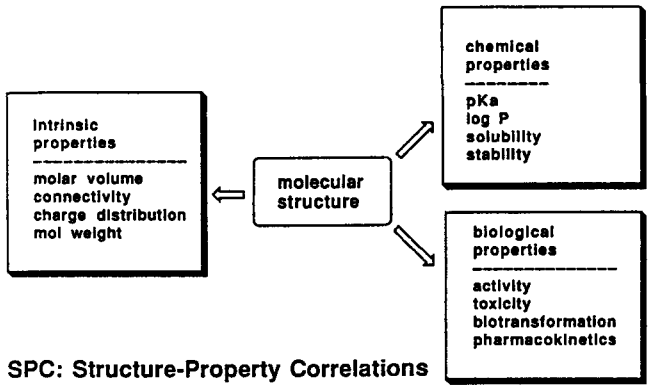


**Figure 1.** The concept of structure-property correlations. Intrinsic properties derive directly from the molecular structure. (Physico) chemical and biological properties emerge from interactions of a molecule with its environment.

A clear trend is emerging that computational chemists, including modelers and SPC experts, become a true collaborative partner, rather than a support function, in multidisciplinary research teams [50]. Since chemistry is an experimental science, also computational chemists do experiments "in computro" by building hypotheses and testing new approaches. Some of them will be rejected, other adjusted or adopted. A close collaboration with other members of a project team is a necessity in order to have a constant challenge of computer-based ideas with experimental facts.

*Outlook*

A further integration between synthetic skills and computer-assisted approaches will further change the live of the medicinal chemist. Workstations bring much computer power and information close to the bench chemist. And even more is to come.

The design and development of bioactive compounds is the result of an active interplay among many different disciplines [51]. In the preclinical phase we mention among others, organic chemists, theoretical medicinal chemists, such as molecular modeler and QSAR specialist, furthermore biologists, biochemists, molecular biologists, (bio)physical chemists, crystallographers and NMR experts. The joined efforts of all these scientists and the progress in each of the fields has made drug discovery to an exciting multidisciplinary endeavour which is indeed different from the time the Hansch approach was introduced.

**Acknowledgement**

The author thanks for critical reading and suggestions Prof. K. Müller, F. Hoffmann-La Roche Ltd. in Basel (Switzerland), Prof. B. Testa, School of Pharmacy in Lausanne (Switzerland), and Dr. D. J. Livingstone, SmithKline Beecham, Welwyn (UK).

**References**

[1] Rekker, R.F., *Quant. Struct.-Act. Relat., 11,* 195 (1992).
[2] Tute, M.S., in *Comprehensive Medicinal Chemistry,* Vol. 4 (Hansch, C., Sammes, P.G. and Taylor, J.B., eds), Pergamon, New York (1990), pp. 1–31.
[3] Hansch, C., Sammes, P.G. and Taylor, J. B. (eds.), *Comprehensive Medicinal Chemistry, Vol. 4, Quantitative Drug Design,* Pergamon, New York (1990).
[4] Ariëns, E.J., *TIPS* 7, 200–205 (1986).
[5] Testa, B. and Mayer, J.M., *Acta Pharm. Jugosl.* 40, 315–350 (1990).
[6] Seydel, J.K. and Schaper, K.-J., *Pharmac. Ther.* 15, 131–182 (1982).
[7] Van de Waterbeemd, H., in Trends in Drug Research, Claassen, V., ed., Elsevier, Amsterdam (1990), pp. 243–252.
[8] Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y. and Kaufman, L., Chemometrics: a textbook, Elsevier, Amsterdam (1988).
[9] Livingstone, D.J., in Molecular Design and Modeling: Concepts and Applications (Langone, J.J., ed.), Methods in Enzymology, 203, Academic Press, London, pp. 615–640 (1991).
[10] Craig, P.N., *J. Med. Chem.* 14, 680–684 (1971).
[11] Austel, V., *Eur. J. Med. Chem.* 17, 9–16, 339–347 (1982).
[12] Hellberg, S., Sjöström, M., Skagerberg, B. and Wold, S., *J. Med. Chem.* 30, 1126–1135 (1987).
[13] Höltje, H.D., *Quant. Struct.-Act. Relat.*, this issue.
[14] Hansch, C., Klein, T., McClarin, J., Langridge, R. and Cornell, N.W., *J. Med. Chem.* 29, 615–620 (1986).
[15] Poulos, T.L. and Howard, A.J., *Biochem.* 26, 8165–8174 (1987).
[16] Marshall, G.R., *Ann. Rev. Pharmacol. Toxicol.* 27, 193–213 (1987).
[17] Van de Waterbeemd, H. and Testa, B., *Adv. Drug Res.* 16, 85–225 (1987).
[18] Van de Waterbeemd, H. El Tayar, N., Carrupt, P.A. and Testa, B., *J. Comput. Aided Mol. Des.* 3, 11–132 (1989).
[19] Cramer, R.D., Patterson, D.E. and Bunce, J.D., *J. Amer. Chem. Soc.* 110, 5959–5967 (1988).
[20] Dowcyko, A.M., in Probing Bioactive Mechanisms, Magee, P.S., Henry, D.R. and Block, J.H., eds. ACS Symposium Series 413 (1989), pp. 82–104.
[21] Kato, Y., Itai, A. and Iitaka, Y., *Tetrahedron* 43, 5229–5236 (1987).
[22] Kubinyi, H., *Progr. Drug Res.* 23, 97–198 (1979).
[23] Van de Waterbeemd, J.Th.M. and Jansen, A.C.A., *Pharmaceut. Weekbl. Sci. Ed.* 3, 71–78 (1981).
[24] Dearden, J.C., in Comprehensive Medicinal Chemistry, Vol. 4 (Hansch, C., Sammes, P.G. and Taylor, J.B., eds.), Pergamon, New York (1990), pp. 375–411.
[25] Leahy, D.E., Taylor, P.J. and Wait, A.R., *Quant. Struct.-Act. Relat.* 8, 17–31 (1989).
[26] Young, R.C., Mitchell, R.C., Brown, T.H., Ganellin, C.R., Griffiths, R., Jones, K.K., Rana, D., Saunders, I.R., Sore, N.E. and Wilks, T.J., *J. Med. Chem.* 31, 656–671 (1988).
[27] El Tayar, N., Tsai, R.-S., Carrupt, P.-A., Testa, B., Hansch, C. and Leo, A., *J. Pharm. Sci.* 80, 744–749 (1991).
[28] El Tayar, N., Tsai, R.-S., Testa, B., Carrupt, P.-A. and Leo, A., *J. Pharm. Sci.* 80, 590–598 (1991).
[29] El Tayar, N., Testa, B. and Carrupt, P.A., *J. Phys. Chem., 96,* 1455–1459 (1992) .
[30] Abraham, M.H., Duce, Ph.P., Prior, D.V., Barratt, D.G., Morris, J.J. and Taylor, P.J., *J. Chem. Soc. Perkin Trans II, 1355–1375 (1989).*
[31] El Tayar, N., Tsai, R.-S., Vallat, P., Altomare, C. and Testa, B., *J. Chromatogr.* 556, 181–194 (1991).
[32] De Bruin, J., Busser, F., Seinen, W. and Hermens, J., *Environ. Toxicol. Chem.* 8, 499–512 (1989).

[33] Kuban, V., *Anal. Chim. Acta* 248, 493–499 (1991).
[34] Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York (1979).
[35] Broto, P., Moreau, G. and Vandycke, C., *Eur. J. Med. Chem.* 19, 66–70 (1984).
[36] Ghose, A.K. and Crippen, G.M., *J. Chem. Inf. Comput. Sci.* 27, 21–35 (1987).
[37] Wireko, F.C., Kellogg, G.E. and Abraham, D.J., *J. Med. Chem.* 34, 758–767 (1991).
[38] Suzuki, T., J. Comput. Aided Mol. Des. 5, 149–166 (1991).
[39] Klopman, G. and Iroff, L.D., *J. Comput. Chem.* 2, 157–160 (1981).
[40] Bodor, N., Gabanyi, Z. and Wong, C.-K., J. Amer. Chem. Soc. 111, 3783–3786 (1989).
[41] Fauchère, J.L., Quarendon, P. and Kätterer, L., *J. Mol. Graph.* 6, 203–206 (1988).
[42] Furet, P., Sele, A. and Cohen, N.C., *J. Mol. Graph.* 6, 182–200 (1988).
[43] Weber, C., Wider, G., von Freyberg, B., Traber, R., Braun, W., Widmer, H. and Wüthrich, K., *Biochemistry* 30, 6563–6574 (1991).
[44] Barcza, S., Kelly, L.A. and Lenz, C.D., *J. Chem. Inf. Comput. Sci.* 30, 243–251 (1990).
[45] Darvas, F., *J. Mol. Graphics* 6, 80–86 (1988).
[46] Testa, B. and Kier, L.B., *Med. Res. Revs.*, 11, 35–48 (1991).
[47] Johnson, M.A. and Maggiora, G.M. (eds.), Concepts and Applications of Molecular Similarity, Wiley, New York (1990).
[48] Zalewski, R.I., Krygowski, T.M. and Shorter, J. (eds.), Similarity Models in Organic Chemistry, Biochemistry and related Fields, Elsevier, Amsterdam (1991).
[49] Johnson, M., Lajiness, M. and Maggiora, G., in: QSAR: Quantitative Structure-Activity Relationships in Drug Design, (Fauchère, J.L., ed.), Alan Liss, New York (1989), pp. 167–171.
[50] Schnyder, J.P., *Med. Res. Revs.* 11, 641–662 (1991).
[51] Spilker, B., Multinational Drug Companies: Issues in Drug Discovery and Development, Raven Press, New York (1989).