



Published in final edited form as:

Curr Top Med Chem. 2010 ; 10(1): 67–83.

Hydrophobicity – Shake Flasks, Protein Folding and Drug Discovery

Aurijit Sarkar and **Glen E. Kellogg***

Department of Medicinal Chemistry & Institute for Structural Biology and Drug Discovery, Virginia Commonwealth University, Richmond, VA 23298-0540, USA.

Abstract

Hydrophobic interactions are some of the most important interactions in nature. They are the primary driving force in a number of phenomena. This is mostly an entropic effect and can account for a number of biophysical events such as protein-protein or protein-ligand binding that are of immense importance in drug design. The earliest studies on this phenomenon can be dated back to the end of the 19th century when Meyer and Overton independently correlated the hydrophobic nature of gases to their anesthetic potency. Since then, significant progress has been made in this realm of science. This review briefly traces the history of hydrophobicity research along with the theoretical estimation of partition coefficients. Finally, the application of hydrophobicity estimation methods in the field of drug design and protein folding is discussed.

Keywords

Hydrophobicity; Drug Design; LogP; HINT; QSAR; Protein Folding; Hydrophobicity Scales; Hydrophobic Interactions

INTRODUCTION

Hydrophobicity (or lipophilicity) is a well-known and extensively studied phenomenon. It is commonly understood to be the tendency of non-polar molecules to form aggregates in order to reduce their surface of contact with polar molecules such as water [1]. Its manifestations include simple observable macroscopic phenomena such as the immiscibility of oil and water or modern techniques such as chromatographic separation. The importance of hydrophobic interactions on atomic or molecular scale has long been recognized in various facets of science [1]. While the concepts have changed and the applications have expanded, the fact remains that hydrophobic interactions are often the driving force in a variety of physical and biological phenomena. In this review, we introduce the concept of hydrophobic interactions, present a very short history and discussion of theoretical and experimental studies on the phenomenon. A key theme of this review is that the same set of forces and interactions that partitions a molecule between polar and hydrophobic solvent phases, i.e., determines its hydrophobicity, is pervasive in all biological interactions including small molecule binding and protein folding. Finally, we present an in-depth perspective on computational studies involving hydrophobic interactions. These studies include methods for estimation of the hydrophobic nature of small and large biological molecules and applications of this in drug discovery or design. Another aim is to suggest future directions in hydrophobicity research, with the hope of stimulating interest in the field.

* Address correspondence to this author at Department of Medicinal Chemistry, Box 980540, Virginia Commonwealth University, Richmond VA, 23298-0540, USA; Tel: +1(804) 828-6452; Fax: +1(804) 827-3664; glen.kellogg@vcu.edu.

A brief historical overview of hydrophobicity

Even before the turn of the 20th century, the importance of hydrophobic interactions in biological phenomena, particularly drug activity, was recognized by the work of Meyer and Overton [2,3,4]. In 1937, Butler showed a linear relationship between heat of hydration and entropy of hydration [5]. He estimated the energies of interaction of different functional groups with water and showed that the heats of hydration are additive in nature. He also explained that the heats of hydration do not determine the free energy of interactions, but that there is a direct proportionality between them. The origins of this phenomena were unclear at this point, but it was hypothesized that entropy might be dependent on the size of the “cavity” that contains the molecule. The importance of H-bonds was also briefly discussed as formation of H-bonds between polar parts of the molecule causes an increase in entropy, which favors dissolution of an otherwise non-polar molecule.

Frank and Evans, in the middle of the 20th century, described the formation of “icebergs” of water around non-polar parts of molecules [6]. Their findings were based on the deviation of entropy of vaporization for certain substances when dissolved in aqueous and non-aqueous solutions. The formation of a regularized lattice-like structure of water molecules surrounding non-polar moieties has been experimentally validated with crystallography [7] and is now more or less taken for granted. This theory was extended to proteins by Klotz, who explained the variation in pK_a, molecular volume, denaturation and the masking of expected behavior of protein functional groups in terms of this “iceberg” formation [8]. In fact, the association of two molecules can become energetically favorable due to the increase in entropy when these ordered water molecules are scattered or disordered (see Figure 1).

Kauzmann first coined the term “hydrophobic bond” in 1959, which caught the attention of many scientists at the time; this notion was supported by a number of research investigations of that era [9]. The work of Némethy, Scheraga and Steinberg also supported the use of this term [10]. Perhaps it was the tendency of the non-polar substances to form aggregates that caused scientists to draw parallels with the definition of a bond – “the tendency of two atoms to stay together in space”. Hydrophobic bonds have been described as endothermic, i.e., as temperature increases their strength increases until a maximum value is reached at approximately 60°C [10]. However, the stability of proteins depends on not only these hydrophobic “bonds,” but also hydrogen bonds. These have an inverse behavior, i.e., they become *weaker* with increasing temperature. Thus, as temperature increases, hydrophobic interactions would decrease in strength and hydrogen bonds would be predominant in maintaining protein structure. Hence, at higher temperatures proteins would unfold.

In the 1970s, Robert Hermann published a series of three papers on “the theory of hydrophobic bonding” [11-13] where the large negative entropy of partitioning a hydrophobic molecule into a non-polar solvent was explained by the loss of order in water molecules in direct contact with the hydrophobic surface. The ordered arrangement of water molecules on the surface of a molecule is due to dipole-dipole interactions with the immediate next layer of waters. In effect, this phenomenon is similar to surface-tension where the first layer arranges itself in order to reduce contact with the hydrophobic air, while less order exists in the second and succeeding layers. Order continues to decrease in layers away from the hydrophobic surface and there is a linear, but inverse, correlation between hydrophobic surface area and its solubility in water [11]. Hermann also determined that the free energy for hydration of a hydrophobic molecule is linearly related to the number of water molecules that can be packed around it. This first study did not take into account cavity curvature and was restricted to small molecules. Later work [12] described a correlation between a molecule’s hydrophobic surface area and its solubility in water. Hermann also introduced a simple mathematical treatment of hydrophobic interactions [13] that addresses hydrophobic interactions at a distance taking into account not only solubility, but also the distances between hydrophobic entities with the Lennard-Jones

potential as has also been suggested by Reynolds *et al* [14]. Leo, Hansch and Jow established a relationship between hydrophobicity and two other factors – the nature of the solute surface and the molecular (CPK) volume [15]. The major innovation of this study is that they used the partition coefficient for 1-octanol/water ($\text{LogP}_{o/w}$) as a measure of hydrophobicity rather than solubility. This parameter has been used almost ubiquitously in studies thereafter. Most importantly, these observations could not be explained by the simple concept of a “hydrophobic bond”, but rather as a complex phenomenon involving the interplay of flexible molecules and solvent under particular conditions.

The argument on semantics over the use of the term “hydrophobic bond” has continued ever since, but the fact that hydrophobic phenomena can explain a multitude of observations in science cannot be ignored. In this review, we attempt to explain how naming and characterizing this effect has changed the realms of computational chemistry and drug design. The preceding paragraphs provide an outline of the sequence of events and were not intended to serve as a comprehensive guide to the principles behind hydrophobic phenomena. However, a comprehensive review of the research on hydrophobicity is available elsewhere [1].

CALCULATIONS OF HYDROPHOBICITY AND THE HYDROPHOBIC EFFECT

Estimation of $\text{LogP}_{o/w}$

Hansch and Leo published their seminal paper on the determination and uses of partition coefficients in 1971 [16]. This paper was, and perhaps continues to be, the most comprehensive article on the subject. It explains the fundamentals of partition phenomena and provides detailed descriptions of the history and theory of the same. It also contains a very comprehensive tabulation of LogP values for various substances. However, most interesting to theoreticians is the discussion on additive-constitutive properties wherein the utilization of the Hammett equation in calculations of partitioning free energy and the effects of various stereoelectronic effects on the partition coefficient are described. Also, various uses of partition coefficients for such diverse research topics as countercurrent distribution, measurement of equilibria, hydrophile-lipophile balance, drug dissolution and “hydrophobic bonding ability” are outlined. Of note, the partitioning of alcohols between water and red cells was compared to their partitioning between water and 1-octanol. The energy of partitioning per methylene group was the same for both cases, i.e., approximately $-690 \text{ cal mol}^{-1}$. The repercussions of this quantification of hydrophobic interaction energies have been key to drug design projects as well as computational chemistry. The Hansch and Leo method for theoretical estimation of molecular LogP values, which is the basis of the C-LOGP method (*vide infra*), is also described in great detail.

A loose categorization of different methodologies for estimation of LogP is provided in Table 1, complete with a few typical examples of each. Here, the discussion on these methods will be limited to a general overview highlighting the application of these methods in drug design and the relevance of accuracy for these prediction methods in that context. Several comprehensive reviews on the computational estimation of octanol-water partition coefficients are available [17-21]. These articles are highly recommended for those who intend to learn about the estimation methods in detail or to understand the nuances of their relative accuracies for particular applications.

To commence, it is a monumental understatement to say that a lot of good research has been done in this field. Many diverse empirical methods exist today that predict LogP of various molecules with different degrees of context-dependent certainty [17,20]. Some of the major types are discussed below.

Fragment-based methods—Rekker's fragment based system was the first fragment-based computational method to estimate LogP [22]. Fragment-based methods implement and statistically deconvolve empirical data from experimental LogP values of compounds. See Appendix 1 for a short explanation of this approach. In order to explain the effect of inter-fragmental interactions, certain additive correction factors are introduced. Several other algorithms of this type exist including the C-LOGP [23-27] and ACD/LogP [28] methods. The criticism most often applied to this methodology is that the fragmentation of the target molecule is “arbitrary”. This is not actually true for C-LOGP as there are a complete and unambiguous set of rules. However, they can be difficult to visualize, and fragments can be much more complex than organic functional groups. Thus, fragments observed in new molecules can be missing from the C-LOGP database library, yielding poor predictions of LogP [17,20]. However, there are also advantages to these methods: significant and complex electronic interactions are automatically taken into account when they exist within a library fragment [20]; when the fragments coincide with real organic functional groups their interpretation is intuitive; the correction factors can be used to understand the relationship between functional groups or the effect of the observed feature on solubility, e.g., factors representing aliphatic chain branching explain the increased water solubility of branched hydrocarbons; and since fragment methods are based on empirical data, their associated algorithms are very fast and practical to implement in software.

Atom-based methods—These are similar to the fragment-based methods, but assume the hydrophobicity of a molecule to be the sum of the individual atomic contributions. Appendix 2 provides an overview of the principle behind atom-based additive methodologies. Again, several methodologies of this type exist, including the well known Ghose-Crippen [29-32] and XLOGP [33-35] methods. Mostly these algorithms avoid correction factors by taking into account these sorts of contributions with a large set of atom types according to the individual environment it exists in within the molecule [20]. In order to somewhat reduce the atom type set the XLOGP algorithm implements a small number of correction factors. The reduced dependence on corrections is the major advantage of these methods. As described by Buchwald and Bodor, the major disadvantage of this method is that often, the molecule is “more than a sum of its parts” [17]. Furthermore, human interpretability is reduced as the size of the atom database set grows and the correspondence with organic and medicinal chemistry principles is lost.

Molecular methods—Over the last two decades quantum mechanical calculations have been increasingly used in applied research including drug discovery, particularly with respect to estimations of interactions between solute and solvent molecules. A number of studies have used quantum chemical principles for estimation of molecular hydrophobicity [17]. Early work includes that of Rogers and Cammarata [36,37] and also that of Hopfinger and Battershell [38]. Klopman and Iroff used charge densities to calculate partition coefficients [39]. More recently, Bravi and Wikel described a method to predict LogP using a technology called Molecular Surface – Weighted Holistic Invariant Parameters (MS-WHIP) [40]. Unfortunately, a relatively large standard deviation between predicted and actual LogP was observed [18]. Toulmin *et al.* described another prediction method for octanol/water partition coefficients [41] that correlated minimized molecular electrostatic potentials with the H-bonding capability of molecules. In this method ΔLogP is defined as the difference between LogP_{oct} (logarithm of the 1-octanol/water partition coefficient) and their predicted LogP_{hxd} (logarithm of the hexadecane/water partition coefficient). H-bonding capability has a profound effect on partition coefficients with a strong correlation between ΔLogP and V_{min} (minimized molecular electrostatic potential). A strong correlation was also reported between ΔLogP or LogP_{hxd} and CNS penetration of compounds, i.e., through the blood-brain barrier. This highlights the importance of H-bond donors and acceptors in normal partitioning phenomena.

Livingstone *et al.* described a method that uses neural networks (NN) to predict LogP values from a training set of electrotopological descriptors [42] of 900 drug and pesticide-like compounds [43]. Other studies involving artificial-intelligence utilize parameters calculated by various methods in unsupervised-learning processes to develop predictive models [44,45]. Taskinen and Yliruuski provide an in-depth analysis of such models in their review on NN modeling [46]. They note that while NN methods are accurate in predicting LogP values of molecules within the size, functional group, etc. confines of the training set, they are less accurate in predictions for molecules outside the training set. However, this is the case for all LogP estimation methods.

Hydrophobicity of amino acids and proteins—Understanding the hydrophobic behavior of amino acids, peptides and proteins has implications far beyond the seemingly simple task of calculating LogP for twenty or so small molecules (the amino acids). Abraham and Leo extended the Hansch and Leo fragment-based method of LogP calculation to amino acid zwitterions and side-chains [47]. Excellent agreement was reported for 19 out of the 20 natural amino acids. Proline, however, was calculated to be more hydrophilic than in reality, probably due to poor fragment parameterization for its secondary cyclic amine. With this method, hydrophobicity values for amino acid side chains were best predicted if a field effect was applied to the alpha-carbon. The field effect is the sum total of polar proximity effects of both the backbone amidic (peptide) bonds surrounding the alpha-carbon atom of any given residue. This field effect parameter accounts for the charge distribution on side-chain atoms and hence directly affects the hydrophobic/hydrophilic nature of the residue. Application of this effect allowed a higher correlation between predicted and calculated values of hydrophobicity for side-chains. In additional studies, Buchwald and Bodor reported a correlation between the van der Waal's volume of peptides with their LogP values [48]. Another approach was adopted by Steinmetz, where 3-D QSAR Comparative Molecular Field Analysis (CoMFA) studies were applied in a similar manner [49]. Experimentally determined LogP values of free and blocked di- and tripeptides were analyzed statistically to produce another set of parameters [50]. Akamatsu's work on the solvent partitioning of peptides using regression analysis of the experimental data to abstract the hydrophobic parameters [51-54] is commonly regarded as the most convincing and accurate [55]. A comparison between software programs in predicting peptide LogPs was published recently [56]. In general, fragment-based methods are sensitive to composition but not to peptide sequence, which can be considered to be a major flaw of these programs. Also, it is important to note that most current programs are inefficient and ineffective in calculating LogP values for long peptides.

Summary of LogP estimation methods—A lot of effort has gone into devising methods for high prediction accuracy for LogP. However, most methods are accurate for members or close relatives of their own training sets but continue to be less accurate outside their training sets. It should also be pointed out that a considerable portion of the predictive inaccuracy may, in fact, lie with the data itself. Such data has often been obtained with experimental procedures whose accuracy varies with the method used [57]. One example brings this into focus: as many drugs and drug-like molecules contain ionizable functional groups, the conditions of measurement, particularly pH, are extremely relevant to measured LogP. Thus, if a user attempts to estimate LogP for a molecule, e.g., by specifying a carboxylic acid-containing species, what LogP value should be reported? The molecule in its acid form? The molecule in its ionized (nominally pH 7) form? Or the weighted average representing the equilibrium between the two forms? (This is what the experiment, as performed on the molecules in the training set, measured.) As other functional groups on the molecule can shift that equilibrium, how does this affect the contribution of the carboxylic acid/carboxylate fragment (or constituent atoms) to the predicted LogP?

While whole-molecule approaches are designed to estimate LogP values with great accuracy without extensive piecemeal (atom- or fragment-wise) empirical parameterization, their predictive nature in the end must also be compared to experimental data, limiting our ability to really judge the accuracy of predictions. This begs the question: do we need to emphasize accuracy of predictions so much? We suggest that when it comes to drug design, it is largely the Δ LogP changes between analogues that will drive the evolution in design with respect to physicochemical properties of the molecule. Virtually all methods of estimating LogP can accurately describe the replacement of a proton by a hydroxyl, the halogenation of an aromatic ring, substitution of an amine for a methyl, or nearly any of the chemical modifications that would be performed in fine-tuning a lead compound. The prediction of LogP for random organic compounds is probably not an important real world exercise. Also, representing such an important physicochemical property as a simple scalar value underutilizes the information content of the molecule's 3-D topology and, particularly, its hydrophobic structure. The combination of topology and hydrophobicity provides us with structural details of immense importance, which play a direct role in intermolecular interactions, e.g., ligand binding, protein-protein associations, etc. However, we do recognize the importance of LogP in QSAR studies, and also in assessing the drug likeness of a compound, both of which will be discussed below.

Can predictive methods for estimating the LogP of a peptide translate into a meaningful number for protein hydrophobicity? The idea that an additive atom-based or fragment-based algorithm (or even a whole molecule approach) could describe the dissolution of a protein into water and/or 1-octanol is probably preposterous. To start, it is likely that a severe conformational change would occur if macromolecules pass from aqueous to organic solvents, e.g., hydrophobic residues would rearrange to the surface while the hydrophilic ones attempt to optimize hydrogen bonding and/or electrostatic interactions at the core. In other words, a protein would be an entirely different chemical species when interacting with solvents of different polarity – if it could actually be solubilized. However, the atomic, fragment or residue-level components of such a total LogP should be useful descriptors for understanding the forces and energetics of protein secondary, tertiary and quaternary structure, and have been used in various schemes of describing and predicting protein folding for more than 20 years.

HYDROPHOBICITY SCALES AND PROTEIN FOLDING

There has long been evidence that protein secondary structure is dependent on the hydrophobic properties of the amino acid residue side chains. There is, in fact, a reproducible pattern of these properties in well-defined secondary structural elements such as α helices and β sheets. Thus, considerable effort has been expended in developing hydrophobicity scales that can aid in predictions of protein folding patterns. Some of these scales are based on water-ethanol transfer free energies [58,59], while others are based on partitioning between the bulk aqueous phase and the air-water interface [60], or on water-vapor partition free energies [61]. Kyte and Doolittle discussed the weaknesses of all three of these in a paper that also introduced their own hydrophobicity scale [62]. In their view, water-ethanol transfer free energy-based methods suffer because some amino acids are known to be insoluble in both water and ethanol and the latter may not be a truly inert solvent. Using partition data from transfer between the aqueous phase and air-water interfaces was also problematical because the hydrogen bonds that must be broken and the charges that must be neutralized to remove a residue from the aqueous phase during the formation of the native structure probably remain unchanged at an air-water interface. Thus, they would not be a factor in the overall reaction.

The “hydrophobicity” parameter of Kyte and Doolittle [62] is an amalgam of water-vapor transfer free energies and the interior-exterior distribution of amino acid side-chains determined by Chothia [63]. A moving-segment approach that continuously determines the average hydrophobicity while it advances through a sequence is used to obtain a plot of hydrophobicity as a

function of sequence. On this plot, any parts of the sequence that are above the average hydrophathy for the sequence are termed hydrophobic and correspond well with experimentally determined “internal” regions of proteins. Conversely, sequence elements with hydrophathy below the average are termed hydrophilic and correspond well with areas of the protein that are “exterior” and likely to be in contact with the polar solvent. The motivation is that analysis of these data may indicate the “folding” pattern of the sequence. To further exploit this, Wimley and White reported a new forcefield derived from partitioning two series of model peptides into the interface of neutral (zwitterionic) phospholipid membranes [64]. An alternative approach was introduced in 1986 by Eisenberg and McLachlan [65] for calculating the stability of protein structures in water based on atomic coordinates. The contribution of each protein atom to the solvation free energy is estimated as the product of the solvent accessibility of the atom and an atomic solvation parameter.

Li and Deber [66] used circular dichroism (CD) data to rank order helical propensity of proteins within membranes. Residues such as Ile, Val and Thr, which usually exist as β -sheets in an aqueous environment, prefer an α -helical conformation in lipid membranes. Thus, the helical propensity of amino acid residues correlates with the hydrophobic nature of the side chain. More recently, Dyson, Wright and Scheraga have explained [67] how strict classification of side chains as polar or non-polar has obscured certain facts about protein folding. For example, methylene groups present in large polar or charged amino acid side chains like the four methylenes in the lysine side chain can be considered non-polar. Interestingly, this fact was imbedded as one of the factors in the Hansch and Leo system for estimating LogP [68] nearly 30 years earlier! These methylenes can aggregate with other non-polar groups and assist in hydrophobic collapse of the sequence.

Felitsky *et al.* introduced the use of a new parameter called “average area buried upon folding” (AABUF) [69] that explains both local contacts and long-range interactions. AABUF was used to study folding of apomyoglobin and provided additional insight into hydrophobic collapse and early folding events. Studies on polyalanine and polyleucine helices in water by MacCallum *et al.* [70] confirmed that in folding many unfavorable enthalpic events are counterbalanced by favorable entropic contributions by the solvent. This indicates a very small free energy barrier for folding. Thus, folding is mainly a desolvation phenomenon. Similarly, the Mardia and Nyirongo procedure for generating virtual protein C_{α} traces simulates the hydrophobic effect during folding [71] and produces models that are globular and compact.

Another related application of hydrophobicity is in the development of algorithms to simulate folding of hydrophobic-polar (HP) models in 2 and 3 dimensions [72]. The concept is to simplify the complex problem of folding by reducing it to representing residues by spheres with H (hydrophobic) and P (polar) character. The ensuing simulations are based on the observation that hydrophobic forces are the major forces determining native conformation of small globular proteins. These model simulations have been used to develop mathematical strategies for solving the combinatorial explosion problem, rather than actually simulating the hydrophobic effect [73-74].

As these studies have progressed over the past 20 years or so, the understanding of the hydrophobic effect and its impact on protein structure has matured. The early emphasis of using hydrophobicity scales to define folding patterns has shifted to algorithms that define protein folding in terms of mathematical approaches to reduce the calculational combinatorial explosion caused by exhaustive sampling of conformational space. However, it must be repeated that the same forces and energetics that drive solvent partitioning in the shake flask are at the core of protein folding. The difficulty is to unravel them and define algorithms that can simulate folding in these terms.

LogP IN DRUG DESIGN

Small molecule hydrophobicity has long been a consideration in drug discovery and design. The relationship between anesthetic effect of certain gases and their hydrophobicity has been extremely well established [4,75,76]. As described by Meyer in 1937 [3], chemically inert substances accumulate in “lipoids” and at a certain concentration, produce narcosis. The concentration itself is dependent on the animal, but independent of the narcotic itself. Hansch *et al.* also confirmed the Meyer-Overton hypothesis about a direct relation between hydrophobic nature of a compound and its anesthetic capabilities [75] through statistical correlations. However, Hansch suggested the additional involvement of a polar factor because molecules with polar hydrogens showed greater anesthetic action. Hansch *et al.* also introduced a similar theory for the hypnotic effect of barbiturates [76]. Other studies also have shown the important correlation of partition coefficients with binding affinities of drugs to receptors [77,78]. McFarland used a very simple probabilistic treatment of drug diffusion from the site of administration to the site of action via a collection of hydrophobic and hydrophilic barriers [79]. Inclusion of the Hammett equation into this study gave an intuitively satisfying parabolic relationship between drug potency and hydrophobicity (Appendix 3): higher doses of drugs with unfavorable partition coefficients (either too high or else too low) are required for them to reach the site of action. Recently, Kier has proposed a general theory of inhaled anesthetics [80].

The Lipinski “Rule of 5”

Hydrophobicity, of course, has also been a key factor in Lipinski's “rule of 5” [81,82]. In simple terms, Lipinski's rule can be stated as such: Poor absorption or permeation is more likely for a chemical entity when: a) there are more than 5 H-bond donors (sum of OHs and NHs); b) the molecular weight is over 500; c) the LogP is over 5; or d) there are more than 10 H-bond acceptors (sum of Ns and Os). The only exceptions to these rules were said to be substrates for biological transporters and natural products, which have a tendency to be highly complex molecules with multiple stereogenic centers and rarely contain nitrogen [83,84]. In 2000 Lipinski introduced changes to address terms such as ‘drug-like’ because it was predicted that ADME (Absorption, Distribution, Metabolism and Excretion) screening of molecules (into drug-like or non-drug-like) would precede screening for activity at biological receptors [82]. The rule of 5 was further extended [85] to define a number of useful parameters: a) the presence of greater than 10 rotatable bonds reduces oral bioavailability; b) $0 < \text{LogD} < 3$ enhances the probability of good intestinal permeability (LogD is logarithm of the distribution coefficient D, which is in turn defined as the ratio of the sum of concentrations of all forms of a substance distributed between two mutually immiscible phases); c) a polar surface area (PSA) of less than 60-70 describes CNS active compounds; d) an N+O count of less than or equal to 5 enhances the probability of passing the blood-brain barrier; e) if $\text{LogP} - (N + O) > 0$, the molecule tends to be CNS active; f) orally-active drugs have lower molecular weight and fewer H-bond donors, acceptors and rotatable bonds; g) pulmonary drugs tend to have a larger PSA; and h) if the molecular weight < 300 , $\text{LogP} < 3$, H-bond donors and acceptors < 3 and rotatable bonds < 3 , the compound can be called “lead-like”.

This revolutionary work, which brilliantly summarized over 100 years of Medicinal Chemistry trial and error, made possible a number of rational filters and screens that, in principle, would improve the likelihood that a compound with promising “activity” could produce a “lead” and eventually yield a “drug”. Muegge described various methods for classification of drug-like compounds in his 2003 publication [86]. Similar publications addressing the terms ‘drug-like’ and ‘tool-like’ were also made [87,88]. Oprea *et al.* reported the presence of a “medicinal chemistry lead-like space” and urged careful use of Lipinski's rules [89]. A very interesting discussion [82] on how the properties of drug candidates from two pharmaceutical companies have varied across time pointed out that stress on rational methods of drug design in Merck

laboratories caused no significant change in MLogP (Moriguchi LogP [90]) values across time. In contrast, there was a measurable increase in MLogP values for candidates from Pfizer since almost 50% of their hits were discovered with high-throughput screening (HTS) methods. Because the easiest method to increase in-vitro potency is to appropriately position a hydrophobic moiety onto a lead compound, HTS methods almost invariably select more hydrophobic candidates. Similar trends were observed [91] in that more than half of the molecules reported to have high-activity towards the end of the last century had a high LogP (> 4.25), high molecular weight (>425) and log of solubility in its neutral state (estimated from its molecular weight and LogP values), i.e., $\text{LogS}_w (< -4.25)$, only about 35% of the true lead compounds had these properties. It was also noted that as these molecules go through clinical trials, there is a distinct decrease in LogP values for compounds that make it to the market. One thing is clear from these studies and an analysis by Proudfoot of drugs currently on the market [92]: the lipophilicity of molecules that make it all the way to commercialization has remained in the same range for a number of years. In other words, there is a delicate balance between the hydrophobic and hydrophilic nature of a molecule that is absolutely essential for it to be transported to the site of action by diffusion across membranes.

Hydrophobicity in QSAR

Similarity between molecules is often perceived by chemists both qualitatively and quantitatively. A synthetic chemist would describe two molecules as similar if they have similar topologies, bond connectivities, functional groups or maybe synthetic strategies. Structure-Activity Relationships (SARs) are based on such comparisons in the context of physiological function, but are mostly limited to qualitative or semi-quantitative treatments of biological phenomena or activities. However, more stringent definitions of similarity have been formulated and can be used with chemical computing software to perceive (and even predict) chemical equivalence provided the likeness is scrutinized critically. Thus, a more mathematical and quantitative approach called the Quantitative Structure-Activity Relationship (QSAR), wherein affinities of ligands for their binding sites, inhibition constants, rate constants and other biological activities are correlated to molecular properties such as lipophilicity, polarizability, electronic and steric properties, was developed. Comprehensive reviews have been published on the subject in the past [93,94], which should be referred to by those wishing to learn about the QSAR concept in depth. Here, we will focus on the key role of hydrophobicity in these studies.

There are many different approaches used in classical QSAR studies, including establishment of relationships between activity and physicochemical properties such as steric properties (Hansch analysis, extrathermodynamic approach), structural features (Free Wilson analysis) [94], or topological descriptors (Kier-Hall indices) [42]. 3D QSAR methods, especially those such as CoMFA, consider three-dimensional ligand structures and use those to propose the binding modes of those ligands at a common protein active site [94]. Data is often analyzed by statistical methods such as Multiple Linear Regression (MLR), Partial Least Squares (PLS) or by use of artificial intelligence (AI) methods such as Neural Networks (NN) or Support Vector Machines (SVM) [95] in order to detect correlations between a target activity and various descriptors (like LogP).

Hansch and Fujita first introduced their method, and coined the term QSAR, for correlation of biological activity to chemical structure in the 1960s [78,96-98]. The method correlated, with the use of regression analysis, ligand structural variations to the biological activities of those ligands. In time, these studies would become a distinct scientific field and a mainstay of drug discovery and design research. Many applications have been reported across the past five decades. A review by Kubinyi has described, in great detail, the various subtleties of the science [99]. Indeed, in the absence of a detailed target or receptor structure, this ligand-based drug

design method gives invaluable quantitative information to drug designers. It is important to note that the first publication on QSAR in 1962 [78] showed the importance of hydrophobicity through LogP. Appendix 4 explains the general concept behind the Hansch Analysis technique where the free energy-based substituent constant π is based on the Hammett function σ . π is dependent on the substituent's chemical nature and, since molecules must repeatedly partition between lipid membranes to be effective drugs, the constituting fragments of π should be such that their additive effect would allow easy partitioning into either membranous or aqueous phases.

It must be noted, however, that hydrophobicity is not always the principal parameter determining activity [100]. For example, when DNA is the drug target, e.g., in binding to the major or minor groove, QSAR analyses often shows negligible hydrophobic terms because the negatively charged phosphate groups of DNA are hydrophilic. On the other hand, DNA intercalation would likely be a hydrophobic effect. Radical reactions also typically lack hydrophobic terms in QSAR analyses, although these studies are mostly on small datasets and more thorough studies would be desirable. Finally, it has been suggested from QSAR studies on Multiple Drug Resistance that this process might be accomplished without hydrophobic assistance, although this conflicts with the fact that efficiency of efflux pumps is often correlated with the hydrophobicity of their substrates [101].

3D QSAR methods like Comparative Molecular Field Analysis (CoMFA) [102] generate 3D field maps around aligned molecules to display zones of steric, electronic and lipophilic tolerance or intolerance. This gives a visual understanding of biological activity that contrasts well with the often messy collection of molecular descriptors in classical QSAR studies, thereby allowing easier interpretation of results. This, in turn, may lead to a better basis for designing novel scaffolds and/or chemical substituents to the existing scaffold. The basic idea behind this method is explained in Figure 2. Kellogg *et al.* introduced a method for hydrophobic field calculation for CoMFA [103] using an empirical force field (Hydrophobic INTeractions or HINT, discussed in detail below). This was one of the first attempts in 3D QSAR to modify the purely enthalpic treatment of ligand-receptor binding by inclusion of an implicit entropic term. References to the use of HINT-CoMFA in drug design are available [103-106]. Another attempt to include hydrophobicity into CoMFA was made by Gaillard, Testa and coworkers in their papers [107,108] describing the use Molecular Lipophilicity Potential (MLP) in 3D QSAR along with its applications. This alternative method of using hydrophobicity in CoMFA studies has found a number of applications in molecular modeling and drug design work [109-111].

QUANTIFICATION OF HYDROPHOBIC INTERACTIONS

Equations calculating energy from structure, a.k.a. force fields, have been in use for many years in computational chemistry and molecular modeling [112]. Generally, force fields have been restricted to enthalpic terms that are simple to correlate with bond formation or bond breaking and simple Newtonian physical phenomena like bond stretches and bends, electrostatics and dispersion. The hydrophobic effect is, in some measure, an entropic phenomenon, and is not easily derivable from these first principles. Nevertheless, a few examples of quantifying lipophilicity and its effect on biomolecular energetics have been reported [12,113-119]. Hermann and Chothia [12,113], among others, proposed that hydrophobicity can be quantified by the calculation of hydrophobic surface area. Oobatake and Ooi present an excellent review of this approach [114]. Cramer and Truhlar introduced a solvation model [115] that included charge distributions on solute molecules, the energetic effects of cavity formation and restructuring of water around such cavities, and even subtle variations in charge distribution due to interactions between solute particles and surrounding solvent molecules. Sharp and coworkers introduced a new solvation model illustrating the dependence of the hydrophobic

effect on curvature of the site [116]. This was an attempt to explain the difference between the calculated energy for hydration of hydrocarbons (about $25 \text{ cal mol}^{-1} \text{ \AA}^{-2}$) and the surface tension at the water-hydrocarbon interface (about $75 \text{ cal mol}^{-1} \text{ \AA}^{-2}$). This altered surface area measurement suggested that the “macroscopic” hydration energy is $47 \text{ cal mol}^{-1} \text{ \AA}^{-2}$. Indeed, the assumption that the energy of hydrophobic interactions is dependent on the area of the hydrophobic-water interface is the mainstay of much research in the area. However, alternative approaches have had some success. Cesari *et al.* presented a model describing the hydrophobic interactions within globular proteins based on analysis of X-ray data [117] where fold definitions were clearly shown to be a function of hydrophobicity. Hummer described the development of a hydrophobic force field as an alternative to surface-area models [118]. The best developed model for quantitating hydrophobic interactions has been the HINT (Hydropathic INTeractions) system that is discussed below.

The HINT paradigm

A notably different approach was taken by Kellogg and Abraham [119,120] in designing the “natural” force field HINT (see Appendix 5). This non-covalent interaction force field is derived from partition coefficients based on the Hansch and Leo LogP estimation method. It is very empirical in nature and approximates all components of biomolecular interactions, including hydrogen bonding, Coulombic interactions along with entropy and solvation/desolvation effects in addition to hydrophobic interactions because all of these effects are inherent in the experiments that measure LogP [68]. Interestingly, the Hansch and Leo method encodes many interaction effects within the “correction” factors. For example, intramolecular hydrogen bonding within a small molecule, which would make the molecule less polar (and seemingly more hydrophobic) because the involved polar hydrogen and its partner acceptor are less able to interact with water solvent, is encoded with a factor that gives an internally calibrated indication of the energetics of hydrogen bonding (0.6 – 1.0 LogP units, i.e., 0.8 – 1.4 kcal mol^{-1}).

A key principle behind HINT is that significant understanding of biological phenomena, particularly interactions, can be revealed by representing hydrophobicity as a 3D “field” property rather than as a simple scalar (number) [121]. For example, consider the Tyrosine molecule shown in figure 3. Hydropathic properties of the molecule are mapped in three dimensions around the structure of Tyrosine, creating a HINT map. The use of these maps creates a visual representation of properties that are often mentioned casually, such as hydrophobic or polar nature of functional groups. Not only does this methodology allow a chemist to form a qualitative understanding of the molecular topology, but also forms the basis for quantitative estimation of physicochemical properties by using the HINT force field. This may further be used in the depiction of molecular interactions, which has a direct repercussion in drug design.

Intermolecular Interaction Analysis—Perhaps the most important application of HINT is in the assessment of intermolecular interactions. HINT calculations derive an interaction score that in numerous studies [122-125] has been shown to correlate with free energy of interaction. Although it is data set dependent, i.e., for specific protein-ligand or polynucleotide-ligand systems, it is estimated that, on average, 515 HINT score units correspond to 1 kcal mol^{-1} free energy of binding [119]. A recent report indicated the value of the HINT score in ligand docking studies by a comparison to the scoring functions within FlexX, AutoDock and GOLD [123]. The most important advantage of the HINT methodology is that it inherently estimates enthalpic as well as entropic contributions to binding (Appendix 5). It has been shown [122,126] that errors in prediction for very diverse sets of protein-ligand complexes are approximately $\pm 2.6 \text{ kcal mol}^{-1}$, although within a family of ligands binding to the same protein this error can approach $\pm 1 \text{ kcal mol}^{-1}$. There often is an order of magnitude difference between

values of K_i measured by different laboratories on the same protein-ligand complex, which corresponds to a possible 1.0-1.5 kcal mol⁻¹ experimental uncertainty. Thus, the error value reported above between experimental results and HINT scores indicates that HINT is a robust method for binding affinity predictions. Further sources of error include uncertainties in positions of atoms in models, incorrectly assigned atom types, or (often) missing solvent molecules in the source crystallographic structure data. The HINT method has been used successfully in quite a number of projects [122,126-129]. In a recent example, Tripathi *et al.* generated a model capable of predicting antiproliferative activity of pyrrole derivatives against cancer cell lines. In this study, experimentally determined IC₅₀s of a number of compounds were correlated with HINT scores from docking these ligands to $\alpha\beta$ -tubulin to generate molecular models that could be scored, and yielded a significant correlation. This correlation could distinguish active molecules from inactive ones by the HINT score value and, thus, provides a basis for design of novel molecules with anticancer activity. In another interesting application the sequence specificity of anthracycline groove-binding intercalators was evaluated and predicted by HINT score [130]. This work illustrated that the HINT score could be parsed into relevant free energy subsets that can be ranked and compared for particular intercalator functional groups and/or nucleotide bases in DNA double helix strands.

Computational Titration—An extension of the HINT force field known as Computational Titration [124] is used to evaluate the ionization states of functional groups on ligands or residues at the binding site. It is well known that these variations can have a strong influence on binding affinities. The method models, in parallel, multiple ionization states for both ligand and protein creating a collection of ionization state ensembles. Each distinct protonation state ensemble is optimized for hydrogen bonding, including water positions, and analyzed by HINT score. The best scoring complex indicates the optimum state for binding and suggests the corresponding pH for that optimum binding. However, the pH at which crystals are grown and analyzed can be different from this optimum pH. The resulting model can help reconcile the differences between *in silico* models and data. However, at room temperature, where binding data is measured, there are likely to be many protonation models of similar, accessible energy. Computational Titration analysis helps develop an understanding of the relationship between these states. There is now a computational titration server for public use at <http://hinttools.isbdd.vcu.edu/CT> [131].

Analysis of Bridging Waters—Another factor relating to the stability of biomolecular complexes is the contribution of water molecules within the binding site and bridging between the ligand and biomolecule [125,132]. The presence of these bridging water molecules can be a very important factor in binding of molecules, but water molecules can play a variety of roles as they facilitate biomolecular interactions and stabilize structure. Often, due to a variety of experimental reasons, positions of water molecules in crystal structures are not well defined, even after x-ray crystallographic analysis. This mischaracterization and non-detection of water positions can be correlated with x-ray crystallographic resolution, with better resolution both locating a larger number of water molecules and placing their positions more accurately. Thus, to thoroughly evaluate structure, it is often necessary to verify water molecules systematically with tools such as the GRID program of Goodford [133]. Concomitantly, it is desirable to know which of these waters are subject to displacement by ligands and which are conserved. Using HINT score combined with a metric based on geometry, Amadasi *et al.* developed a robust method to calculate the relevance of binding site waters; those with particularly high relevance score would be expected to yield extra entropy if a ligand was designed to displace it, i.e., similar to the cyclic HIV-1 protease inhibitors [134]. In another study, the contribution of bridging water molecules to overall free energy of binding has been derived and quantitated [132].

3D QSAR with HINT—A very early application of the HINT force field was the introduction of field hydrophobicity parameters into 3D QSAR technology, to complement the original steric and electronic fields in CoMFA [103]. The steroid data set originally reported by Cramer *et al.* [102] was reexamined with the addition of a HINT-derived field. While this study provided little advantage in terms of statistical improvement due to a variety of reasons described previously [119], it provided a distinct advantage in chemical interpretability for chemists aiming to design new molecules based on such a QSAR study. Quite a few reports of studies based on the HINT-CoMFA methodology have been reported since then [104-106, 135-140] and some, particularly where the ligands or active sites are particularly non-polar, do show significant statistical improvement when hydrophobic fields are included. Fields in 3D QSAR are another class of descriptor that often needs to be optimized for the data set [141] in that each data set has forces and structures that may be best represented hydrophobically, sterically, electrostatically, or with other types of fields.

CONCLUSIONS AND FUTURE DIRECTIONS

Hydrophobicity impacts every aspect of drug design and even delivery, as has been repeatedly pointed out over the past century. Studies of this phenomenon have resulted in multiple theories, algorithms and tools for applying the concept. A large amount of effort has been put forth into studying the partition coefficient both experimentally, especially in terms of its prediction because of its importance in “druggability” of compounds. Many theoretical methods are robust in estimating LogP for molecules similar to their training set, but large errors are fairly common for compounds with large chemical and structural differences from that set.

Since Hansch and Fujita introduced the QSAR method, drug design projects have repeatedly found use for hydrophobic parameters. This dependence of drug design on lipophilicity is intuitive arising from drugs and proteins coming together, or proteins folding, in order to reduce the surface area in contact with polar water molecules. Quantification of this phenomenon has taken many forms, such as calculating of hydrophobic surface contact area to represent hydrophobic interactions, supplementing 3D QSAR with hydrophobic fields (HINT and MLP) [103,107-109], and with direct quantification of intermolecular interactions with HINT [123, 126-129]. While there are numerous force fields available, most are Newtonian in origin and concentrate on H-bonding, Coulombic interactions, van der Waal's interactions and London forces for estimating the strength of molecular interactions, all of which are mostly if not entirely enthalpic. HINT is different in that it accounts for both hydrophobic and hydrophilic interactions, and is derived from a free energy experiment. The availability of the HINT toolkit [142,143] makes it possible to develop application programs for computer-aided drug discovery and design.

The complex phenomena of hydrophobicity and hydrophobic interactions are still only poorly understood and remain quite difficult to simulate. Considerable resources and effort will continue towards development of accurate experimental and theoretical methods to determine/predict the partition coefficient. However, understanding and exploiting the hydrophobic effect in drug design, e.g., docking, target structure prediction, etc., will undoubtedly be increasingly important in the future. It is generally accepted that proteins fold in such a way that hydrophobic groups are largely shielded from water by hydrophilic groups. While evaluating the factors that affect prediction of protein folds, Park *et al.* noted that hydrophobicity of residues is the largest force defining protein structure, but that other factors were involved as well [144]. Accurate hydrophobicity measurements and estimation of hydrophobic interactions could have a tremendous impact on the modeling of protein folding and side chain orientation. Better modeling and representation of hydrophobic interactions will also contribute to the understanding of biological processes implicating hydrophobicity such as drug resistance

caused by efflux. There is no doubt in our mind as to the applications of this phenomenon in computational life sciences and computer-aided drug design. Hydrophobicity may not be the “Holy Grail” of biomolecular phenomena, but it is definitely the one of the “Commandments”.

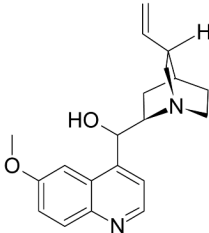
Acknowledgments

The authors would like to acknowledge the help of Mr. Chenxiao Da, Dr. Philip Mosier and Dr. Alexander Bayden for assistance with the preparation of this manuscript. HINT was originally conceived by Dr. Donald J. Abraham and G.E.K. The research groups led by Dr. Andrea Mozzarelli and Dr. Pietro Cozzini of the University of Parma have contributed significantly to the development of HINT. Research into the HINT model has also involved the invaluable contributions of a large number of collaborators whose names are listed in the references. The work in the G.E.K. laboratory is partially supported by U.S. NIH grant R01GM071894.

Appendix

Appendix 1 Fragmental methods for determination of LogP values

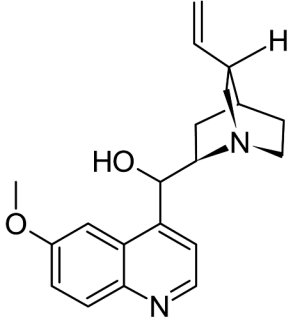
Rekker's method is highlighted with an example adapted from Mannhold and van de Waterbeemd [18].

Fragmentation Methods	 quinidine																																								
This approach breaks a molecule into fragments and assumes that the total LogP of a molecule is the sum total of all contributions of each fragment. However, the molecular environment affects the contributions by each fragment. Hence, correction factors are included in the calculation as shown by the following equation:	<table border="0"> <tr> <td colspan="3" data-bbox="927 1045 1427 1075"><i>Fragments:</i></td> </tr> <tr> <td data-bbox="927 1096 1081 1121">1 quinolinyl (-1H)</td> <td></td> <td data-bbox="1333 1096 1398 1121">+1.617</td> </tr> <tr> <td data-bbox="927 1123 1052 1148">1 O (aromatic)</td> <td></td> <td data-bbox="1333 1123 1398 1148">-0.450</td> </tr> <tr> <td data-bbox="927 1150 1062 1176">1 OH (aliphatic)</td> <td></td> <td data-bbox="1333 1150 1398 1176">-1.448</td> </tr> <tr> <td data-bbox="927 1178 1049 1203">1 N (aliphatic)</td> <td></td> <td data-bbox="1333 1178 1398 1203">-2.074</td> </tr> <tr> <td></td> <td style="text-align: right;">SUM</td> <td data-bbox="1333 1203 1398 1228"><u>-2.355</u></td> </tr> <tr> <td data-bbox="927 1230 1122 1255"><i>CH residual:</i> C₁₁H₁₈</td> <td></td> <td data-bbox="1333 1230 1398 1255">+4.893</td> </tr> <tr> <td></td> <td style="text-align: right;">SUM</td> <td data-bbox="1333 1255 1398 1281"><u>2.538</u></td> </tr> <tr> <td colspan="3" data-bbox="927 1283 1427 1308"><i>Corrections:</i></td> </tr> <tr> <td data-bbox="927 1310 1133 1335">Proximity effect (+2C_M),</td> <td></td> <td data-bbox="1333 1310 1398 1335">+0.438</td> </tr> <tr> <td data-bbox="927 1337 1235 1362">Electronegativity facing bulk (-2C_M),</td> <td></td> <td data-bbox="1333 1337 1398 1362">-0.438</td> </tr> <tr> <td data-bbox="927 1365 1052 1390">O-C-Ar (+1C_M)</td> <td></td> <td data-bbox="1333 1365 1398 1390">+0.219</td> </tr> <tr> <td></td> <td style="text-align: right;">LogP</td> <td data-bbox="1333 1411 1398 1436">2.757</td> </tr> </table>		<i>Fragments:</i>			1 quinolinyl (-1H)		+1.617	1 O (aromatic)		-0.450	1 OH (aliphatic)		-1.448	1 N (aliphatic)		-2.074		SUM	<u>-2.355</u>	<i>CH residual:</i> C ₁₁ H ₁₈		+4.893		SUM	<u>2.538</u>	<i>Corrections:</i>			Proximity effect (+2C _M),		+0.438	Electronegativity facing bulk (-2C _M),		-0.438	O-C-Ar (+1C _M)		+0.219		LogP	2.757
<i>Fragments:</i>																																									
1 quinolinyl (-1H)		+1.617																																							
1 O (aromatic)		-0.450																																							
1 OH (aliphatic)		-1.448																																							
1 N (aliphatic)		-2.074																																							
	SUM	<u>-2.355</u>																																							
<i>CH residual:</i> C ₁₁ H ₁₈		+4.893																																							
	SUM	<u>2.538</u>																																							
<i>Corrections:</i>																																									
Proximity effect (+2C _M),		+0.438																																							
Electronegativity facing bulk (-2C _M),		-0.438																																							
O-C-Ar (+1C _M)		+0.219																																							
	LogP	2.757																																							
$\text{LogP} = \sum_{i=1}^n a_i f_i + \sum_{j=1}^m b_j F_j$																																									
where, LogP = log of the partition coefficient a = the number of fragments, f = fragmental constant b _j = frequency of F _j F _j = correction factor for the j th fragment																																									
A simple calculation by Rekker's fragmental method is illustrated in the panel at the right. The experimentally determined value of LogP for quinidine's is 3.44.																																									

Appendix

Appendix 2 Atom contribution methods

the calculation of LogP for quinidine by atom contributions is shown (adapted from Mannhold and van de Waterbeemd [18]).

Atom Contribution Methods			
<p>This is an extension of the fragmental contribution method. It is assumed that the total LogP of the molecule is a contribution by each individual atom comprising it (instead of a contribution by fragments). Calculation as shown by the following equation:</p> $\text{LogP} = \sum n_i a_i$ <p>where, n_i = the number of atoms of type i a_i = fragmental constant</p> <p>A simple calculation by the Ghose / Crippen method is illustrated in the panel at the right. The experimentally determined value of Log P for quinidine is 3.44.</p>		 <p style="text-align: right;">Quinidine</p>	
Type	Description	Frequency	Contribution
2	C in CH ₂ R ₂	2	-0.9748
3	C in CHR ₃	2	-0.7266
5	C in CH ₃ X	1	-1.0824
6	C in CH ₂ RX	2	-1.6740
8	C in CHR ₂ X	2	-1.0420
15	C in =CH ₂	1	-0.1053
16	C in =CHR	1	-0.0681
24	C in R--CH--R	4	+0.0272
25	C in R--CR--R	2	+0.3200
26	C in R--CX--R	2	-0.2066
27	C in R--CH--X	1	+0.0598
46	H attached to C ⁰ _{sp3} with no X next to C	1	+0.4410
47	H attached to C ¹ _{sp3} or C ⁰ _{sp3}	16	+5.3488
48	H attached to C ² _{sp3} , C ¹ _{sp3} or C ⁰ _{sp3}	1	+0.3161
50	H attached to heteroatom	1	-0.3260
52	H attached to C ⁰ _{sp3} with one X next to C	5	+1.8475
56	O in alcohol	1	+0.1402
60	O in Al - Al, Ar ₂ O, R:R or R-O-C=X	1	+0.2712
68	N in Al ₃ N	1	+0.3954
75	N in R--N--R or R--N--X	1	-0.1106
LogP			2.852
<p>Key symbols: R is group connected to C; X is heteroatom; "=" is double bond; ":" is an aromatic single bond such as the C-N bond in pyrrole; subscripts give the hybridization state and superscripts the formal oxidation number.</p>			

Appendix

Appendix 3 A parabolic relationship exists between drug potency and hydrophobicity

An equation relating probability and partition coefficient was derived by McFarland [79].

Model of n alternating aqueous phases and lipophilic membranes, in a hypothetical biological system is shown below:

Aq ₀	Lip ₁	Aq ₂	Lip ₃	Aq ₄Lip _{$n-2$}	Lip _{$n-1$}	Aq _{n}
-----------------	------------------	-----------------	------------------	------------------------------------------------------	---------------------------------	------------------------------

Assume that the rate of passage of molecules from aqueous to lipophilic zones is k , while the rate of passage of molecules in the opposite direction is given by l . Then, the partition coefficient of the molecule shall be given by k/l .

If $P_{x,y}$ is the probability of moving a molecule from layer x to layer y , the probability of moving a molecule from aq_0 to aq_n is given by:

$$P_{0,n} = P_{0,1} \cdot P_{1,2} \cdot P_{2,3} \cdot P_{3,4} \cdots P_{n-2,n-1} \cdot P_{n-1,n}$$

Although the probability of a molecule passing through one layer to another would depend on a number of things, the least of which happens to be the partition coefficient, we can assume an unbiased passage of the molecule. This reduces the entire problem to one of pure probability. So, considering no bias in the passage of molecules between layers, we shall have:

$$P_{0,1} = P_{2,3} = P_{n-2,n-1}$$

Similarly, the following equation can also be obtained:

$$P_{1,2} = P_{3,4} = P_{n-1,n}$$

Combining all three above equations, we have:

$$P_{0,n} = (P_{0,1})^{n/2} \cdot (P_{1,2})^{n/2}$$

Now, the number of molecules being transferred from aq_0 to lip_1 is proportional to k . Since molecules entering lip_1 may return to aq_0 , the number of molecules staying in lip_1 is proportional to the sum of k and l . The probability of a molecule moving from layer 0 to 1 is given by $P_{0,1}$, which is mathematically defined as:

$$P_{0,1} = \frac{k}{k+l}$$

Dividing both numerator and denominator on the right hand side of the equation by l , we get:

$$P_{0,1} = \frac{k/l}{k/l+1}$$

If $P_{1,0}$ is the probability of a molecule passing from lip_1 to aq_0 ,

$$P_{1,0} = 1 - P_{0,1}$$

Since we have assumed equal probabilities, we have:

$$P_{1,2} = 1 - P_{0,1}$$

Substituting this equation into the fourth equation described above, we have:

$$P_{0,n} = (P_{1,2})^{n/2} \cdot (1 - P_{1,2})^{n/2}$$

$$P_{0,1} = \left(\frac{k/l}{k/l+1} \right)^{n/2} \cdot \left(1 - \frac{k/l}{k/l+1} \right)^{n/2}$$

$$P_{0,1} = \left(\frac{k/l}{k/l+1} \right)^{n/2} \cdot \left(\frac{1}{k/l+1} \right)^{n/2}$$

$$P_{0,1} = \frac{(k/l)^{n/2}}{(k/l+1)^n}$$

This relationship will be parabolic, indicating that there is an optimum range of LogP with respect to drug potency.

Appendix

Appendix 4 Hansch Analysis

a method to relate physicochemical parameters to drug potency. For details, refer to Hansch and Fujita [95].

The 'extrathermodynamic approach' relates various free energy-like descriptors in a model:

$$\log \frac{1}{C} = a(\log P)^2 + b \log P + c\sigma + \dots + k$$

(C, LogP and σ are the inhibition constant, log of the partition coefficient and steric parameters respectively, while a,b,c and k are constants.) This model explains that drug transport from site of application to the site of action depends on the lipophilicity of the drug and is non-linear under typical conditions. Although special conditions could reduce this equation into simpler forms [73], this equation surmises the behavior of any molecule under normal diffusion conditions.

A novel parameter π defines the lipophilicity of substituent X:

$$\pi_X = \log P_{RX} + \log P_{RH}$$

where LogP is the log of the partition coefficient. This equation was a variant of the Hammett Equation,

$$\rho\sigma = \log K_{RX} + \log K_{RH}$$

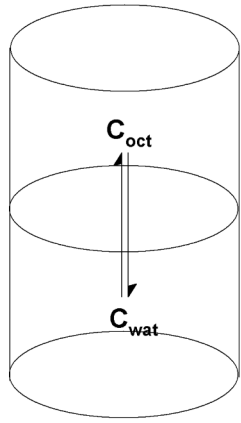
where the reaction equilibrium constants have been substituted with partition coefficients.

Values for LogP and σ of different molecules may be correlated to their IC₅₀ or K_i values by statistical analysis such as multiple linear regressions (MLR) or partial least squares (PLS). Artificial intelligence methods such as Neural Networks (NN) and Support Vector Machines (SVM) have also been used. These analysis methods have pros and cons: While MLR and PLS do a good job of finding linear relationships between variables, they tend to oversimplify. On the other hand, AI methods tend to pick up on minute non-linear trends and tend to over-fit models.

Appendix

Appendix 5 The HINT Paradigm

A “natural” free energy force field based on LogP. It is available as a toolkit, allowing flexibility in development of applications. Refer to Kellogg and Abraham [118] for details.

 <p>A representation of a shake flask. Substances distribute themselves between the water and octanol layers with concentrations C_{wat} and C_{oct}, respectively, in a particular ratio called the partition coefficient.</p>	<p>By definition,</p> $P = \frac{C_{oct}}{C_{wat}}$ <p>LogP can be considered the sum total of individual lipophilic propensities of each atom called hydrophobic atom constant (a_i), i.e.</p> $\text{Log}P = \sum a_i$ <p>The values of a_i are readily available from various methods, as described earlier; HINT itself uses an adaptation of the Hansch and Leo C-LOGP approach [23-27]. The HINT method calculates scores (b_{ij}) of each atom against all other atoms according to the equation</p> $b_{ij} = a_i S_i a_j S_j T_{ij} R_{ij} + r_{ij}$ <p>Where, a_i is the hydrophobic atom constant for the i^{th} atom and S_{ij} is the solvent accessible surface area. T_{ij} is a variable which takes on the values of +1 or -1, depending on the acid and base properties of the pair of atoms being considered. For example, if the atoms under consideration are both amino nitrogens, the interaction is unfavorable and T_{ij} is -1. In contrast, if one is a polar (amine) hydrogen and the other is a carboxyl oxygen, their interaction would be favorable and T_{ij} is +1. R_{ij} is the exponential term e^{-r}, where r is the distance between the i and j atoms. r_{ij} is a van der Waal's term.</p> <p>The total HINT score would be the sum total of each atom-atom score thus calculated, i.e.</p> $\text{HINT Score} = \sum b_{ij} = \sum \sum a_i S_i a_j S_j T_{ij} R_{ij} + r_{ij}$
<p>If ΔG is the change in Gibb's Free Energy, R is the gas constant and T is the absolute temperature,</p> <p>we know that $\text{Log}P = \sum a_i$ and $\text{Log}P = -\Delta G/2.303RT$;</p> <p>thus, $\text{HINT score} \propto f(a_i)$.</p> <p>Hence, HINT scores reflect free energy by taking into account both enthalpic and entropic factors. Because the absolute value of HINT score may not be predictive of a discrete biological association event, these absolute scores are not as important or relevant as differences in HINT score values between analogous systems, much the same as differences in free energy, i.e., $\Delta\Delta G$. The difference in Gibb's Free Energy between two states is a very important parameter as it tells us about the spontaneity or likelihood of the change, whereas the value for any one state itself is often of less consequence.</p>	

References

1. Meyer EE, Rosenberg KJ, Israelachvili J. Recent progress in understanding hydrophobic interactions. Proc. Nat. Acad. Sci. USA 2006;103(43):15739–15746. [PubMed: 17023540]
2. Meyer H. Zur theorie der alkoholnarkose: I. Welche eigenschaft der anaesthetika bedingt ihre narkotische wirkung? Arch. Exp. Pathol. Pharmacol 1899;42(2-4):109–118.
3. Meyer KH. Contributions to the theory of narcosis. Trans. Faraday Soc 1937;33:1062–1064.
4. Overton E. Über die allgemeinen osmotischen eigenschaften der zelle, ihre vermutlichen ursachen und ihre bedeutung für die physiologie. Vierteljahrsschr. Naturforsch. Ges. Zürich 1899;44:87–136.
5. Butler JAV. The energy and entropy of hydration of organic compounds. Trans. Faraday Soc 1937;33:229–236.

6. Frank HS, Evans MW. Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J. Chem. Phys* 1945;13(11):507–532.
7. Teeter MM. Water structure of a hydrophobic protein at atomic resolution: pentagon rings of water molecules in crystals of crambin. *Proc. Nat. Acad. Sci. USA* 1984;81(19):6014–6018. [PubMed: 16593516]
8. Klotz IM. Protein hydration and behavior. *Science* 1958;128(3328):815–822. [PubMed: 13592259]
9. Kauzmann W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem* 1959;14:1–63. [PubMed: 14404936]
10. Scheraga HA, Némethy G, Steinberg IZ. The contribution of hydrophobic bonds to the thermal stability of protein conformations. *J. Biol. Chem* 1962;237(8):2506–2508. [PubMed: 14498257]
11. Hermann RB. Theory of hydrophobic bonding. I. Solubility of hydrocarbons in water, within the context of the significant structure theory of liquids. *J. Phys. Chem* 1971;75(3):363–368.
12. Hermann RB. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem* 1972;76(19):2754–2759.
13. Hermann RB. Theory of hydrophobic bonding. III. Method for the calculation of the hydrophobic interaction based on liquid state perturbation theory and a simple liquid model. *J. Phys. Chem* 1975;79(2):163–169.
14. Reynolds JA, Gilbert DB, Tanford C. Empirical correlation between hydrophobic free energy and aqueous cavity surface area. *Proc. Nat. Acad. Sci. USA* 1974;71(8):2925–2927. [PubMed: 16578715]
15. Leo A, Hansch C, Jow PYC. Dependence of hydrophobicity of apolar molecules on their molecular volume. *J. Med. Chem* 1976;19(5):611–615. [PubMed: 1271402]
16. Leo A, Hansch C, Elkins D. Partition coefficients and their uses. *Chem. Rev* 1971;71(6):525–616.
17. Buchwald P, Bodor N. Octanol-water partition: searching for predictive models. *Curr. Med. Chem* 1998;5(5):353–380. [PubMed: 9756979]
18. Mannhold R, van de Waterbeemd H. Substructure and whole molecule based approaches for calculating LogP. *J. Comput.-Aid. Mol. Des* 2001;15(4):337–354.
19. Mannhold R, Petrauskas A. Substructure versus whole-molecule approaches for calculating LogP. *QSAR Comb. Sci* 2003;22(4):466–475.
20. Manhold R, Poda GI, Ostermann C, Tetko IV. Calculation of molecular lipophilicity: state-of-the-art and comparison of LogP methods on more than 96000 compounds. *J. Pharm. Sci* 2009;98(3):861–893. [PubMed: 18683876]
21. Livingstone DJ. Theoretical property predictions. *Curr. Top. Med. Chem* 2003;3(10):1171–1192. [PubMed: 12769715]
22. Nys GG, Rekker RF. The concept of hydrophobic fragmental constants (f-values). II. Extension of its applicability to the calculation of lipophilicities of aromatic and hetero-aromatic structures. *Chem. Ther* 1974;9(4):361–374.
23. Leo A, Jow PY, Silipo C, Hansch C. Calculation of hydrophobic constant (LogP) from π and f constants. *J. Med. Chem* 1975;18(9):865–868. [PubMed: 1159707]
24. Leo AJ. Some advantages of calculating octanol-water partition coefficients. *J. Pharm. Sci* 1987;76(2):166–168. [PubMed: 3572757]
25. Leo AJ. Hydrophobic parameter: measurement and calculation. *Methods Enzymol* 1991;202:544–591. [PubMed: 1784188]
26. Leo AJ. Calculating LogP_{oct} from structures. *Chem. Rev* 1993;93(4):1281–1306.
27. Leo AJ, Hoekman D. Calculating logP (oct) with no missing fragments; the problem of estimating new interaction parameters. *Perspect. Drug Discov. Des* 2000;18:19–38.
28. Petrauskas A, Kolovanov EA. ACD/LogP method description. *Perspect. Drug Discov. Des* 2000;19(1):99–116.
29. Ghose AK, Crippen GM. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships. I. Partition coefficients as a measure of hydrophobicity. *J. Comp. Chem* 1986;7(4):565–677.

30. Ghose AK, Crippen GM. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships. II. Modeling dispersive and hydrophobic interactions. *J. Chem. Inf. Comput. Sci* 1987;27(3):21–35. [PubMed: 3558506]
31. Ghose AK, Pritchett A, Crippen GM. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships. III. Modeling hydrophobic interactions. *J. Comp. Chem* 1988;9(1):80–90.
32. Ghose AK, Vishwanadhan VN, Wendoloski JJ. Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: an analysis of ALOGP and CLOGP methods. *J. Phys. Chem. A* 1998;102(21):3762–3772.
33. Wang RX, Fu Y, Lai LH. A new atom-additive method for calculating partition coefficients. *J. Chem. Inf. Comput. Sci* 1997;37(3):615–621.
34. Wang RX, Gao Y, Lai LH. Calculating partition coefficient by atom additive method. *Perspect. Drug. Discov. Des* 2000;19(1):47–66.
35. Cheng T, Zhao Y, Li X, Lin F, Xu Y, Zhang X, Li Y, Wang R, Lai L. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model* 2007;47(6):2140–2148. [PubMed: 17985865]
36. Rogers KS, Cammarata A. A molecular orbital description of the partitioning of aromatic compounds between polar and non-polar phases. *Biochim. Biophys. Acta* 1969;193(1):22–29. [PubMed: 5349616]
37. Rogers KS, Cammarata A. Superdelocalizability and charge density. a correlation with partition coefficients. *J. Med. Chem* 1969;12(4):692–693. [PubMed: 5793164]
38. Hopfinger AJ, Battershell RD. Application of SCAP to drug design. 1. Estimation of octanol-water partition coefficients using solvent-dependent conformational analyses. *J. Med. Chem* 1976;19(5):569–573. [PubMed: 1271397]
39. Klopman G, Iroff LD. Calculation of partition coefficients by the charge density method. *J. Comput. Chem* 1980;2(2):157–160.
40. Bravi G, Wikel JH. Application of MS-WHIM parameters: 3. Prediction of molecular properties. *Quant. Struct.-Act. Relat* 2000;19(1):39–49.
41. Toulmin A, Wood JM, Kenny PW. Toward prediction of alkane/water partition coefficients. *J. Med. Chem* 2008;51(13):3720–3730. [PubMed: 18558667]
42. Kier, LB.; Hall, LH. *Molecular Structure Description*. Academic Press; New York, NY: 1999.
43. Livingstone DJ, Ford MG, Huuskonen JJ, Salt DW. Simultaneous prediction of aqueous solubility and octanol/water partition coefficient based on descriptors derived from molecular structure. *J. Comput.-Aid. Mol. Des* 2001;15(8):741–752.
44. Molnár L, Keserű GM, Papp Á, Gulyás Z, Darvas F. A neural network based prediction of octanol-water partition coefficients using atomic5 fragmental descriptors. *Bioorg. Med. Chem. Lett* 2004;14(4):851–853. [PubMed: 15012980]
45. Liao Q, Yao J, Yuan S. SVM approach for predicting LogP. *Mol. Divers* 2006;10(3):301–309. [PubMed: 17031534]
46. Taskinen J, Yliruusi J. Prediction of physicochemical properties based on neural network modelling. *Adv. Drug Del. Rev* 2003;55(9):1163–1183.
47. Abraham DJ, Leo AJ. Extension of the fragment method to calculate amino acid zwitterions and side chain partition coefficients. *Prot. Struct. Funct. Genet* 1987;2(2):130–152.
48. Buchwald P, Bodor N. Octanol-water partition of non-zwitterionic peptides: Predictive power of a molecular size-based model. *Prot. Struct. Funct. Genet* 1998;30(1):86–89.
49. Steinmetz WE. A CoMFA analysis of selected physical properties of amino acids in water. *Quant. Struct.-Act. Relat* 1995;14(1):19–23.
50. Sotomatsu-Niwa T, Ogino A. Evaluation of the hydrophobic parameters of the amino acid side chains of peptides and their application in QSAR and conformational studies. *J. Mol. Struct. (Theochem)* 1997;392:43–54.
51. Akamatsu M, Yoshida Y, Nakamura H, Asao M, Iwamura H, Fujita T. Hydrophobicities of di- and tri-peptides having unionizable side-chains and correlation with substituent and structural parameters. *Quant. Struct.-Act. Relat* 1989;8(3):195–203.

52. Akamatsu M, Okutani S, Nakao K, Hong NJ, Fujita T. Hydrophobicities of N-acetyl-di- and tripeptide amides having unionizable side chains and correlation with substituent and structural parameters. *Quant. Struct.-Act. Relat* 1990;9(3):189–194.
53. Akamatsu M, Fujita T. Quantitative analyses of hydrophobicity of di- to pentapeptides having unionizable side chains with substituent and structural parameters. *J. Pharm. Sci* 1992;81(2):164–174. [PubMed: 1545357]
54. Akamatsu M, Katayama T, Kishimoto D, Kurokawa Y, Shibata H, Ueno T, Fujita T. Quantitative analyses of the structure-hydrophobicity relationship for N-acetyl di- and tripeptide amides. *J. Pharm. Sci* 1994;83(7):1026–1033. [PubMed: 7965659]
55. Tao P, Wang R, Lai L. Calculating partition coefficients of peptides by the addition method. *J. Mol. Model* 1999;5(10):189–195.
56. Thompson S, Hattotuwigama CK, Holliday JD, Flower DR. On the hydrophobicity of peptides: comparing empirical predictions of peptide logP values. *Bioinformatics* 2006;1(7):237–241. [PubMed: 17597897]
57. Berthod A, Carda-Broch S. Determination of liquid-liquid partition coefficients by separation method. *J. Chromatogr. A* 2004;1037(1-2):3–14. [PubMed: 15214657]
58. Nozaki Y, Tanford C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J. Biol. Chem* 1971;246(7):2211–2217. [PubMed: 5555568]
59. Rose GD. Prediction of chain turns in globular proteins on a hydrophobic basis. *Nature* 1978;272(5654):586–590. [PubMed: 643051]
60. Bull HB, Breese K. Surface tension of amino acid solutions: a hydrophobicity scale of the amino acid residues. *Arch. Biochem. Biophys* 1974;161(2):665–670. [PubMed: 4839053]
61. Wolfenden R, Andersson L, Cullis PM, Southgate CCF. Water, protein folding and the genetic code. *Science* 1979;206(4418):575–577. [PubMed: 493962]
62. Kyte J, Doolittle RF. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol* 1982;157(1):105–132. [PubMed: 7108955]
63. Chothia C. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol* 1976;105(1):1–14. [PubMed: 994183]
64. Wimley WC, White SH. Experimentally determined hydrophobicity scale or proteins at membrane interfaces. *Nat. Struct. Biol* 1996;3(10):842–848. [PubMed: 8836100]
65. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. *Nature* 1986;319(6050):199–203. [PubMed: 3945310]
66. Li S-C, Deber C. A measure of helical propensity for amino acids in membrane environments. *Nat. Struct. Biol* 1994;1(6):368–373. [PubMed: 7664049]
67. Dyson HJ, Wright PE, Scheraga HA. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Nat. Acad. Sci. USA* 2006;103(35):13057–13061. [PubMed: 16916929]
68. Hansch, C.; Leo, AJ. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. J. Wiley and Sons; New York, NY: 1979.
69. Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE. Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc. Nat. Acad. Sci. USA* 2008;105(17):6278–6283. [PubMed: 18434548]
70. MacCallum JL, Moghaddam MS, Chan HS, Tieleman DP. Hydrophobic association of α -helices, steric dewetting and enthalpic barriers to protein folding. *Proc. Nat. Acad. Sci. USA* 2007;104(49):6206–6210. [PubMed: 17404236]
71. Mardia KV, Nyirongo VB. Simulating virtual protein Ca traces with applications. *J. Comput. Biol* 2008;15(9):1209–1220. [PubMed: 18973436]
72. Lau KF, Dill KA. Lattice statistical mechanics model of the conformation and sequence space of proteins. *Macromol* 1989;22(10):3986–3997.
73. Gupta A, Manüch J, Stacho L. Structure-approximating inverse protein folding problem in the 2D HP model. *J. Comput. Biol* 2005;12(10):1328–1345. [PubMed: 16379538]

74. Shmygelska A, Hoos HH. An ant colony optimization algorithm for the 2D and 3D hydrophobic polar protein folding problem. *BMC Bioinformatics* 2005;6:30. [PubMed: 15710037]
75. Hansch C, Vittoria A, Silipo C, Jow PY. Partition coefficients and the structure-activity relation of the anesthetic gases. *J. Med. Chem* 1975;18(6):546–548. [PubMed: 1151965]
76. Hansch C, Steward AR, Anderson SM, Bentley D. Parabolic dependence of drug action upon lipophilic character as revealed by a study of hypnotics. *J. Med. Chem* 1968;11(1):1–11. [PubMed: 5637185]
77. Lobl TJ, Tindall DJ, Cunningham GR, Kemp PL, Campbell JA. The correlation of steroid partition coefficients with binding affinities to the rat cytoplasmic androgen receptor, rat androgen-binding protein and human testosterone estradiol-binding globulin. *Ann. N. Y. Acad. Sci* 1982;383:477–478. (JUN).
78. Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficient. *Nature* 1962;194(4824):178–180.
79. McFarland JW. Parabolic relation between drug potency and hydrophobicity. *J. Med. Chem* 1970;13(6):1192–1196. [PubMed: 5479865]
80. Kier LB. A theory of inhaled anesthetic action by disruption of ligand diffusion chreodes. *Am. Assoc. Nurse Anesthet. J* 2003;71(6):422–428.
81. 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;46(1-3):3–26. [PubMed: 11259830]
82. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Meth* 2000;44(1):235–249.
83. Clardy J, Walsh C. Lessons from natural molecules. *Nature* 2004;432(7019):829–837. [PubMed: 15602548]
84. Keller TH, Pichota A, Yin Z. A practical view of ‘drugability’. *Curr. Op. Chem. Biol* 2006;10(4):357–361.
85. Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol* 2004;1(4):337–341.
86. Muegge I. Selection criteria for drug-like compounds. *Med. Res. Rev* 2003;23(3):302–321. [PubMed: 12647312]
87. Walters WP, Murcko AA, Murcko MA. Recognizing molecules with drug-like properties. *Curr. Op. Chem. Biol* 1999;3:384–387.
88. Lipinski C, Hopkins A. Navigating chemical space for biology and medicine. *Nature* 2004;432(7019):855–861. [PubMed: 15602551]
89. Oprea TI, Davis AM, Teague SJ, Leeson PD. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci* 2001;41(5):1308–1315. [PubMed: 11604031]
90. Moriguchi I. Quantitative-structure activity studies. I. parameters relating to hydrophobicity. *Chem. Pharm. Bull* 1975;23(2):247–257.
91. Oprea TI. Current trends in lead discovery: are we looking for the appropriate properties? *Mol. Divers* 2000;5(4):199–208. [PubMed: 12549672]
92. Proudfoot JR. The evolution of synthetic oral drug properties. *Bioorg. Med. Chem. Lett* 2005;15(4):1087–1090. [PubMed: 15686918]
93. Nikolova N, Jaworska J. Approaches to measure chemical similarity – a review. *QSAR Comb. Sci* 2003;22(9-10):1006–1026.
94. Kubinyi H. QSAR and 3D QSAR in drug design. Part 1. Methodology. *Drug Discov. Today* 1997;2(11):457–467.
95. Dudek AZ, Arodz T, Gálvez J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Comb. Chem. High T. Scr* 2006;9(3):213–228.
96. Hansch C, Fujita T. ρ - σ - π analysis: a method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc* 1964;86(8):1616–1626.
97. Muir RM, Fujita T, Hansch T. Structure-activity relationship in the auxin activity of mono-substituted phenylacetic acids. *Plant Physiol* 1967;42(11):1519–1526. [PubMed: 6080870]

98. Fujita T, Hansch C. Analysis of the structure-activity relationship of the sulfonamide drugs using substituent constants. *J. Med. Chem* 1967;10(6):991–1000. [PubMed: 6056065]
99. Kubinyi H. QSAR and 3D QSAR in drug design. Part 2. Applications and problems. *Drug Discov. Today* 1997;2(12):538–546.
100. Hansch C, Kurup A, Garg R, Gao H. Chem-bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms. *Chem. Rev* 2001;101(3):619–672. [PubMed: 11712499]
101. Li X-Z, Nikaido H. Efflux-mediated drug resistance in bacteria. *Drugs* 2004;64(2):159–204. [PubMed: 14717618]
102. Cramer RD III, 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(18):5959–5967.
103. Kellogg GE, Semus SF, Abraham DJ. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J. Comput.-Aid. Mol. Des* 1991;5(6):545–552.
104. Debnath AK. Application of 3D-QSAR techniques in anti-HIV1 drug design – an overview. *Curr. Pharm. Des* 2005;11(24):3091–3110. [PubMed: 16178747]
105. Oprea TI, Waller CL, Marshall GR. 3D-QSAR of human immunodeficiency virus (I) protease inhibitors. III. Interpretation of CoMFA results. *Drug Des. Discov* 1994;12:29–51. [PubMed: 7578806]
106. Nayyar A, Malde A, Jain R, Coutinho E. 3D-QSAR study of ring-substituted quinoline class of anti-tubercular agents. *Bioorg. Med. Chem* 2006;14(3):847–856. [PubMed: 16214351]
107. Testa B, Carrupt P-A, Gaillard P, Billois A, Weber P. Lipophilicity in molecular modeling. *Pharm. Res* 1996;13(3):335–343. [PubMed: 8692723]
108. Gaillard P, Carrupt P-A, Testa B, Boudon A. Molecular lipophilicity potential, a tool in 3D-QSAR: methods and applications. *J. Comput.-Aid. Mol. Des* 1994;8(2):83–96.
109. Barreca ML, Carotti A, Carrieri A, Chimirri A, Monforte AM, Calace MP, Rao A. Comparative molecular field analysis (CoMFA) and docking studies of non-nucleoside HIV-1 RT inhibitors (NNIs). *Bioorg. Med. Chem* 1999;7(11):2283–2292. [PubMed: 10632038]
110. Carrieri A, Carotti A, Barreca ML, Altomare C. Binding models of reversible inhibitors to type-B monoamine oxidase. *J. Comput.-Aid. Mol. Des* 2002;16(11):769.
111. Efremov RG, Chugunov AO, Pyrkov TV, Priestle JP, Arseniev AS, Jacoby E. Molecular lipophilicity in protein modeling and drug design. *Curr. Med. Chem* 2007;14(4):393–415. [PubMed: 17305542]
112. Ponder JW, Case DA. Force fields for protein simulation. *Adv. Prot. Chem* 2003;66:27–85.
113. Chothia C. Hydrophobic bonding and accessible surface area in proteins. *Nature* 1974;248(5446):338–339. [PubMed: 4819639]
114. Oobatake M, Ooi T. Hydration and heat stability effects on protein unfolding. *Prog. Biophys. Mol. Biol* 1993;59(3):237–284. [PubMed: 8441810]
115. Cramer CJ, Truhlar DG. An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. *Science* 1992;256(5054):213–217. [PubMed: 17744720]
116. Sharp KA, Nicholls A, Fine RF, Honig B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* 1991;252(5002):106–109. [PubMed: 2011744]
117. Cesari G, Sippl MJ. Structure derived hydrophobic potential. Hydrophobic potential derived from X-ray structures of globular proteins is able to identify native folds. *J. Mol. Biol* 1992;224(3):725–732. [PubMed: 1569551]
118. Hummer G. Hydrophobic force fields as a molecular alternative to surface-area models. *J. Am. Chem. Soc* 1999;121(26):6299–6305.
119. Kellogg GE, Abraham DJ. Hydrophobicity: is $\text{Log}P_{o/w}$ more than the sum of its parts? *Eur. J. Med. Chem* 2000;35(7-8):651–661. [PubMed: 10960181]
120. Kellogg GE, Burnett JC, Abraham DJ. Very empirical treatment of solvation and entropy: a force field derived from $\text{Log}P_{o/w}$. *J. Comput.-Aid. Mol. Des* 2001;15(4):381–393.
121. Abraham DJ, Kellogg GE. The effect of physical organic properties on hydrophobic fields. *J. Comput.-Aid. Mol. Des* 1994;8(1):41–49.

122. Cozzini P, Fornabaio M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. *J. Med. Chem* 2002;45(12):2469–2483. [PubMed: 12036355]
123. Spyraakis F, Amadasi A, Fornabaio M, Abraham DJ, Mozzarelli A, Kellogg GE. The consequences of scoring docked ligand conformations using free energy correlations. *Eur. J. Med. Chem* 2007;42(70):921–933. [PubMed: 17346861]
124. Fornabaio M, Cozzini P, Mozzarelli A, Abraham DJ, Kellogg GE. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 2. Computational titration and pH effects in molecular models of neuraminidase-inhibitor complexes. *J. Med. Chem* 2003;46(21):4487–4500. [PubMed: 14521411]
125. Fornabaio M, Spyraakis F, Mozzarelli A, Cozzini P, Abraham DJ, Kellogg GE. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 3. The free energy contribution of structural water molecules in HIV-1 protease complexes. *J. Med. Chem* 2004;47(18):4507–4516. [PubMed: 15317462]
126. Tripathi A, Fornabaio M, Kellogg GE, Gupton JT, Gewirtz DA, Yeudall WA, Vega NE, Mooberry S. Docking and hydrophobic scoring of polysubstituted pyrrole compounds with antitubulin activity. *Bioorg. Med. Chem* 2008;16(5):2235–2242. [PubMed: 18083520]
127. Simoni D, Invidiata FP, Eleopra M, Marchetti P, Rondanin R, Baruchello R, Grisolia G, Tripathi A, Kellogg GE, Durrant D, Lee RM. Design, synthesis and biological evaluation of novel stilbene-based antitumor agents. *Bioorg. Med. Chem* 2009;17(2):512–522. [PubMed: 19117761]
128. Porotto M, Fornabaio M, Kellogg GE, Moscona A. A second receptor binding site on human parainfluenza virus type 3 hemagglutinin-neuraminidase contributes to activation of fusion mechanism. *J. Virol* 2007;81(7):3216–3228. [PubMed: 17229690]
129. Burnett JC, Botti P, Abraham DJ, Kellogg GE. Computationally accessible method for estimating free energy changes resulting from site-specific mutations of biomolecules: systematic model building and structural/hydrophobic analysis of deoxy and oxy hemoglobins. *Prot. Struct. Funct. Genet* 2001;42(3):355–377.
130. Cashman DJ, Kellogg GE. A computational model for anthracycline binding to DNA: Tuning groove-binding intercalators for specific sequences. *J. Med. Chem* 2004;47(6):1360–1374. [PubMed: 14998326]
131. Bayden AS, Fornabaio M, Scarsdale JN, Kellogg GE. Web application for studying the free energy of binding and protonation states of protein-ligand complexes based on HINT. *J. Comput.-Aid. Mol. Des* 2009;23(9):621–632.
132. Amadasi A, Spyraakis F, Cozzini P, Abraham DJ, Kellogg GE, Mozzarelli A. Mapping the energetics of water-protein and water-ligand interactions with the “natural” HINT forcefield: predictive tools for characterizing the roles of water in biomolecules. *J. Mol. Biol* 2006;358(1):289–309. [PubMed: 16497327]
133. Goodford P. Multivariate characterization of molecules for QSAR analysis. *J. Chemometrics* 1996;10(2):107–117.
134. Amadasi A, Surface JA, Spyraakis F, Cozzini P, Mozzarelli A, Kellogg GE. Robust classification of “relevant” water molecules in putative protein binding sites. *J. Med. Chem* 2008;51(4):1063–1067. [PubMed: 18232647]
135. Pajeva I, Wiese M. Molecular modeling of phenothiazines and related drugs as multidrug resistance modifiers. A comparative molecular field analysis study. *J. Med. Chem* 1998;41(11):1815–1826. [PubMed: 9599232]
136. Tsakovska IM, Pajeva IK. Molecular modeling of triazine type MDR modulators using CoMFA and CoMSIA approaches. *SAR & QSAR Environ. Res* 2002;13(3-4):473–484. [PubMed: 12184388]
137. Welch W, Ahmad S, Airey JA, Geron K, Humerickhouse RA, Besch HR Jr, Deslongchamps P, Sutko JL. Structural determinants of high-affinity binding of ryanoids to the vertebrate skeletal muscle ryanodine receptor: a comparative molecular field analysis. *Biochemistry* 1994;33(20):6074–6085. [PubMed: 8193121]
138. Grell W, Hurnaus R, Griss G, Sauter R, Rupprecht E, Mark M, Luger P, Nar H, Wittneben H, Muller P. Repaglinide and related hypoglycemic benzoic acid derivatives. *J. Med. Chem* 1998;41(26):5219–5246. [PubMed: 9857091]

139. Bursi R, Grootenhuis PDJ. Comparative molecular field analysis and energy interaction studies of thrombin-inhibitor complexes. *J. Comput.-Aid. Mol. Des* 1999;13(3):221–232.
140. Iskander MN, Coupan IM, Winkler DA. Investigation of 5-HT₄ agonist activities using molecular field analysis. *J. Chem. Soc. Perkin Trans* 1999;(2):153–158.
141. Kellogg GE. Finding optimum field models for 3D QSAR. *Med. Chem. Res* 1997;7(6-7):417–427.
142. Kellogg GE, Fornabai M, Chen DL, Abraham DJ. New application design for a 3D hydrophobic map-based search for potential water molecules bridging between protein and ligand. *Internet Electron. J. Mol. Des* 2005;4(3):194–209.
143. The HINT toolkit at the Edu-Soft, LLC. Home page. <http://www.edusoft-lc.com/toolkits/>
144. Park BH, Huang ES, Levitt M. Factors affecting the ability of energy functions to discriminate correct from incorrect folds. *J. Mol. Biol* 1997;266(4):831–846. [PubMed: 9102472]

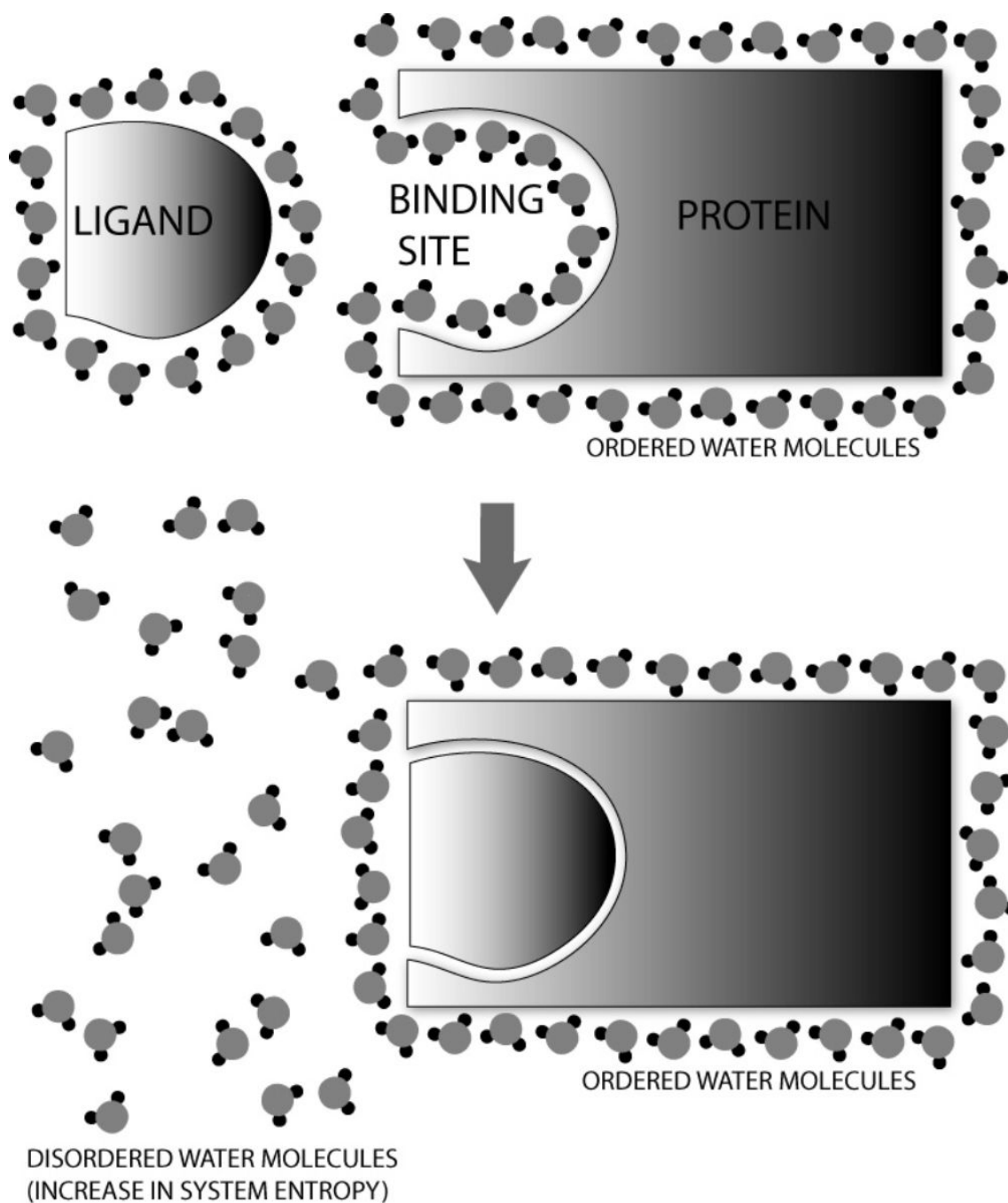


Figure 1. The Hydrophobic Effect

Hydrophobic molecules are surrounded by an ordered cage of water molecules. When two such molecules come together, they aggregate in order to reduce their surface area in contact with the polar water molecules. This causes a number of water molecules to be removed from their ordered formation, thus increasing disorder (increased entropy) and potentially making the process energetically favorable.

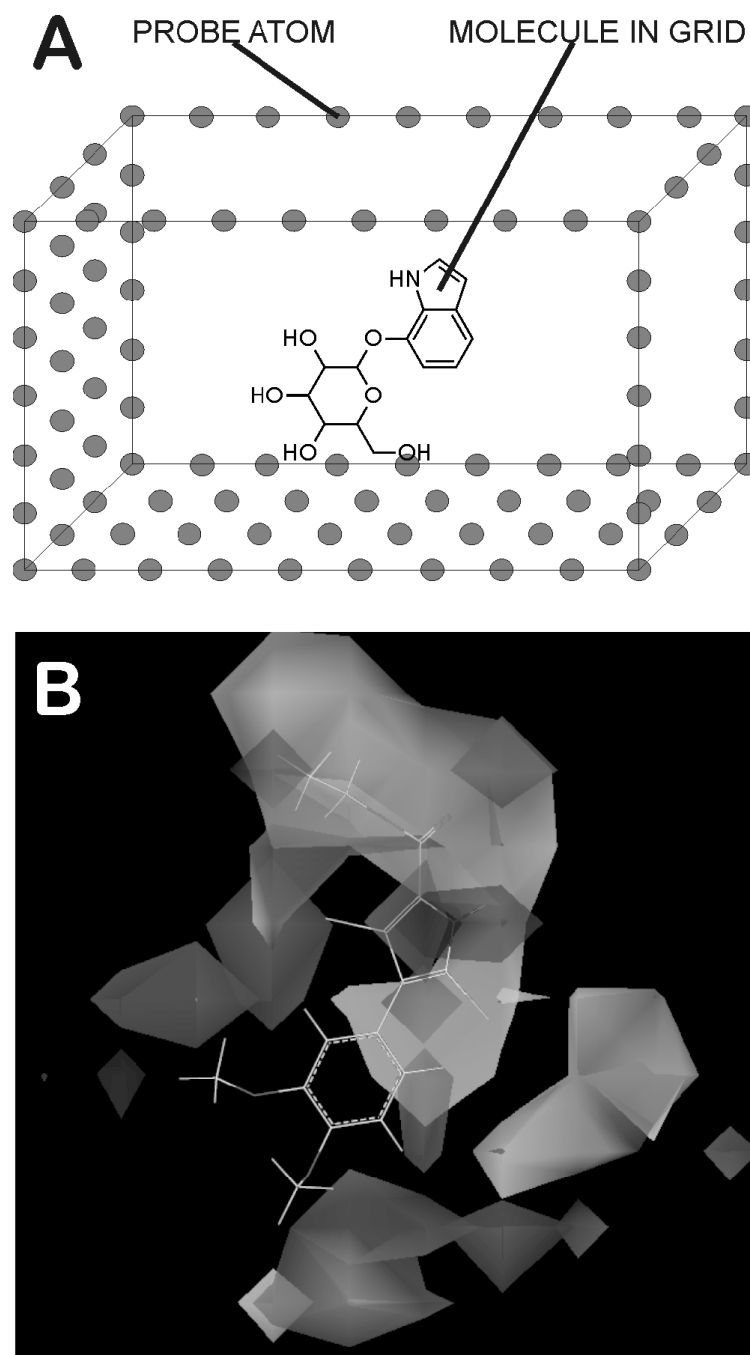


Figure 2. 3D QSAR

3-Dimensional Quantitative Structure Activity Relationships (3D QSAR) are models generated by taking into account the 3-dimensional positions of various physicochemical characteristics of a set of overlapped molecules and the effect they have on drug potency [94,99]. A) Molecules are overlapped and placed in a superimposed grid with extents of several Angstroms in all directions. Each grid point is treated as a probe; neutral carbon atoms probe for Van der Waal's interactions, while charged atoms probe Coulombic interactions. Fragments can also be used as probes for elucidating H-bond donors or acceptors. Simple physics equations calculate the energy (data value) at each grid point. These data are examined for trends using PLS, MLR or AI algorithms fit to the biological properties of the molecules. B) The resultant map indicates

regions where certain physicochemical parameters, e.g., charge or steric bulk, are tolerated (or not tolerated), which serves as an aid to chemists designing analogs with improved properties.

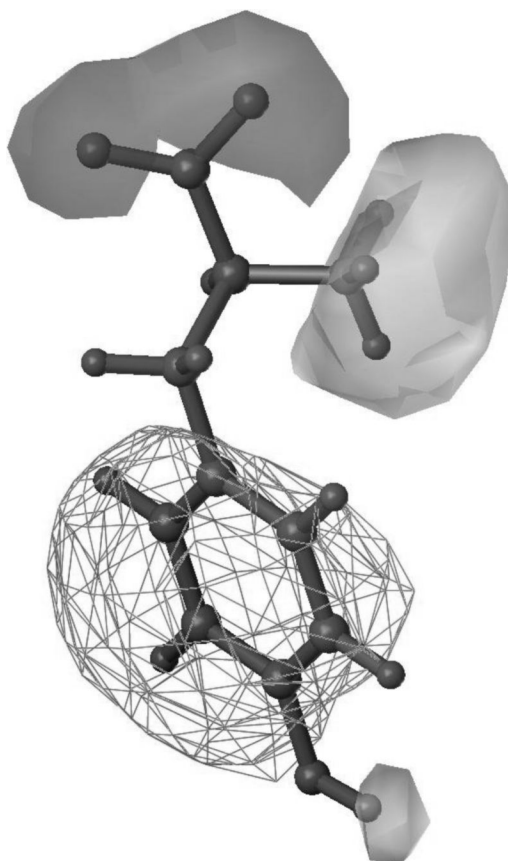


Figure 3. HINT map for the molecule of Tyrosine

This map shows a hydrophobic area on the molecule represented as a cage around the benzene ring. The polar areas on the map are further depicted: acidic (light grey lobes) and basic (dark grey lobe).

Table 1

Various types of methods for LogP calculations This table shows a rough classification of methods used for theoretical prediction of LogP for compounds. Examples of all the different types are included.

Approach	Methodology	Example(s)
Substructure approaches	Fragment-based methods	Rekker's method [22], Leo's C-LOGP method [23-27], ACD/LogP method [28]
	Atom-based methods	XLOGP method [33-35], Ghose-Crippen method [29-32]
Whole molecule approaches	Molecular Lipophilicity Potential and related approaches	MLP [107-108]
	Topology descriptions	MS-WHIP [40]
	Molecular Property descriptions	Toulmin's Δ LogP method [41]