Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 551–557
Bioorganic & Medicinal Chemistry Letters

Unified Markov thermodynamics based on stochastic forms to classify drugs considering molecular structure, partition system, and biological species: distribution of the antimicrobial G1 on rat tissues

Humberto González-Díaz,[a,b,*] Guillermin Agüero,[c] Miguel A. Cabrera,[b,d,e] Reinaldo Molina,[b,f] Lourdes Santana,[a] Eugenio Uriarte,[a] Giovanna Delogu[g] and Nilo Castañedo[b]

[a]*Department of Organic Chemistry, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain*
[b]*Department of Drug Design, CBQ, Central University of Las Villas, 54830 Santa Clara, Cuba*
[c]*Faculty of Pharmacy and Chemistry, CAP, Central University of Las Villas, 54830 Santa Clara, Cuba*
[d]*Department of Pharmacy and Pharmaceutical Technology, University of Valencia, Valencia, Spain*
[e]*Innovation Place Research Park, 3-411 Downey Road, Saskatoon, Saskatchewan, Canada S7N 4L8*
[f]*Universität Rostock, FB Chemie, Albert-Einstein-Str. 3a, D 18059 Rostock, Germany*
[g]*Dipartimento Farmaco Chimico Tecnologico, Universitá Degli Studi di Cagliari, Via Ospedale 72, 09124 Cagliari, Italy*

Received 7 October 2004; revised 15 November 2004; accepted 18 November 2004
Available online 22 December 2004

Abstract—To date, molecular descriptors do not commonly account for important information beyond chemical structure. The present work, attempts to extend, in this sense, the stochastic molecular descriptors (González-Díaz, H. et al., *J. Mol. Mod.* **2002**, *8*, 237), incorporating information about the specific biphasic partition system, the biological species, and chemical structure inside the molecular descriptors. Consequently, MARCH-INSIDE molecular descriptors may be identified with time-dependent thermodynamic parameters (entropy and mean free energy) of partition process. A classification function was developed to classify data of 423 drugs and up to 14 different partition systems at the same time. The model has shown a high overall accuracy of 92.1% (293 out of 318 cases) in training series and 90% (36 out of 40 cases) in predicting ones. Finally, we illustrate the use of the model by predicting a high probability (%) for G1 (a novel antibacterial drug) to undergo partition on different biotic systems (rat organs): liver (97.7), spleen (97.5), lung (97.4), and adipose tissue (97.6). These theoretical results coincide with herein reported steady state plasma concentrations (c) and partition coefficients (P) in liver ($c = 42.25 \pm 7.86/P = 4.75$), spleen ($11.47 \pm 4.43/P = 1.29$), lung ($17.04 \pm 3.58/P = 1.91$), and adipose tissue ($28.19 \pm 11.82/P = 3.17$). All values were relative to ^{14}C-labeled-radioactive-G1 in plasma ($c = 8.9 \pm 3.05$) after 3 h of oral administration. In closing, the present stochastic forms derive average thermodynamic parameters fitting on a more clearly physicochemical framework with respect to classic vector–matrix–vector forms, which include, as particular cases, quadratic forms such as Wiener index, Randic invariants, Zagreb descriptors, Harary index, Balaban index, and Marrero-Ponce quadratic molecular indices.

1. Introduction

Classifying organic chemicals and drugs according to their behavior on biphasic system partition phenomena constitutes one of the foremost problems for industrial, pharmaceutical, bioorganic, and medicinal chemistry nowadays. This importance relies on the broad distribution of such phenomena. For instance, partition of organic chemicals in organic solvent/water systems is of major importance in bioorganic chemistry when selecting the correct extraction strategy.[1] While much work has been done by many groups in developing models for water, as reviewed elsewhere,[2–4] much less effort has been devoted to developing models for nonaqueous solvents.[5–11] However, a large body of data is available for 1-octanol[12] and *n*-hexadecane.[13,14] A significant

Keywords: Markov model; QSAR; Entropy; Free energy; Partition coefficients.
* Corresponding author. Tel.: +53 42281473; fax: +53 42281130; e-mail: humbertogd@vodafone.es

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.11.059

amount is also available for other alkanes, cyclohexane, benzene, toluene, xylenes, diethyl ether, chloroform, carbon tetrachloride, and chlorobenzene.[15]

Another interesting partition system is the so-called aqueous biphasic system (ABS) formed by the addition of two (or more) water-soluble polymers or a polymer and salt to an aqueous solution above certain critical concentrations or temperatures. Due to its highly aqueous and hence mild nature, which is consonant with the maintenance of macromolecular structure, ABS has been employed for the separation of biological macromolecules for over 40 years.[16,17] ABS systems are also used in industrial biotechnology quality control for the detection of denaturation and degradation of proteins.[18] Recently, partitioning in ABS has found applications in many different fields of science and technology due to ABS media being non-volatile, non-toxic, and non-flammable. ABS thus represents unique alternatives to traditional solvent-based biphasic systems for the separation of metal ion species,[19–22] small organic molecules,[23,24] and lignins from cellulose in the paper and pulping process.[25,26] ABS partition coefficients have been suggested as viable alternatives to log P values in QSAR (quantitative structure–activity relationships) applied to biotechnological products. Proteins as biomacromolecules and nucleic acids are prone to denaturation in alkyl alcohols, whereas suitable polymer/water compositions can more closely resemble the native living cell conditions.[27]

The classifying of pollutants according to their partition in water/soil systems can be of major interest too, especially for environmental sciences.[28] However, in addition to providing some indication of how likely it is for the solute to penetrate an organic system or suffer partition on ABS it is very interesting for medicinal chemistry studies to classify drugs according to their capacity to penetrate lipid bilayer, skin, brain, central nervous system, or other biophase or to bind to a nonpolar site in or on a protein.[29,30]

If one takes data from all possible systems, a very large number of data are available. Unfortunately, the data come from highly diverse sources and have been determined by very different methods with different errors. So instead of directly predicting partition coefficients it may be more consistent in terms of accuracy to classify molecules into two groups namely: partition-like (PL = 1, when partition coefficient >1) and partition-unlike compounds (PL = −1, otherwise) for a specific system. In this article, it is shown that it is possible to analyze these data as a whole and develop a model that encompasses a large number of partition systems and drugs in a single stochastic framework. For the sake of simplicity all the partition coefficients were written in such a way that partition phenomena are studied from a more organized system to a less organized one, that is, from water to air, or from tissue to plasma.

In this sense, we adapted former stochastic molecular descriptors to describe not only atom connectivity (chemical structure) but the partition phenomena for a specific system. Stochastic molecular descriptors based on the theory of Markov Chains (MC)[31] have been largely used by our group. This method called the **Mar**kovian **Ch**emicals **In Si**lico **De**sign (**MARCH-INSIDE**) approach has been applied in bioinformatics, proteomics, virology as well as in bioorganic medicinal chemistry for the rational design of anticoccidial, fluckicidal, and anticancer drugs too. Briefly, the present generalization of MARCH-INSIDE methodology considers as states of the MC, any atom, nucleotides, or aminoacids in the molecule in dependence of the kind of molecule to be described: small-to-medium sized drug, a nucleic acid, or a protein, respectively.[32–43] Due to this work dealing with small-to-medium sized molecules we are going to use in the present definition only atoms from now on.

One can consider a hypothetical situation in which a molecule is, as a whole, on one side of the partition system at an arbitrary initial time (t_0). In this case, it is of interest to develop a simple stochastic model for the step-by-step passing of all the atoms inward to the other phase, that is, liquid, air, or tissue, of the partition system. It can be supposed that, after this initial situation, the atoms begin to pass throughout the interface at discrete intervals of time (t_k with $k = 0,1,2\ldots$).[33,35–37,40]

Thus, by using the MC theory it is possible to develop a simple model of the absolute probabilities ${}^{A}p_k(j)$, entropies $\Theta_k(j)$, and atomic free partition energy $\Delta\mathbf{G}_k(j)$ with which the atoms a_j pass to the other phase of the partition system in subsequent intervals of time t_k. Thence, we can write classic MC models that may enable us to calculate the average free partition energy $\Delta\mathbf{G}_k$ (Eq. 1) and the average entropy Θ_k (Eq. 2) with which specific molecules undergo partition in a given system until a stationary or steady-state distribution arises:[32–43]

$$\Delta\mathbf{G}_k = {}^{A}\boldsymbol{\varphi}_0^{\mathrm{T}} \cdot {}^{k}\boldsymbol{\Pi} \cdot \mathbf{g} = {}^{A}\boldsymbol{\varphi}_0^{\mathrm{T}} \cdot ({}^{1}\boldsymbol{\Pi})^{k} \cdot \mathbf{g} = {}^{A}\boldsymbol{\varphi}_k \cdot \mathbf{g} = \sum_{j=1}^{n} {}^{A}p_0(j) \cdot \Delta\mathbf{G}_0(j) \quad (1)$$

$${}^{A}\boldsymbol{\varphi}_0^{\mathrm{T}} \cdot {}^{k}\boldsymbol{\Pi} \cdot \mathbf{u} = {}^{A}\boldsymbol{\varphi}_0^{\mathrm{T}} \cdot ({}^{1}\boldsymbol{\Pi})^{k} \cdot \mathbf{u} = \sum_{j=1}^{n} {}^{A}p_0(j) = 1 \quad (2)$$

where

- ${}^{A}\boldsymbol{\varphi}_0$ is the vector listing the absolute initial probabilities with which an atom in the molecule undergo partition.[41]
- ${}^{A}\boldsymbol{\varphi}_k$ are time dependent vectors, whose components are the absolute probabilities ${}^{A}p_k(j)$ with which an atom in the molecule undergoes partition.[32]
- ${}^{1}\boldsymbol{\Pi}$ is the Markov or stochastic matrix, the ${}^{1}\boldsymbol{\Pi}$ matrix built up as a squared matrix $n \times n$ (n number of atoms in the molecule). The elements of this matrix are probabilities (${}^{1}p_{ij}$) with which the atoms a_j undergo partition themselves at time $t_1 = 1$ given that the atoms a_i have passed throughout the interface at time $t_0 = 0$.[32–43]
- As a result of the Chapman–Kolmogorov equations the ${}^{k}\boldsymbol{\Pi}$ matrices are composed by probabilities (${}^{k}p_{ij}$)

with which atoms a_j undergo partition themselves at any further time $t_k > 1$ given that atoms a_i have passed throughout the interface at time $t_{k-1} < t_k$.

- Finally, vectors **g** and **u** are the vectors of the atomic free energy of partition for a given atom in a specific system and the unitary vector, respectively.[31,32,34,40]

In this context, the present MC model summarized by relationship (1) expresses that the different values in the time $\Delta\mathbf{G}_k$ for the average free partition energy with which a specific molecule undergoes partition on a specific system depends on

(a) The initial probabilities with which each atom tends to undergo partition.
(b) The starting atomic free partition energy $\Delta\mathbf{G}_0(j)$ with which the atoms a_j pass to the other phase of the partition system at time $t_0 = 0$ (elements of **g**).
(c) The subsequent short-range probabilities ${}^1p_{ij}$ (elements of ${}^1\mathbf{\Pi}$) with which each atom tends to undergo partition given that other atoms in the molecule have undergone the same partition process. See below how these terms encode molecular connectivity.

The transition probabilities ${}^1\mathbf{p}_{ij}$ (elements of ${}^1\mathbf{\Pi}$) and the absolute partition probabilities ${}^A p_k(j)$, elements of ${}^A\boldsymbol{\varphi}_k$ can be calculated as follows:

$$ {}^1\mathbf{p}_{ij} = \frac{\alpha_{ij}\cdot\Delta\mathbf{G}_0(j)}{\sum_{m=1}^{\delta+1}\alpha_{im}\cdot\Delta\mathbf{G}_0(m)} = \frac{\alpha_{ij}\cdot(\mathbf{c}_{aj}+\mathbf{c}_{psj}+\mathbf{c}_{bj}+\mathbf{c}_0)}{\sum_{m=1}^{\delta+1}\alpha_{im}\cdot(\mathbf{c}_{am}+\mathbf{c}_{psm}+\mathbf{c}_{bm}+\mathbf{c}_0)} \quad (3) $$

$$ {}^A\mathbf{p}_0(j) = \frac{\alpha_{ij}\cdot\Delta\mathbf{G}_0(j)}{\sum_{m=1}^{n}\alpha_{im}\cdot\Delta\mathbf{G}_0(m)} \quad (4) $$

where δ is the number of atoms covalently bound to the ith atom plus 1 accounting for the atom itself, see also Eq. 1. In consonance with (**c**) α_{ij} are the elements of the atom adjacency matrix, which are equal to 1 if atoms a_i and a_j are bond connected and 0 otherwise. $\Delta\mathbf{G}_0(j)$ is expanded here as a series of linear contributions to the atom free energy of partition including atom specific contribution $\mathbf{c}_{aj}$, partition system contribution $\mathbf{c}_{psj}$, biologic species contribution $\mathbf{c}_{bj}a$ and a scale factor $\mathbf{c}_0$ to avoid negative probability values. All these contributions were determined directly from the experimental data by linear regression techniques using dummy variables for each atom, system, and biological species. These contributions have been depicted in Table 1. Note that, in expression (3), we summed up to δ atoms bound to a_j while in Eq. 4 we have summed up to all the n atoms in the molecule.

On the other hand, given that probabilities must be normalized and then can be considered as absolute ones ${}^A p_k(j)$ equality (2) holds itself. Accordingly, the entropies (Θ_k) with which the model evolves in time can be derived from these probabilities as a classic relationship, k_B being the Boltzmann constant:[32–38]

$$ \Theta_k = -k_B\sum_{j=1}^{n}{}^A p_k(j)\log{}^A p_k(j) \quad (5) $$

Both Eqs. 1 and 2 derived therefore by a vector–matrix–vector approach include as particular cases the quadratic forms. Panoply of these transformations has been previously used in QSAR studies for a long time. For instance, the first molecular descriptor defined in a chemical context, the Wiener index, W (Eq. 6), is a quadratic form.[44] In addition, several other classic Zagreb indices M_1 (7) and M_2 (8), Harary number H (9), Randic invariant χ (10), valence connectivity index χ^v (11), the Balaban index J (12), the MTI index (13), and Marrero-Ponce $q_k(X)$ quadratic indexes (14), just to mention a few examples, may all be expressed with the exception of Eq. 7 as quadratic forms:[45,46]

Table 1. Linear contributions to partition of different atoms, biologic species, and biphasic systems

Atoms	Contribution	Std. error	System	Contribution	Std. error
C	0.061	0.035	S1 (water/air)	1.557	0.939
N	−0.153	0.070	S2 (vegetable oil/water)	−1.187	0.907
O	0.019	0.072	S3 (muscle/plasma)	−1.591	0.526
S	0.254	0.189	S4 (bone/plasma)	−1.599	0.692
H–X	0.085	0.112	S3 (brain/plasma)	−1.867	0.558
H–C	−0.066	0.025	S6 (heart/plasma)	−1.687	0.552
F	−0.206	0.222	S7 (intestine/plasma)	−1.876	0.596
Cl	−0.225	0.131	S8 (skin/plasma)	−1.709	0.591
Br	0.532	1.361	S9 (lung/plasma)	−1.920	0.549
Biological species			S10 (spleen/plasma)	−1.351	0.702
Rat	−0.742	0.950	S11 (liver/air)	−0.056	0.493
Rabbit	−1.296	0.970	S12 (muscle/air)	−0.018	0.459
Mouse	−0.795	0.257	S13 (adipose tissue/air)	0.672	0.459
Human	−0.912	0.050	S14 (blood/air)	0	Zeroed[a]
Abiotic systems	0.000	Zeroed[a]	Intercept	2.873	1.928
		R	df	*F*	*p*
	log *P*	0.62	26.00	11.17	0.00

**The more negative value scaling factor ($\mathbf{c}_0$) (Cl, rabbit, lung/plasma) = −0.5678.
[a] Zeroed predictors after sigma-restricted parameterization.

$$W = \frac{1}{2}(\mathbf{u}\cdot\mathbf{D}\cdot\mathbf{u}^{\mathrm{T}}) \tag{6}$$

$$M_1 = \mathbf{v}\cdot\mathbf{A}\cdot\mathbf{u}^{\mathrm{T}} \tag{7}$$

$$M_2 = \frac{1}{2}(\mathbf{v}\cdot\mathbf{A}\cdot\mathbf{v}^{\mathrm{T}}) \tag{8}$$

$$H = \frac{1}{2}(\mathbf{u}\cdot\mathbf{D}^{-k}\cdot\mathbf{u}^{\mathrm{T}}) \tag{9}$$

$$\chi = \mathbf{v}'\cdot\mathbf{A}\cdot\mathbf{v}'^{\mathrm{T}} \tag{10}$$

$$\chi^{v} = \mathbf{v}''\cdot\mathbf{A}\cdot\mathbf{v}''^{\mathrm{T}} \tag{11}$$

$$J = \frac{1}{2}\cdot\mathbf{C}\cdot(\mathbf{d}'\cdot\mathbf{A}\cdot\mathbf{d}'^{\mathrm{T}}) \tag{12}$$

$$\mathrm{MTI} = \mathbf{v}\cdot(\mathbf{A}+\mathbf{D})\cdot\mathbf{u}^{\mathrm{T}} \tag{13}$$

$$q_k(X) = \mathbf{x}^{\mathrm{T}}\cdot M\cdot\mathbf{x} \tag{14}$$

where **D** represents the topologic distance matrix, **A** the atom adjacency matrix, $\mathbf{D}^{-k}$ is a matrix whose elements are calculated as the inverse of the topologic distance, and **M** is the pseudograph matrix. On the other hand, **u** is a unit vector with all elements equal to 1, **v**, **v**′, and **v**″ are the atoms vertex degree, the atom's vertex degree raised to the power of −1/2, and the atom's valence vertex degree vectors; while **d** is a vector whose elements are the sum of the row distance raised to the power of −1/2 and **x** are the atom's electronegativity vectors. All the vectors and matrices used in expressions (6)–(14) have been exhaustively explained in the literature, reported and references cited therein.[45,46] A seminar work introduced the Γ matrix unifying almost all of them and other quadratic forms reported by these authors.[47] In the present work we propose to call all these forms the deterministic vector–matrix–vector forms by opposition to our stochastic forms. The main advantage of the present stochastic forms is the possibility of deriving average thermodynamic parameters depending on the probability of the states of the MC, which fit more clearly in the physicochemical sense with respect to classic quadratic forms. Specifically, this work introduces for the first time a Markov quadratic form to calculate thermodynamic parameters of the partition process considering in a unified scheme: time, chemical structure, system including biological tissues, and biological species. The method approaches a drug classification scheme taking into consideration all the aforesaid aspects. We studied a large data series of 423 drugs and up to 14 different partition systems collected from different sources[29,30,48–50] at the time of seeking the following linear discriminate analysis classification function:[51–53]

$$\begin{aligned}\mathrm{PL} = {} & 3.03\times\Theta_0(T) + 87.49\times\Theta_2(\mathrm{Het}) - 220.41\\ & \times\Theta_3(\mathrm{Het}) + 133.87\times\Theta_5(\mathrm{Het}) + 5.81\times\Theta_0(\mathrm{Hx})\\ & - 7.98\times\Theta_1(\mathrm{Hx}) + 3.85\times\Delta G_0(\mathrm{Hx})\\ & - 3.55\times\Delta G_1(\mathrm{Hx}) - 1.32\\ N(\mathrm{T}) = 318 \quad & N(\mathrm{PL}) = 106 \quad \%(\mathrm{PL}) = 90.6\\ N(\mathrm{PUL}) = 212 \quad & \%(\mathrm{PUL}) = 92.9\end{aligned} \tag{15}$$

where the terms *T*, Het, and Hx express that we are summing up the entropies or free energies at a given time *k* for all the atoms in the molecule or alternatively only for heteroatoms or labile hydrogens. *N*(T), *N*(PL), *N*(PUL) and its % analogues are, respectively, the total number of cases and accuracy as well as for partition-like (PL) and partition-unlike (PUL) drugs in training. Besides this high overall accuracy of 92.1% (293/318 cases) in training series, the models have shown 90% (36 out of 40 cases) in predicting ones, which are excellent values.[54] The overall accuracy was calculated for compounds in training series used to fit the model. On the other hand, the predictability was determined as the rate of good classified compounds in external series, which were never used to derive the model. In statistics terms, the model showed Fisher ratio values that determined a *p*-level <0.01. This means that accepting the model as valid presupposes an error lower than 1%. The names, system, biological species, and posterior probabilities for all studied compounds have been depicted in the supplementary material. The calculation of all indices was carried out with the MARCH-INSIDE 2.0[55] package while all statistical calculations were developed with the STATISTICA 6.0 software.[56] Regrouping all the terms for a specific parameter without considering time allowed estimation of the overall effect of every aspect in overall partition-related classification of the drug.[32]

$$\begin{aligned}\mathrm{PL} = {} & 3.03\times\Theta(T) + 0.95\times\Theta(\mathrm{Het}) - 2.17\\ & \times\Theta(\mathrm{Hx}) + 0.3\times\Delta G(\mathrm{Hx}) - 1.32\end{aligned} \tag{16}$$

For instance, the higher the total entropy of the process the higher will be by 3.03 times the tendency of the drug to undergo partition from a more organized system to a less organized one. This fact coincides with the expected overall increase in entropy when the system passes from a more organized state to a less one. However, according to the second term of the present model the contribution of 3.03 must be corrected in 0.95 units for heteroatoms, which possibly reflects the presence of stronger electrostatic interactions for these atoms. On the contrary, a negative influence of −2.17 is predicted by the model for hydrogen atoms bound to heteroatoms. This aspect may be connected to the association of these atoms by hydrogen bonds in the more organized state. The last term of the equation predicts the negative change in free energy for these atoms in coincidence with the previous idea. In fact the lack of fit to $R = 0.62$ beyond the error due to collecting data from very different sources is the aspect that determines the necessity in deriving a step-by-step thermodynamic model with respect to the molecular descriptors nature.

In the introduction section we defined the system of reference in such a way that partition is studied from the more organized system to the less organized one. To ensure physical coherence, the method must be invariant[37,41] to the inversion of the system of reference for the partition phenomena. That is to say, an experiment measuring the probability of partition in one direction has to give the same result independently of the system of reference. This fact is straightforward to realize by inverting the signs on the probability definition and noting that both results for the same probability measured from different systems of reference coincide:

$$
\begin{aligned}
{}^{1}p_{ij}(\Rightarrow) &= \frac{\alpha_{ij}\cdot(-\mathbf{c}_{aj}-\mathbf{c}_{psj}-\mathbf{c}_{bj}-\mathbf{c}_{0})}{\sum_{m=1}^{\delta+1}\alpha_{im}\cdot(-\mathbf{c}_{am}-\mathbf{c}_{psm}-\mathbf{c}_{bm}-\mathbf{c}_{0})} \\
&= \frac{-\alpha_{ij}\cdot\Delta\mathbf{G}_{0}(j)}{-\sum_{m=1}^{\delta+1}\alpha_{im}\cdot\Delta\mathbf{G}_{0}(m)} \\
&= \frac{\alpha_{ij}\cdot(\mathbf{c}_{aj}+\mathbf{c}_{psj}+\mathbf{c}_{bj}+\mathbf{c}_{0})}{\sum_{m=1}^{\delta+1}\alpha_{im}\cdot(\mathbf{c}_{am}+\mathbf{c}_{psm}+\mathbf{c}_{bm}+\mathbf{c}_{0})} \\
&= \frac{\alpha_{ij}\cdot-\Delta\mathbf{G}_{0}(j)}{\sum_{m=1}^{\delta+1}\alpha_{im}\cdot-\Delta\mathbf{G}_{0}(m)} = {}^{1}p_{ij}(\Leftarrow) \qquad (17)
\end{aligned}
$$

Finally, we illustrate the use of the model by predicting a high probability (%) for G1 (a novel antibacterial drug) to undergo partition on different biotic systems (rat organs): liver (97.657), spleen (97.51), lung (97.44), and adipose tissue (97.64). These theoretical results coincide with those herein reported by the first time steady-state plasma concentrations (c) and partition coefficients (P) in liver ($c = 42.25 \pm 7.86/P = 4.75$), spleen ($11.47 \pm 4.43/P = 1.29$), lung ($17.04 \pm 3.58/P = 1.91$), and adipose tissue ($28.19 \pm 11.82/P = 3.17$). All values were relative to ^{14}C-labeled-radioactive-G1 in plasma ($c = 8.9 \pm 3.05$) after 3 h of oral administration.[57] In closing, the work leaves a new door open to the development of novel molecular descriptors encoding environment properties but molecular structure too. The present work confirms the increasing utility of mathematical concepts on bioorganic medicinal chemistry research (Table 2).[58,59]

2. Experimental section

2.1. Compound

The [^{14}C]-G1 was obtained from Dalton Chemical Laboratories Inc. (North York, Ontario, Canada) (specific activity, 2.3 mCi/mmol; radiochemical purity, 95% as determined by TLC on silica gel (ethyl acetate and hexane, BDH, UK, 1:10, v/v). The non-radiolabeled compound was supplied by Bioactive Chemical Center (Cuba). To prepare the radiolabeled dose formulation, appropriate amounts of radiolabeled and non-radiolabeled compounds were dissolved in Miglyol® 810N, a caprilic triglyceride (USP, purchased from Hüls America Inc., Germany) to achieve the desired concentration (100 mg/kg) of G1 with a radioactivity level of 8 μCi/mL per rat (9). The mean concentration of [^{14}C]-G1, in the dosing formulation was determined by liquid scintillation counting (Tri-Carb 2000 CA, Canberra Packard, Germany) of aliquots prior to and following administration.

2.2. Animals

Thirty-two male Sprague–Dawley rats, weighing between 280 and 300 g were obtained from an animal care unit (Faculty of Veterinary, University of Saskatchewan, Canada) and were cared for in accordance with the Canadian Council on Animal Care and the University of Saskatchewan guidelines. They were housed in polycarbonate cages and were allowed food (standard laboratory diet) and water ad libitum. The rats were maintained under a 12-h light/12-h dark cycle for at least 1 week prior to treatment at constant room temperature (22 ± 2 °C). All rats were fasted overnight before dosing (16 h) and also 4 h after it. The animals were separated in eight experimental groups (four rats/time point). The rats from the last experimental group (48 h) were housed singly in glass metabolism chambers (Nalgene®, Italy).

2.3. Drug administration and collection of samples

The rats (280–300 g) previously weighed received a single oral dose of [^{14}C]-G1 in Miglyol® 810N (100 mg/kg) by oral-gavage-elongated needle (16 G, 3 in.). The dose volume was 5 mL/kg. At 0.5, 1, 2, 3, 6, 12, 24, and 48 h after drug administration, rats were anesthetized with an intramuscular injection (ketamine/xylazine, SIGMA, USA, 100 mg/kg), and blood

Table 2. Comparison between experimental and theoretically predicted main partition direction for G1, a novel microcidal drug

System	C[a]	PC[b]	PC error	logPC	(Observed/predicted) class	P%[c]
Observed and predicted values for G1						
Liver	42.25	4.75	7.86	0.676	1/1	97.7
Spleen	11.47	1.29	4.43	0.110	1/1	97.5
Lung	17.04	1.91	3.58	0.282	1/1	97.4
Adipose tissue	28.19	3.17	11.82	0.501	1/1	97.6
Plasma	8.9		3.05			

[a] C: plasma concentration mg mol^{-1}.
[b] PC: partition coefficient.
[c] Posterior probability predicted (%).

collected in heparinized tubes (Becton & Dickinson Vacutainer Systems, UK) by cardiac puncture. Samples of tissues were rapidly removed, rinsed, dried, weighed, and frozen. Red blood cells and plasma were separated by centrifugation (3000 rpm, 15 min, 4 °C, Hettich Rotanta, Germany). All tissues and fluids were kept frozen at −20 °C until further processing. Radioactivity was also determined in whole blood.

2.4. Analysis of samples

Three fractions of each tissue (~0.1 g) were weighed and homogenized. An aliquot of 1 mL of Toluene 350 (Packard Instrument Company, Meriden, CT) was added to samples and heated at 50 °C overnight. Samples were decolorized with 180 μL of 30% hydrogen peroxide (BDH, UK) and 30 μL of glacial acetic acid (ACS, BDH, UK) were added and left 48 h in a dark place. Finally 2.4 mL of scintillation cocktail (Ready Value™, Beckman, USA) was added, mixed in vortex and counted with a liquid scintillation counter (TRI-CARB 2000 CA, Canberra Packard, Germany). Blood samples (100 μL) were mixed with 1 mL of water in order to hemolyze the red cells. The total radioactivity was determined as described above for tissue samples. Samples of urine (1 mL) and plasma (100 μL) were counted as gel, adding 2 mL of scintillation cocktail. The radioactivity in these fluids was determined in a liquid scintillation counter. The radioactivity was expressed as disintegration per minute (dpm). The aforementioned experimental procedures were carried out according to established laboratory guidelines (University of Saskatchewan).[60,61]

Acknowledgements

H. Gonzalez-Diaz would like to express his most sincere gratitude to the editor Prof. L. Ghosez and his assistant H. Vanliverde for professional and kind attention as well as to the unknown referees for useful comments, which improved the final presentation of the manuscript.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2004.11.059.

References and notes

1. Katritzky, A. R.; Kuanar, M.; Fara, D. C.; Karelson, M.; Acree, W. E., Jr. *Bioorg. Med. Chem.* **2004**, *12*, 4735.
2. Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101.
3. Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
4. Cramer, C. J.; Truhlar, D. G. *Structure and Reactivity in Aqueous Solution*; American Chemical Society: Washington, DC, 1994.
5. Bonaccorsi, R.; Floris, F.; Tomasi, J. A. *J. Mol. Liq.* **1990**, *47*, 25.
6. Bonaccorsi, R.; Floris, F.; Palla, P.; Tomasi, J. *Thermochim. Acta* **1990**, *162*, 213–222.
7. Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057.
8. Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1995**, *99*, 7137.
9. Sitkoff, D.; Ben-Tal, N.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 2744.
10. Luque, F. J.; Zhang, Y.; Aleman, C.; Bachs, M.; Gao, J.; Orozco, M. *J. Phys. Chem.* **1996**, *100*, 4269.
11. Luque, F. J.; Aleman, C.; Bachs, M.; Orozco, M. *J. Comput. Chem.* **1996**, *17*, 806.
12. Leo, A. J. *Chem. Rev.* **1993**, *93*, 1281.
13. Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. *J. Chem. Soc., Perkin Trans. 2* **1990**, 291.
14. Zhang, Y.; Dallas, A. J.; Carr, P. W. *J. Chromatogr.* **1993**, *638*, 43.
15. Leo, A. J. *Masterfile MedChem. Software*; BioByte Corp: Claremont, CA, 1994.
16. Albertsson, P.-A. *Partition of Cell Particles and Macromolecules*; Wiley & Sons: New York, 1986.
17. Rogers, R. D.; Eiteman, M. A. *Aqueous Biphasic Separations: Biomolecules to Metal Ions*; Plenum: New York, 1995.
18. Zaslavsky, B. Y. *Aqueous Two-Phase Partitioning: Physical Chemistry and Bioanalytical Applications*; Marcel Dekker: New York, 1994.
19. Rogers, R. D.; Bond, A. H.; Bauer, C. B. *Sep. Sci. Technol.* **1993**, *28*, 139.
20. Rogers, R. D.; Bond, A. H.; Baucer, C. B.; Zhang, J.; Griffin, S. T. *J. Chromatogr. B* **1996**, *680*, 221.
21. Rogers, R. D.; Bauer, C. B. *J. Chromatogr. B* **1996**, *680*, 237.
22. Rogers, R. D.; Griffin, S. T. *J. Chromatogr. B* **1998**, *711*, 277.
23. Rogers, R. D.; Willauer, H. D.; Griffin, S. T.; Huddleston, J. G. *J. Chromatogr. B* **1998**, *711*, 255.
24. Willauer, H. D.; Huddleston, J. G.; Griffin, S. T.; Rogers, R. D. *Sep. Sci. Technol.* **1999**, *34*, 1069.
25. Li, M.; Willauer, H. D.; Huddleston, J. G.; Rogers, R. D. *Sep. Sci. Technol.* **2001**, *36*, 835.
26. Willauer, H. D.; Huddleston, J. G.; Li, M.; Rogers, R. D. *J. Chromatogr. B* **2000**, *743*, 127.
27. Zaslavsky, B. Y.; Mestechkina, N. M.; Miheeva, L. M.; Rogozhin, S. V.; Bakalkin, G. Ya.; Rjazhsky, G. G.; Chetverina, E. V.; Asmuko, A. A.; Bespalova, J. D.; Korobov, N. V.; Chichenkov, O. N. *Biochem. Pharmacol.* **1982**, *31*, 3757.
28. Gawlik, B. M.; Sotiriou, N.; Feicht, E. A.; Schulte-Hostede, S.; Kettrup, A. *Chemosphere* **1997**, *34*, 2525.
29. Poulin, P.; Krishnan, K. A. *Hum. Exp. Toxicol.* **1995**, *14*, 273.
30. Abraham, M. H.; Kamlet, M. J.; Taft, R. W.; Doherty, R. M.; Weathersby, P. K. *J. Med. Chem.* **1985**, *28*, 865.
31. Freund, J. A.; Poschel, T. Stochastic Processes in Physics, Chemistry, and Biology. In *Lect. Notes Phys.*; Springer: Berlin, Germany, 2000.
32. González-Díaz, H.; Marrero, Y.; Hernández, I.; Bastida, I.; Tenorio, I.; Nasco, O.; Uriarte, E.; Castañedo, N. C.; Cabrera-Pérez, M. A.; Aguila, E.; Marrero, O.; Morales, A.; González, M. P. *Chem. Res. Tox.* **2003**, *16*, 1318.
33. González-Díaz, H.; Ramos de, A. R.; Molina, R. R. *Bioinformatics* **2003**, *19*, 2079.
34. González-Díaz, H.; Molina, R. R.; Uriarte, E. *Polymer* **2004**, *45*, 3845.
35. González-Díaz, H.; Molina, R. R.; Uriarte, E. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4691–4695.

36. Ramos de, A. R.; González-Díaz, H.; Molina, R. R.; Uriarte, E. *Proteins, Struct. Func. Bioinf.* **2004**, *56*, 715.
37. González-Díaz, H.; Bastida, I.; Castañedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H. S.; Ramos de, A. R. *Bull. Math. Biol.* **2004**, *66*, 1285.
38. Ramos de, A. R.; González Díaz, H.; Molina, R.; González, M. P.; Uriarte, E. *Bioorg. Med. Chem.* **2004**, *12*, 4815.
39. González-Díaz, H.; Olazábal, E.; Castañedo, N.; Hernández, S. I.; Morales, A.; Serrano, H. S.; González, J.; Ramos de, A. R. *J. Mol. Mod.* **2002**, *8*, 237.
40. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernádez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J. A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L. A.; Cabrera, M. A. *J. Mol. Mod.* **2003**, *9*, 395.
41. González-Díaz, H.; Hernández, S. I.; Uriarte, E.; Santana, L. *Comp. Biol. Chem.* **2003**, *27*, 217.
42. González-Díaz, H.; Ramos de, A. R.; Molina, R. R. *Bull. Math. Biol.* **2003**, *65*, 991.
43. González-Díaz, H.; Ramos de, A. R.; Uriarte, E. *Online J. Bioinf.* **2002**, *1*, 83.
44. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
45. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley VCH: Weinheim, Germany, 2000.
46. Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2004**, *12*, 5331.
47. Estrada, E. *Chem. Phys. Lett.* **2001**, *336*, 248.
48. Kamlet, M. J.; Abraham, D. J.; Doherty, R. M.; Taft, R. W.; Abraham, M. H. *J. Pharm. Sci.* **1986**, *75*, 350.
49. Paterson, S.; Mackay, D. *Br. J. Ind. Med.* **1989**, *46*, 321.
50. Abraham, M. H.; Weathersby, P. K. *J. Pharm. Sci.* **1994**, *83*, 1450.
51. Cronin, M. T. D.; Aynur, A. O.; Dearden, C. J.; Deffy, C. J.; Netzeva, I. T.; Patel, H.; Rowe, H. P.; Schultz, T. W.; Worth, P. A.; Voutzolidis, K.; Schüürmann, G. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 869.
52. González, M. P.; González-Díaz, H.; Molina, R.; Cabrera-Pérez, M. A.; Ramos de, A. R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192.
53. Cabrera-Pérez, M. A.; García, A. R.; Teruel, C. F.; Álvarez, I. G.; Sanz, M. B. *Eur. J. Pharm. Biopharm.* **2003**, *56*, 197.
54. Van Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Timmerman, K-L. H., Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H., Ed.; VCH: Weinhiem, 1995; Vol. 2, pp 265–282.
55. González-Díaz, H.; Molina, R.; Hernández, I. *MARCH-INSIDE*, 2002, version 2.0. This is a preliminary experimental version. A future professional version shall be available to the public. For any information about it send an e-mail to the corresponding author humbertogd@vodafone.es or humbertogd@usc.es.
56. Statsoft inc. *Statistica*, **2002**, *6.0*.
57. Blondeau, J. M.; Castañedo, N.; Gonzalez, O.; Medina, R.; Silveira, E. *Antimicrob. Agents Chemother.* **1999**, *11*, 1663.
58. González, M. P.; Moldes, M. C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3077.
59. González, M. P.; Moldes, M. C. T. *Bioorg. Med. Chem.* **2004**, *12*, 2985.
60. Francis, L. S.; James, M. J. *Preclinical Drug Disposition: A Laboratory Handbook*; Marcel Dekker Inc: New York, **1991**; pp 17–49.
61. Mast, R. W.; Jeffcoat, A. R.; Sadler, B. M.; Kraska, R. C.; Friedman, M. A. *Fd. Chem. Toxic.* **1983**, *21*, 807.