

## EXERCISE II.4

### COMPARING LOG *P* CALCULATIONS BY THE GHOSE–CRIPPEN AND VILLAR METHODS

**Laura Giurato and Salvatore Guccione**

*Department of Pharmaceutical Sciences, University of Catania, viale Andrea Doria 6, Ed. 2 Città  
Universitaria, I-95125 Catania, Italy*

Phone: +39 095 738-4020; Fax: +39 095 443604; E-mail: edufarm@unict.it (LG); guccione@unict.it (SG)

*Scientific computation is not an end itself. It must be implemented  
in the context of problems to be solved.*

#### **Introduction**

Biological activity is not an invariant property of a geometrical arrangement of a subset of ligand atoms, but it is contingent on parameters of the whole molecule and on interactions external to the molecules themselves, i.e., with the receptor.

There are three major forces that are important in biochemical ligand binding: hydrophobic, dispersive, and electrostatic interactions. Molar refractivity is related to dispersive forces, and the molecular orbital charge distribution or the electrostatic potential at the van der Waals radius may be used for modeling the electrostatic interaction. However, the hydrophobic interaction, although probably the most important factor for biochemical interaction, is least understood. The term “hydrophobic interaction” refers to the force or the corresponding energy that operates between two or more nonpolar solutes in liquid water. Although the theoretical work on hydrophobic interactions led to a clear understanding of the molecular structure of aqueous solution, it has hardly begun to build a satisfactory theoretical description of the process that has a wide range of practical applicability. In such a situation, medicinal chemists try to model this interaction using a physicochemical property which closely parallels hydrophobicity, namely, the partition coefficient of the ligand molecules between water and a nonpolar solvent (usually *n*-octanol). This property in fact represents nonregiospecific dispersive and electrostatic forces and the consequent entropic factor.

For many years, the rational design of novel compounds with therapeutic activity was largely based on the use of free energy expressions and regression analysis techniques (QSAR) to relate structural and physicochemical properties of a set of compounds to their activities or affinities at a given binding site. Among other properties, octanol/water partition coefficients led to successful correlations in a variety of fields related to pharmacokinetics and drug design.

The importance of the partition coefficient as a parameter for drug design is due to its relevance to a number of steps in the pathway between the administration of a drug and its biological endpoint such as the drug transport process (*Subcellular Pharmacokinetic* according to Balaz). Understanding the kinetics of ADMET (*A*dministration, *D*istribution, *T*oxicity, *E*xcretion and *T*oxicity) in terms of drug structure and properties is a key step for rational drug development. More-over, log *P* is the only readily accessible physicochemical property that can be related to the entropic change that accompanies the interaction between a drug and a receptor, which in most cases is the dehydration process that precedes it. Finally, it can also serve as a measure of the interaction between the drug and the receptor.

Log *P* is a property exceedingly difficult to measure in the real world. The issue becomes even more complicated if the measurements were done at a pH such that the log *P* had to be corrected for ionization.  $pK_a$  values vary as much as log *P*s, if not more: they are sometimes very temperature-sensitive. Log *D* values measured at the same pH but in different laboratories often vary widely. Another problem concerns compounds that can tautomerize or equilibrate between zwitterion and neutral form. Experimental values jump all over, especially for compounds with high log *P* (usually because of solubility issues, micelle formation, and a troublesome tendency of some

molecules to stick to the vessel walls; this is compounded by the accuracy of the assay procedure itself). So, when a purported method to estimate  $\log P$  claims accuracy to within 1 log unit, it must surely scare and, as any estimation method that claims to be better than experimental error, seems somehow surreal. Using the classic shake-flask method,  $\log P$  measurements may be very time-consuming. Beginning with the “easy” solutes in the mid-1960s, a technician in Hansch’s group usually measured a half-dozen per week, each over at least a five-fold concentration range and with at least three measurements deviating no more than 0.05 log units. But the solutes of current interest are generally much more difficult to measure, especially those of very high or very low  $\log P$  where the solvent ratios are 1000 to 1 or greater.

For extremely lipophilic solutes, it has been shown that measured  $\log P$  values can asymptotically approach a high  $\log P$  value due to the global dissolving effect of octanol in the aqueous phase.

Thus, the accuracy of measurements of these excessively lipophilic solutes is not as much a problem as is the fact that the octanol/water model no longer reflects real-life situations.

Theoretical approaches for  $\log P$  calculation are required to estimate its value for hypothetical structures, such as those that may be designed based on the QSARs developed.

Several methods that allow the computation of partition coefficients have been proposed. Each of these methods varies considerably in its rationale.

The partition coefficient of a solute in two solvents may be approximated as the ratio of its solubilities in the two solvents. The logarithm of partition coefficient, on the other hand, is directly related to the change in free energy during the transfer of the solute from one solvent to the other. One way of understanding the octanol/water partition coefficient is to correlate it with more fundamental physicochemical properties, like molar volume, formal charge density, and polarizability. With some modification, the atomic values of these fundamental properties can then be used to get the atomic contributions to partition coefficient. Although this approach is scientifically attractive, there are several problems with it, particularly the conformational dependency of these fundamental properties for conformationally flexible molecules.

While a variety of different approaches have been proposed in the past for the computation of this property, most of them are based on two-dimensional molecular topology (*vide infra*). This feature renders them inadequate to describe the dependence of hydrophobicity on the three-dimensional arrangement of atoms. However, since flexible compounds can adopt different conformations in solvents of different polarity, this dependence of the hydrophobic character on structure is a significant property. For instance, dependency of the conformation on the environment can be used in conjunction with other properties to identify the bioactive form for a ligand, *i.e.*, the form in which the drug binds to the receptor.

The hydrophobicity of a molecular system depends on the nature of the groups exposed to interaction with the environment, and, therefore, it is dependent on the conformation of the system, whereas  $\log P$  is not. Then, when three-dimensional structural properties of the ligands are included either in QSAR, COMFA analysis, or in mechanistic-based deduction of molecular determinants of receptor recognition, the use of  $\log P$  is insufficient. A simple method to extend its usefulness would be to modify  $\log P$  to relate to structure. However, it is possible to develop this relationship only for rigid analogs since in this case the interactions between the solute and the solvent are uniquely defined. Then, the relation thus established for rigid analogs can be used to characterize the hydrophobicity of particular conformations of flexible analogs. Such a conformationally dependent hydrophobicity parameter would be a useful addition to other properties routinely used to characterize the bioactive structure of a ligand when no structural information about the receptor is known. Without this information, the form of the ligand that is recognized by the receptor, *i.e.*, its bioactive form is then deduced for each ligand based on the similarities in steric and electronic properties for analogs with high binding affinities and on the dissimilarities with all other analogs. The lack of conformationally dependent hydrophobicity criteria prevents the use of this property, among those analyzed, in this crucial step toward the development of a pharmacophore.

The characterization of a bioactive form is already one of the most challenging tasks in the development of pharmacophores for the computer-assisted design of novel drugs. Except for totally rigid analogs, the conformation in which the drug interacts with the receptor does not need to be its minimum energy structure. In most cases, candidate bioactive forms result from the analysis of steric and electronic similarities and dissimilarities among analogs with pharmacological profiles. Hence, inclusion of information relative to influence of conformational dependence on key physicochemical properties of these compounds should facilitate the characterization of the bioactive form.

Hopfinger and Bartell developed a solvent-dependent conformational analysis that allowed the computation of partition coefficients. The method was based on the precise evaluation of the first solvation shell for each solvent, making it computationally expensive.

A variation of the atomic contributions method was developed by Klopman and coworkers: a quantum mechanical parameterization based on the computation of charge densities.

The rationale behind this latter approach is that the partition coefficient depends on the relative solubility of a substrate in a polar and nonpolar solvent. Since solubility in a polar environment can in turn be related to the electrostatic forces that involve charge densities, they postulated that the partition coefficient of a molecule should also be dependent on the charge densities.

More recently, Bodor et al. expanded the method originally developed by Klopman and coworkers to include several other quantities such as the total molecular surface and its globularity, properties that are conformationally dependent, as well as the molecular weight, dipole moment, a polynomial of the sum of the atomic charges per atom, and an empirical parameter that assumed different values depending on the chemical family. Richards and coworkers have also developed a program called HYDRO that allows the computation of a conformation-dependent hydrophobic index. The method consists of a series of fragment transfer free energies, which are a function of the solvent-accessible surface area.

An atom-based method for the computation of a conformationally dependent hydrophobic index ( $p$ ) is that by Villar (see below).

The alternate approach is to express the octanol/water partition coefficient in terms of the chemical structure of the ligand. Rekker et al. first gave some fragmental values for calculating the partition coefficient from the chemical structure of the molecules. Hansch and Leo followed the same direction and gave a thorough list of fragmental values and also a large number of correction factors to account for various intramolecular interactions. The implementation of these latter values in the CLOGP program is the most reliable and widely used technique for the theoretical determination of the partition coefficient. The calculation of the partition coefficient is performed by adding the contributions of molecular fragments. The major task in this method is the adequate definition of the fragments needed to generate a consistent parameterization. It is due to these correction factors that regional contributions toward the partition coefficient are difficult to evaluate.

Alternative methods to determine this property are based on the addition of atomic contributions, such as the methods developed by Broto et al. and Ghose et al. In this approach, tables of contributions for atoms in different topologic environments are used. The major advantage of this technique is that it lends itself to simple automation.

The solvatochromic approach pioneered by Kamlet, Taft, or that by Taylor, Leahy, Abraham points out that other *partitioning pairs* can add valuable information that  $\log P$  octanol lacks.

A good program should provide some “distillation” of the hard-won wisdom contained in some hundreds of papers most researchers will not have read. The future challenge in improving  $\log P(\text{oct})$  calculation programs is to convey to the user as much understanding as possible of what determines the final parameter value, for instance, (i) predicting when a lipophilic environment encourages an intramolecular hydrogen bond to form when the solute is passing through a membrane or in a lipophilic pocket at the active site; (ii) the same sort of prediction when tautomerism is possible, e.g., encouraging the more lipophilic enol over the keto.

We are still a long way off being able to characterize molecules very well at all. There is no “best” set of descriptors, although there are some that we know to be useful, and our feeling is the “best” set will always be problem-dependent—and will probably feature properties from different “sets”.

“What does this calculation tell me about how my promising molecule will behave when acted upon by the physical forces in a biological setting?” If several  $\log P$  prediction programs give similar estimates (not absolute values, but trends: that is some indication of the hydrophobicity of a molecule without having a perfect algorithm) of  $\log P$  for a given compound, then one can have more confidence in these predictions. The most appropriate methods are based on fragmentation schemes, where explicit account is taken of the local environment of these fragments.

Different models are only useful in this context if they contain different information (different views) of the molecules concerned. So one would have to try and assess this point before accepting that similar (or different) predictions for a property by different models actually have some meaning. The added value of a new method for estimating  $\log P$  is difficult to estimate because it is always possible to find failures in comparison exercises. It is not annoying if the models have been well designed and if we keep in mind that a universal simulator does not exist and, consequently, all models fail for specific structures. Although it is desirable to get a perfect algorithm, it should just be remembered that  $\log P$  must give some indication of the hydrophobicity of a molecule.

With this preamble, suppose now that we want to estimate the  $\log P$  value of a new chemical. Combinations can be used, but the idea is to use models well designed and based on different methodologies. If these different models provide a rather similar  $\log P$  value for our new chemical, we can suppose that the simulation is acceptable. Conversely, high differences act as a warning message on the quality of the simulation.

Indeed, nowadays, by the combined use of models based on different methodologies, it is possible to secure simulation results, while 20 years ago it was not possible. If this difference is important for QSPR models (e.g.,  $\log P$ , solubility), it is crucial in QSAR for toxicological endpoints (e.g., mutagenicity, carcinogenicity) where the consequences of bad simulations can be dramatic for the launch of new molecules or for health.

In that case, the judicious use of different models for the same endpoint allows us to reduce the problems of *false positives* and *false negatives*. One question to ask is what are we estimating  $\log P$  (or whatever) for? If we just have a few molecules, then one might as well just go and measure it. While in the context of CombiChem, *in silico* screening we want quick estimates of  $\log P$  to help choose/eliminate structures. In this case, really accurate  $\log P$  estimates are not necessary, we just need an approximation to test against the rule of five or something similar; we also need some indication of when the estimate is likely to be very wrong (back to ClogP again). Of course, in terms of science, it is interesting to be able to calculate accurately and precisely  $\log P$ , but if one works in industry good science is a tool rather than an end in itself.

Consequently, it is desirable that the number of new QSAR and QSPR models designed from very different methodologies will continue to increase in the future.

Furthermore, the risk with all these “property prediction” modules that are now more and more advertised is that in the end they produce meaningless numbers.

### ***Isosterism, Bioisosterism, and Bioanalogy***

Confusion and fundamental problems with the continued use of both the terms “isosterism” and “bioisosterism” remain, which calls for their revision and/or renaming. The word *bioanalogue* was proposed (Floersheim et al., 1992) to take the place of *bioisostere* and adopt the definition of Hansch, but expands its scope to accommodate groups or partial molecular structures as well as entire molecules. Thus, *bioanalogs* are molecules or groups which, in the context of a given biological parameter, elicit *analogous* responses. For this term, the structural connotations present in the previous term *bioisostere* are removed. Ascribing a *bioanalogous* relationship to different partial molecular structures or groups is associated with problems. Thus, for a given set of groups, *bio-*

*analogy* may not necessarily be conserved over a wide range of molecules even with regard to a single biological parameter. Nevertheless, the concept of interchangeable groups which are potentially *bioanalogous* with regard to a single or many biological parameter(s) is indisputably very useful. Indeed, for the medicinal chemist, the greatest potential of the concept lies in systematic searching of a database to find group replacements which, starting from a lead molecule, could yield novel or useful active molecules.

The term “isostere” is so widespread that it can be maintained and continue to be used but only with regard to structure. It should, however, be redefined to be consistent with the current meaning of the adjective steric. Thus, isosteric literally means identical in size and shape, but to be useful, isosteric must be *subjectively* defined to mean similar in size and shape. With this classification, the terms “isosteric” and “non-isosteric *bioanalogs*” were proposed (Floersheim et al., 1992) to replace the terms “classical” and “nonclassical bioisosteres”.

Any consideration of isosterism has to take conformational parameters into account. Clearly, two conformations of the same molecule will not necessarily be isosteric; both, however, may be bioanalogous, if they can elicit similar biological activity.

Thus, it must be considered whether a chemical modification of a compound affects its conformational preferences, and through this mechanism, its biological activity. An additional difficulty may arise in attempting to relate molecules with analogous biological activity when racemic mixtures are used. As has been amply recognized by others, enantiomers interacting with chiral biomolecules may differ greatly in their activities. Enantiomers are not isosteric, as they are by definition non-superimposable, but they may be bioanalogous. This is in agreement with *Pfeiffer's* rule, which refers to the generalization that the greater the biological activity of the racemate, the larger the difference in the activity of the enantiomers. This could be interpreted as the effect of a closer interaction between each of the enantiomers and the receptor, enhancing the stereodifferentiation. In other cases, however, *Pfeiffer's* rule does not hold.

In conclusion, for constructing structure–activity relationships where biological activity is a consequence of a molecular recognition process, while certain empiricisms may be useful, structural considerations grounded in first principles are of prime importance. Thus, isosterism, carefully defined on purely structural considerations, may point the way to bioanalogous molecules with a desirable pharmacological profile.

- Floersheim P., Pombo-Villar E., Shapiro, G. “Isosterism and Bioisosterism Case Studies with Muscarinic Agonists”, *Chimia* **46**, 323–334 (1992) and references cited therein.

## Material and Methods

Molecular mechanics calculations and quantum chemical calculations play a multiple role in modern-day computational chemistry. Molecular mechanics calculations on complex molecules may even be performed on personal computers and have spread widely throughout the chemical community. Quantum chemical calculations, even semiempirical molecular orbital calculations, but especially *ab initio* molecular orbital calculations and density functional calculations, are much more time-demanding. Recently, with the availability of fast workstations and efficient graphics-based programs, these methods have begun to be widely applied.

Quantum chemical methods are also useful for furnishing information about the mechanisms and product distributions of chemical reactions, either directly by calculations on transition states, or indirectly by modeling the steric and electronic demands of the reactants. Quantum chemical calculations are also able to supply information needed as input for other techniques, for example, atomic charges for QSAR analyses. *Ab initio* Hartree–Fock and correlated molecular orbital calculations, and density functional calculations are also able to provide accurate intra- and intermolecular potentials. This kind of information is required both by molecular mechanics and by molecular dynamics techniques used to describe a wide variety of phenomena, ranging from interactions between an enzyme and a drug to the physical properties of polymeric materials.

SPARTAN is a program that embraces molecular mechanics as well as ab initio and semiempirical molecular orbital and density functional methods. SPARTAN is intended to provide a convenient environment to carry out individual molecular mechanics calculations, molecular orbital calculations, and density functional calculations on diverse molecular systems.

SPARTAN presently comprises seven independent program modules: a graphical user interface and ab initio, density functional, semi-empirical, mechanics, properties, and graphics modules.

SPARTAN's architecture makes a clear separation between **tasks** and **methods**. Tasks indicate what is to be done (e.g., perform a geometry optimization or search conformational space), while methods dictate how the tasks are to be done (e.g., use MMFF molecular mechanics or the PM3 semiempirical method). Both tasks and methods are specified in the **graphical user interface**. SPARTAN's **graphical user interface** provides a number of functions, among them: construction and editing of molecular structures, preparation of input designating the quantum chemical or molecular mechanics calculation to be performed by the **ab initio**, **density functional**, **semiempirical**, or **mechanics** modules; the preparation of input for Gaussian 94, preparation of input designating molecular properties to be calculated using the **properties** module.

SPARTAN's **ab initio** module provides for calculation of the energy and wavefunction for a given nuclear configuration, of equilibrium or transition-state geometries and of normal-mode vibrational frequencies. The module is presently limited to Hartree–Fock and MP2 correlated models, for both closed- and open-shell systems. SPARTAN's **semi-empirical** module provides for the calculation of heats of formation, equilibrium, and transition-state geometries and normal-mode vibrational frequencies, as well as for searching of conformation space of both acyclic and cyclic molecules. The MNDO (**M**odified **N**eglect of **D**ifferential **O**verlap), AM1, and PM3 models are supported.

### ***AM1 and PM3 Semi-empirical Models***

Ab initio and semi-empirical models are based on the Hartree–Fock set of ideas. Both the “paradigms” are strongly rigorous in regards to the solution of the Schrödinger equation, but significant approximations are made to make the semi-empirical methods faster than their ab initio counterparts. Semi-empirical methods are simplified versions of the Hartree–Fock theory using empirical methods that are derived from experimental data corrections in order to improve performance. These methods are usually referred to through acronyms encoding some of the underlying theoretical assumptions. The most frequently used methods (MNDO, AM1, PM3) are all based on the Neglect of **D**ifferential **D**iatomic **O**verlap (NDDO) integral approximation, while older methods use simpler integral schemes such as CNDO and INDO. All the three approaches belong to the class of **Z**ero **D**ifferential **O**verlap (ZDO) methods, in which all two-electron integrals involving two-center charge distributions are neglected. A number of additional approximations are made to speed up calculations, and a number of parameterized corrections are made in order to correct for the approximate quantum mechanical model. How the parameterization is performed characterizes the particular semi-empirical method. For MNDO, AM1, and PM3, the parameterization is performed such that the calculated energies are expressed as heats of formations instead of total energies.

Both AM1 (**A**ustin **M**odel **1**) and PM3 (**P**arameterization **M**ethod **3**) are based on the basic NDDO theory by Michael Dewar at the University of Texas, Austin. The substantial difference between the two methods is in the parameters used to partly replace the full ab initio implementation of the Hartree–Fock theory and in the less pronounced *chemical sense* of the PM3 model, built on a largely undirected mathematical optimization process, when compared to AM1. The stronger chemical character of the AM1 method reflects on the better *external* performing of its parameters, that is, the capability to yield useful results for situations not specifically included in the molecular basis set for parameterization (MBSP). Generally, AM1 represents differences between compounds more reliably. PM3 works very well, for instance, when nitro-derivatives, extensively parameterized in the MBSP, are calculated. PM3 performs sometimes better for geometries that are guessed

with high precision. A big drawback in PM3 concerns the charges as a consequence of quite *fuzzy* parameter values. Based on our experience, both models perform in the same way on data such as heats of formation, while problems were encountered with the PM3 model in minimum energy conformations. In terms of the actual NDDO model, PM3 allows most of the parameter values to float, resulting in substantially more parameters whether AM1 has a different “view” in the application of the Gaussian functions considered as patches and introduced to adjust the core-electron/core-electron repulsion function. The addition of Gaussian functions to some of the core repulsion functions (CRFs) for certain elements is a substantial difference between the two models. These functions have a position, a width, and an intensity, all of which can be taken as parameters. They were initially intended to adjust the SHAPE of the CRF so that it would more closely correspond to “reality”, whatever that is. In essence, they were used as patches for specific types of systems that could not be handled in general parameterization without disrupting other chemically important items. To avoid failures common in all the models, PM3 includes de novo two Gaussian functions for each element, which strongly belittles the “chemical” reality.

- Dewar, M.J.S., Healy, E.F., Holder, A.J., Yuan, Y-C. “Comments on a comparison of AM1 with the recently developed PM3 method”, *J. Comp. Chem.* **11**, 541–542 (1990).
- Holder, A.J., Dennington, R.D., Jie, C. “Addendum to SAM1 results previously published”, *Tetrahedron* **50**, 627–638 (1994).
- Stewart, J.J.P. “MOPAC: A Semiempirical Molecular Orbital Program”, *J. Computer-Aided Mol. Des.* **4**, 1–105 (1990).
- Gundertofte, K., Palm, J., Pettersson, I., Stamvik, A. “A comparison of conformational energies calculated by molecular mechanics (MM2 (85), Sybyl 5.1, Sybyl 5.21, and ChemX)—and semiempirical (AM1 and PM3) methods”, *J. Comp. Chem.* **12**, 200–208 (1991).

Among SPARTAN’s **properties** module’s functions, there is the calculation of  $\log P$ .

Two  $\log P$  models are available: **Villar** and **Ghose–Crippen**. The first is a model that only works on semi-empirical wavefunctions with no d-orbitals. The Villar method examines the overlap matrix, searching for the type and number of lone pairs as well as the surface area of each atom. It is parameterized for H, C, N, O, F, S, and Cl.

Ghose–Crippen is the SPARTAN default method of calculating  $\log P$ . This method depends only on the connectivity of the molecule, and it is independent of the wavefunction (i.e., one will get the same results for semi-empirical, HF, and DFT methods but this depends on how the molecule is drawn/connected). The Ghose–Crippen model is parameterized for 110 atom types, including common bonds of H, C, N, O, S, and the halogens. Avoiding correction factors was obtained evaluating the hydrophobicity on an individual atom basis, accounting for the undeniable intramolecular interactions by employing a large number of atom types.

The Villar method is an alternative atom-based method for the computation of a conformationally dependent hydrophobic quantity ( $p$ ). The parameters used in this method are the molecular surfaces and atomic charges, both of which have some dependency on the conformation adopted by the system, as well as a set of adjustable parameters that only depend on the atomic number. These adjustable parameters were determined by linear regression using experimental values of the octanol/water partition coefficient.

The quantity  $p$  differs from the actual octanol/water partition coefficient, which is a macroscopic property. Only if one conformer is accessible as in rigid compounds, are these two quantities equal. Even though the calculation of partition coefficients can be made using the hydrophobic indices described here, their computation would require a cumbersome statistical averaging for flexible analogs. In general, this procedure is not recommended for the computation of the property when the methods proposed in the past can provide similar or better accuracies without the computational effort. One possible exception could be some types of *isomery* (see above, **Isosterism**, **Bioisosterism**, and **Bioanalogy** and the exercise below) which current methods for calculating

partition coefficient do not distinguish. Nevertheless, the Villar hydrophobic indices are useful in themselves since they can provide important additional insights into drug–receptor interactions, leading to the characterization of a bioactive form.

### Description (exercise)

Log *P* calculation for the compounds in **Table I**.

The first thing you have to do is draw the molecules. FIRST STEP:

Using SPARTAN software, build the molecules reported in **Table I** assigning the right (*R* or *S*) chirality. You can also retrieve the one of the enantiomers from the SMD (Spartan Molecular Data-base).

Make a copy of the enantiomeric entry and paste it into a list by:

**Edit > Copy -- File > New Molecule -- then Edit > Paste**

Go into the build mode [+] and holding down the CTRL key, double click on the Chiral Centre to INVERT from *R* to *S* (or *S* to *R*). Structure can also be retrieved from the CSD (**C**ambridge **S**tructural **D**atabase). An interface UNIX to the Cambridge Structural Database (CSD)\* has been provided in SPARTAN. This provides access to >225 K experimental X-ray crystal structures for organic and organometallic molecules together with their literature references.

In case you should download the ibuprofen from the CSD, assign the chirality to the two entries using the Fischer rule and retrieve the one that is correct (see the footnote in **Table II**).

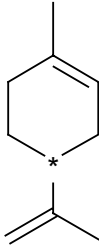
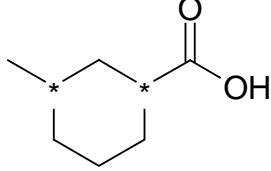
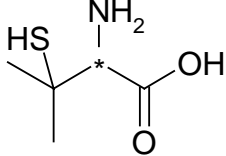
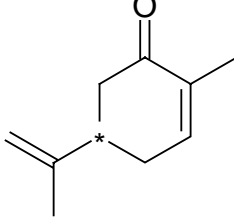
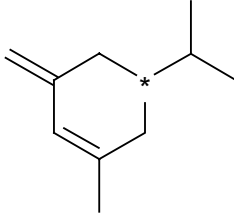
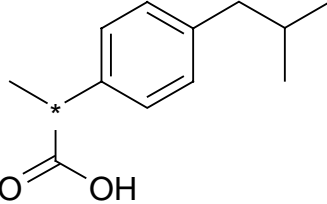
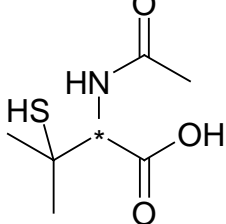
Then, you have to obtain the equilibrium conformer for every drawn structure using “Equilibrium Geometry” at the “lowest” QM level, which is a Hartree–Fock calculation with the 3-21G basis set. Start from the MMFF (**M**erck **M**olecular **F**orce **F**ield) conformer and the MMFF geometry. Once the calculation has completed, you can view and/or print text and/or graphical output (Display → Output). SPARTAN keeps track of what has already been done and will not repeat calculations unnecessarily. While one (or more) job is running, input for another job can be constructed or output corresponding to yet another examined. These jobs may be derivative (i.e., based on information resulting from earlier jobs) or completely independent.

“LogP methods” can be selected with the **LOGP=keyword: LOGP=VILLAR** and **LOGP=GHOSE**, respectively.

Now you can submit the job as a Single Point Energy calculation using Semi-Empirical at AM1 or PM3 level and starting from the initial geometry (this is from the HF calculation), one time writing LOGP=VILLAR as keyword, another time LOGP=GHOSE.



Table I<sup>(a)</sup>

| Molecule   | Formula  |
|--|--|
| Limonene   |    |
| <i>Trans</i> -3-Methylcyclohexanecarboxylic acid |    |
| Penicillamine                                    |    |
| Carvone  |   |
| Homocamfin                                       |  |
| Ibuprofen  |  |
| <i>N</i> -acetyl-penicillamine                   |  |

<sup>(a)</sup>Asymmetric carbon is labeled by an asterisk.

Table II<sup>(1)</sup>

| Molecule  | Ghose/Crippen | Villar |
|---|---------------|--------|
| Limonene <i>S</i>   | 3.01          | 3.75   |
| Limonene <i>R</i>   | 3.01          | 3.76   |
| <i>Trans</i> -3-Methylcyclohexanecarboxylic acid <i>S,S</i> | 2             | 1.10   |
| <i>Trans</i> -3-Methylcyclohexanecarboxylic acid <i>R,R</i> | 2             | 1.12   |
| Penicillamine <i>S</i>                                      | -0.38         | -0.50  |
| Penicillamine <i>R</i>                                      | -0.38         | -0.30  |
| Carvone <i>S</i>  | 2.41          | 2.33   |
| Carvone <i>R</i>  | 2.41          | 2.38   |
| Homocamfin <i>S</i>   | 2.84          | 2.33   |
| Homocamfin <i>R</i>   | 2.84          | 2.34   |
| <i>N</i> -acetyl-penicillamine <i>S</i>                     | -0.45         | 0.30   |
| <i>N</i> -acetyl-penicillamine <i>R</i>                     | -0.45         | 0.27   |
| Ibuprofen <i>S</i> <sup>(2)</sup>                           | 3.72          | 3.75   |
| Ibuprofen <i>R</i> <sup>(2)</sup>                           | 3.66          | 3.75   |

<sup>(1)</sup>Using the Ghose–Crippen and the Villar methods in different environments or their implementation in different SPARTAN versions may lead to different absolute values but the trend should be the same (see text).

<sup>(2)</sup>Every time you use crystallographic files carefully check their correctness (see the papers by Cozzini et al. and Spyrikis et al.). An ibuprofen entry with a wrong chirality is deposited in the CSD. \*CSD needs to be licensed from the Cambridge Crystallographic Data Centre or one of its distributors.

As you can see, the Villar method is quite discriminating of the two enantiomers of the same molecule, while the Ghose–Crippen method gives the same value in both cases.

1) Compare and comment on the results, observing how hydrophobicity changes according to the different structure and substituent chemical nature. Check for possible changes due to the different “*semi-empirical*” (AM1 or PM3) used.

2) Search the Internet for databases (see references) from which you can download experimental log *P* values. Whether it is possible (the available files are in a “chemical format”) to cover the dataset to a format suitable to be read in SPARTAN or “de novo”, sketch the structure in a suitable format for SPARTAN. If necessary, you can download the software BABEL (<http://www.eyesopen.com/products/applications/babel.html>). Be sure all the log *P* values you use to build your dataset are *homogeneous*, which is determined by the same or quite close methods (*homogeneity* also applies to biological activity tests, which is one of the prerequisites to make consistent not “illusive” QSARs). Split your dataset, selecting rigid structures for calculations by the Ghose–Crippen method and conformationally flexible structures for the ones by the Villar method. Calculate the log *P*s using both the Ghose–Crippen and the Villar methods. Prepare a diagram of experimental vs. calculated values (Microsoft Excel can be used) and find for the best method to calculate your *split* dataset. Validate your “model”, mixing rigid and flexible structures and calculate again using both the methods. Log *P*s for flexible structures should be worst calculated by the Ghose–Crippen method. Extend and compare the analysis by the HINT software (see the paper

**Practice I.12** by Cozzini et al. and **Practice III.1** by Spyrakis et al., respectively). You need to convert the files to the pdb or mol2 format for HINT. The differences in values do not matter, provided that the overall trend is the same. Partition coefficient is some indication of the hydrophobicity of a molecule.

“Take home” messages: *Scientific computation is not an end itself. It must be implemented in the context of problems to be solved*, therefore, “What does this calculation tell me about my problem...that is, how much information will it give me?”

## Bibliography

- Dobson, C.M. “Chemical space and biology”, *Nature* **432**, 824–828, 2004 and quoted articles in the same issue and the references list.
- Scherrer, R.A. In *Pharmacokinetic Optimizations in Drug Research*, B. Testa, H. van de Waterbeemd and G. Folkers (Eds.), Verlag HCA, Zurich, 2001, 351–381.
- Scherrer, R.A., and Crooks, S.L. “Titrations in water-saturated octanol: a guide to partition coefficients of ion pairs and receptor-site interactions”. Fauchere, J.L. (Ed.). *Quant. Struct.-Activ. Rel. Drug Design* **8**, 59–62 (1989).
- Scherrer, R.A. “The treatment of ionizable compounds in QSAR studies with special consideration of ion partitioning”. In: Magee, P.S. et al. (Eds.), *Pesticide Synthesis through Rational Approaches*, ACS Symposium Series 225, ACS: Washington DC, 1984, 225–246.
- *Pharmacokinetic Optimization in Drug Research: Biological, Physicochemical, and Computational Strategies*. B. Testa, H. van de Waterbeemd, G. Folkers, R. Guy (Eds.), March 2001, WILEY-VCH.
- *Pharmacokinetics and Metabolism in Drug Design*. D. A. Smith, D. K. Walker, H. van de Waterbeemd, R. Mannhold, H. Timmerman, H. Kubinyi (Eds.), October 2001, WILEY-VCH.
- *Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability*. H. van de Waterbeemd, H. Lennernäs, P. Artursson, R. Mannhold, H. Kubinyi, G. Folkers (Eds.), June 2003.
- *Drug-Membrane Interactions: Analysis, Drug Distribution, Modeling*. J. K. Seydel, M. Wiese, R. Mannhold, H. Kubinyi, G. Folkers (Eds.), March 2002, WILEY-VCH.
- Balaz S. “Lipophilicity in trans-bilayer transport and subcellular pharmacokinetics”, *Persp. Drug Discovery Design* **19**, 157–177 (2000) and references cited therein.
- Abraham, D.J., Leo, A.J. “Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients”, *Proteins* **2**, 130–152 (1987).
- Muñoz-Muriedas, J., Perspicace, S., Bech, N., Guccione, S., Orozco, M., Luque, F.J. “Molecular Alignment Based on Hydrophobic Similarity as a Tool in Drug Design”, *J. Computer-Aided Mol. Des.* **19**, 401–419 (2005) and references cited therein.
- Vicentini, C.B., Guccione, S., Giurato, L., Ciaccio, R., Mares, D., Forlani, G. “Pyrazole Derivatives as Photosynthetic Electron Transport Inhibitors: New Leads and Structure-Activity Relationship”, *J. Agric. Food Chem.* **53**, 3848–3855 (2005).
- Uccello-Barretta, G., Balzano F., Paolino, D., Ciaccio, R., Guccione, S. “Combined NMR-crystallographic and modelling investigation of the inclusion of molsidomine into  $\alpha$ -,  $\beta$ - and  $\gamma$ -cyclodextrins”, *Bioorg. Med. Chem.*, in press.
- Hopfinger, A. J., Battershell, R. D. “Application of SCAR to drug design1. Prediction of octanol-water partition coefficients using solvent dependent conformational analyses”, *J. Med. Chem.* **19**, 569–573 (1976).
- Klopman, G., Iroff, L.D., “Calculation of partition coefficients by the charge density method”, *J. Comput. Chem.* **2**, 157–160 (1981).
- Klopman G., Namboodiri, K., Schochet, M. “Simple method of computing the partition coefficient”, *J. Comput. Chem.* **6**, 28–38 (1985).

- Bodor, N., Gabanyi, Z., Wong, C.-K. "A new method for the estimation of partition coefficient", *J. Am. Chem. Soc.* **111**, 3783–3786 (1989).
- Bodor, N., Huang, M.-J. "Predicting partition coefficients for isomeric diastereoisomers of some tripeptide analogs", *J. Comput. Chem.* **12**, 1182–1186 (1991).
- Richards, N.G.J., Williams, P.B., Tute, M.S., "Empirical methods of computing molecular partition coefficients: II. Inclusion of conformational flexibility within fragment-based approaches", *Int. J. Quantum Chem.* **44**, 219–233 (1992).
- Alkorta, I., Villar, H.O. "Quantum Mechanical Parameterization of a Conformationally Dependent Hydrophobic Index", *Int. J. Quant. Chem.* **44**, 203–218 (1992) and references cited therein.
- Kantola A., Villar, H.O., Loew, G.H. "Atom Based Parametrization for a Conformationally Dependent Hydrophobic Index", *J. Comp. Chem.* **12**, 681–689 (1991) and references cited therein.
- Rekker, R.F. *The Hydrophobic Fragmental Constant*, Elsevier, New York (1977).
- Hansch, C., Leo, A.J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York (1977).
- Broto, P., Moreau, G., Vanduycke, C. "Molecular Structure: perception, autocorrelation descriptor and SAR studies", *Eur. J. Med. Chem.* **19**, 71–78 (1984).
- Ghose A.K., Pritchett A., Crippen, G.M. "Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity relationships III: Modelling Hydrophobic Interactions", *J. Comp. Chem.* **9**, 80–90 (1988) and references cited therein.
- Platts, J.A., Abraham, M.H., Hersey, A., Butina, D. "Estimation of molecular linear free energy relationship descriptors by a group contribution approach. 2. Prediction of partition coefficients", *J. Chem. Inf. Comp. Sci.* **40**, 71–80 (2000).
- Leahy D.E, Taylor P.J., Wait, A.R. "QSAR: Rational Approaches to the design of Bioactive Compounds", Silipo & Vittoria (Eds.), Elsevier, Amsterdam, 75–82 (1991).
- User's Guide and Tutorial and references therein included. SPARTAN '02 (UNIX version) or '04 WINDOWS, Wavefunction, 18401 Von Karman, Suite 307, Irvine, CA 92612, <http://www.wavefun.com/>.
- Experimental log *P* values can be found at:  
[http://www.syrres.com/esc/est\\_kowdemo.htm](http://www.syrres.com/esc/est_kowdemo.htm)  
[http://www.acdlabs.com/pro\\_ucts/phys\\_chem\\_lab/logp/](http://www.acdlabs.com/pro_ucts/phys_chem_lab/logp/)
- A Guide to Molecular Mechanics and Quantum Chemical Calculations, W.J. Hehre, Wavefunction Inc., 2003, <http://www.wavefun.com/>
- Doweyko, M.A. "3D-QSAR illusions", *J. Computer-Aided Mol. Des.* **18**, 587–596 (2004).
- See the papers by Cozzini et al. (Practice I.12) and Spyrakis et al. (Practice III.1) for further useful readings on log *P*.

## Notes

The authors can be contacted for each additional information and support.

Wavefunction Inc. is gratefully acknowledged for having provided SPARTAN'04 (WINDOWS version). Laura Giurato thanks Dr. Sean Ohlinger (Vice President of Business Development, Wavefunction, Inc.) for helpful discussion and suggestions.

**Salvatore Guccione**  
 guccione@unict.it

High standards in safety measures should be maintained in all work carried out in Medicinal Chemistry Laboratories.

The handling of electrical instruments, heating elements, glass materials, solvents and other inflammable materials does not present a problem if the supervisor's instructions are carefully followed.

This document has been supervised by Prof. Salvatore Guccione (guccione@unict.it) who have informed that no special risk (regarding toxicity, inflammability, explosions), outside of the standard risks pertaining to a Medicinal Chemistry laboratory exist when performing this exercise.

If your exercise involves any "special" risks, please inform the editor.