

Information Content in Organic Molecules: Aggregation States and Solvent Effects

Daniel J. Graham*

Department of Chemistry, Loyola University of Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

Received April 1, 2005

The information content of organic molecules has been the subject of a series of papers from this lab. The investigation continues as aggregation states and some of their attendant solvent effects are examined. An organic molecule operates as an information source not in a vacuum but rather in conjunction with one or more solvent compounds. Accordingly, solvents (ethanol, acetone, etc.) furnish both a source and a channel that modify molecular information. In this paper, Brownian techniques are developed further so as to quantify molecular information with aggregation states taken into account. Several applications follow concerning organic acid ionization in solution, tautomerization reactions, and molecular activity at biological receptor sites. The goal is to advance Brownian processing as a means of probing molecular information and its communication. To this end, practical examples are offered relating structure and function along informatic lines.

I. INTRODUCTION

The second paragraph of a 1948 paper by Claude Shannon declares the following:¹

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

Shannon proceeded to develop a mathematical model for communication, following in the footsteps of Hartley, Nyquist, and others.^{2–4} His revolutionary theorem proved that it is possible to construct a code whereby a transmitted message sequence can be predicted via the received sequence with an error that does not exceed an arbitrarily small number. Remarkably, this possibility holds true however noisy the operating conditions. It has been the task of researchers over the years to design and implement effective codes along with vehicles for transmission and reception. The practical realizations of Shannon's contributions are diverse with immeasurable impact. Further, they exhibit features in common: all incorporate a source of information that is coupled to encoding and decoding devices. These are mediated by a physical channel that is injected, to some degree, with noise. A block diagram appears in the upper half of Figure 1 so as to illustrate the essential elements of physical communication.^{1,5}

Organic molecules embody a communication system of their own. Shannon's words thus inspire a paraphrasing:

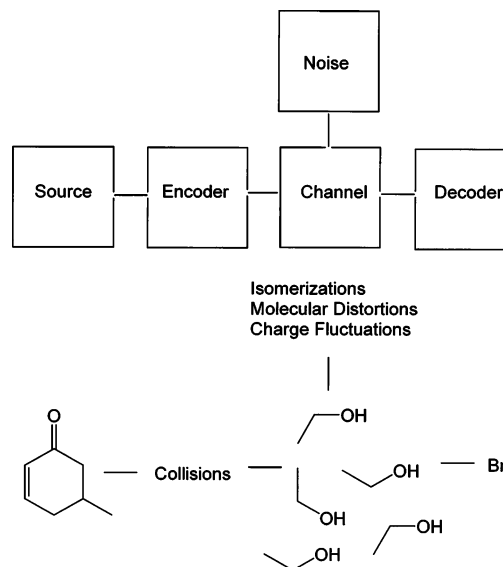


Figure 1. Elements of Communication. Upper portion illustrates block assembly of the essential elements in physical communication. Lower portion illustrates the analogous entities involving three types of molecules in solution.

The fundamental chemical problem is that of reproducing at one physical site either exactly or approximately a molecular message composed at another site. Frequently, the messages have *meaning*; that is, they refer to or are correlated according to modern chemical structure theory. These semantic aspects of chemical communication are irrelevant, however, to many situations. The critical aspect is that a significant message conveyed by atoms and chemical bonds is *selected from a set* of possible messages. For a molecular system to be useful, it must operate for the possible message selections, not just the ones with transforming activity, since these are largely unknown at the time of design.

The paraphrase and its variations provoke thought. The “one physical site” could refer to a compound as routine as

* Corresponding author e-mail: dgraha1@luc.edu; fax: 1-773-508-3086; phone: 1-773-508-3169.

acetone or as complicated as a globular protein. The term “another site” brings to mind laboratory solutions, cellular media, and combinatorial libraries, depending on the application. The “message” communicated by a molecule is composed via atom–bond–atom states, the precise interaction details of which are unknown in the vast majority of applications.⁶ The robust chemical designs are, rather, those that support a variety of messages including the one(s) with transforming capacity.

Organic compounds are the principal vehicles for chemical message transmission and processing. Accordingly, the issues surrounding functional group activity, skeletal integrity, and molecular discrimination are unequivocally relevant to communication and information theory. Over the years, this statistical approach has been well-ingrained in the chemistry realm through the work of many researchers. To list only a few, Bonchev and Trinajstić began applying information theory in the 1970s to diverse sets of topological indices (TIs) composed via the line/vertex graphs for organic molecules.⁷ Later, Bertz enhanced the connection between graph and information theory so as to formulate a general index of molecular complexity.⁸ Also along graph theoretic lines, Mekenyan and Basak employed information theory in order to correlate TIs and chemical reactivity.⁹ More recently, in this journal, Bajorath and co-workers have explored numerous Shannon entropy (SE) measures affiliated with large chemical databases and other sets of molecules.¹⁰ In many cases, the SE measures proved acutely sensitive to specific interaction and molecular class properties, most notably concerning natural products and pharmaceuticals. Importantly, information measures based on TIs for individual compounds and the descriptor statistics for databases have predicated new generations of models for quantitative prediction. The reader is referred to two review articles and one text providing overviews of the graph theoretic and SE approaches to organic molecules.^{11–13}

Our aim the past few years has likewise been to advance the informatic perspective in organic chemistry—to understand how carbon-based systems, even ones lacking biological relevance, express and respond to information. This research was initiated by an examination of the base- and regioinformation content of organic molecules.^{14,15} Most recently, a Brownian processing model was developed to quantify the information encoded in covalent bond networks, taking both structural and molecular communication details into account.^{16,17} This approach departs from ones contributed by previous researchers because the energetics, collision, and random walk peculiarities of organic molecules are incorporated into the discussion. In particular, the Brownian methods such as those used for this paper are patterned after the real-life information processing of polymerase enzymes and RNA molecules.¹⁸ For our systems of interest, as with biopolymers, chemical messages are communicated by lengthy strings of discrete, fundamental code units. The overriding challenge is to understand how the message space connects with actual chemical behavior.

In its most general sense, information is any physical entity capable of altering a probability distribution.¹⁹ Accordingly, a molecule can act as both a source and a receiver. Vacuum environments, however, are of scant utility in real-life circumstances. Most typically, a solvent such as acetone, water, and so forth is necessary during chemical message

transmission and reception. In communication terms, solvents operate, in part, as channels that support and modify information. A schematic for these ideas appears in the lower half of Figure 1. When 3-one-5-methyl-cyclohexene—or any other molecule—serves as an information source, the processing entails compounds such as ethanol.²⁰ The electronic messages derive from the functional groups and skeletal frameworks of all parties and are communicated during collisions. All the while, a variety of charge fluctuations intersperse noise. If a reagent such as bromine is introduced, this compound can discriminate the solvent and solute compounds. Its role is analogous to that of a decoding device in a communication network.

Our previous efforts have focused on molecules as information sources under isolated conditions.^{14–17} The investigation continues as an account is taken of components that are integrated with the sources and are every bit as vital to the communication and reception enterprise. Brownian techniques are, thus, developed further so as to quantify information, allowing for aggregation states and attendant solvent effects.

This paper is organized as follows. In Section II, the process of molecular information and communication is examined from both analogue and digital points of view. We show how the random walk (Brownian) techniques of previous work can be adapted to aggregation states whence sample data are presented in Section III. In Section IV, digital methods are developed further so as to enable informatic comparisons of different systems. A series of applications is presented in Section V concerning acid ionization, tautomerization chemistry, and the action of select molecules at biological receptor sites. The overriding goal is to advance Brownian processing as a means of probing molecular information and its communication. Accordingly, the Section V examples are practical ones relating structure and function along informatic lines. As discussed in Section VI, it is hoped that the tools of this paper can assist in a variety of process chemistry and design situations.

II. MOLECULES, AGGREGATION STATES, AND ELECTRONIC MESSAGES

A. An Analogue Perspective. How does a molecule communicate and what is its message? From an analogue perspective, the answers lie in an electric force field $E(r,t)$, which derives from a charge density function $\rho(r,t)$ dictated by the Schrödinger equation; r and t denote position and time coordinates, respectively.²¹

The upper panel of Figure 2 depicts 3-one-5-methyl-cyclohexene in stationary, isolated form. The structure graph mimics only a crude, time-averaged version of ρ . As is well-known, however, ρ can be obtained at a high order of approximation using an arsenal of computational tools.²² In Figure 2, a point charge q has been depicted at an arbitrary position about 7 Å (scale-wise) from the oxygen center. If the molecule rotates about the carbonyl bond, the Coulombic field at q varies approximately as shown in the upper right-hand side plot. Note that the forces are calculated in the limit $q \rightarrow 0$ and are expressed in complete form via three vector components. For simplicity, however, only the magnitude of E has been represented over a single 2π -rotation cycle. If

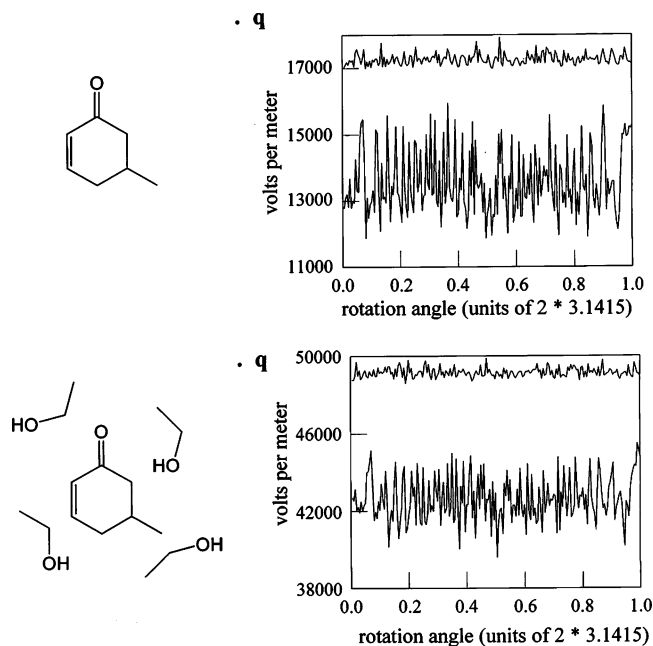


Figure 2. Molecules and analogue force fields. The upper left-hand side features 3-one-5-methyl-cyclohexene and a point charge q . If the molecule rotates about the carbonyl axis, the force at q varies as shown in the upper right-hand-side plot. The offset graph follows from q placed at 28 Å from the source. Only the magnitudes have been represented over one 2π rotation cycle. The lower portion depicts 3-one-5-methyl-cyclohexene plus four ethanol units. Two force field examples are illustrated in the corresponding right-hand-side plot. The point charge locations are the same as those used in the isolated molecule calculations.

the point charge is located further away, a field of lesser intensity is recorded. One example of this is included here, with the plot offset for clarity. For the data shown, q was fixed ca. 28 Å from the oxygen center, along the same direction as in the 7 Å case.

Matters are different when a molecule is accompanied by solvent compounds such as ethanol. One scenario is represented in the lower part of Figure 2, whereby four $\text{CH}_3\text{CH}_2\text{OH}$ units have been placed at arbitrary positions about 3-one-5-methyl-cyclohexene. In this case, the force at q derives from several ρ 's; each contribution is uniquely time- and space-dependent given molecular translations, rotations, and collisions. The results of two calculations are illustrated in the lower plot of Figure 2. The point charge positions are the same as those used for the single-molecule examples.

Different forces are observed if the q sites and molecular rotation axes are altered. The fields are, moreover, sensitive to the number, identity, and motion (translations and rotations) of the solvent compounds. More extensive computations take molecular vibrations, excited states, and induced fields into account. In all cases, the fields demonstrate a marked dependence on the approximations applied to ρ and the placement of q .

Figure 2 makes several points. The first is that the messages transmitted by even relatively simple molecules are complex—the carrier fields are acutely variable along space, time, and environmental dimensions. This underscores a crux issue in molecular communication because the spatial and temporal complexity is all but impossible to intuit via the source structure.

Second, there appear two facets of a molecular message. One is the set of frequencies and *relative* amplitudes of the force fields; the members of this set depend on the source charges and their electronic environment. The frequencies and amplitudes constitute a spectrum, one which furnishes every molecule with an electronic signature. As a field is spatially dependent, even pedestrian compounds express multiple signatures. It is a collection of signatures that distinguishes 3-one-5-methyl-cyclohexene in a vacuum from the same charge assembly mixed with ethanol. Field signatures, moreover, enable discrimination of 3-one-5-methyl-cyclohexene from a related compound such as 3-one-4-methyl-cyclohexene.

A second facet enters by way of the field magnitudes. If the source fields at distances 7 and 28 Å are compared, the frequencies and relative amplitudes prove similar. The magnitudes are significantly disparate, however, owing to the spatial scaling of the Coulombic terms. Clearly, the messages communicated by a molecule depend not only on the source charges and environment but also on the observer's location.

Figure 2 makes the additional point that the messages posed by a molecule are enormously variable. On one hand, this indicates the depth and nuance surrounding noncovalent interactions. On another, the complexity degree is in sharp contrast with the nature of molecular representations. Structure graphs and other simple icons emphasize bond order, anisotropy, and chemical valence. As blueprints, the graphs such as those in Figures 1 and 2 enable chemists to synthesize the real angstrom-scale entities and to communicate the properties to other chemists. It is difficult to envision chemical practice without elementary icons for ρ . Comprehending the molecular messages in analogue terms, however, is very complicated.

B. A Digital Perspective. A digital approach views an organic compound as a statistical data file that is approximated by a structure graph and communicated by a random walk process.¹⁶ Thus, the pictorial for 3-one-5-methyl-cyclohexene (or any other molecule) serves as an information source; the communication can be represented via an alphabet of atom–bond–atom code units. As shown in our previous work, graphs for 3-one-5-methyl-cyclohexene, ethanol, and so forth offer an *infinite* number of messages that can be expressed via the code unit format.¹⁶ To chart the message space, one applies random walk or Brownian methods to the source file. These query the different regions of the graph with the results tabulated as lengthy strings of code units. As yet another computational tool, Brownian processing simulates the collisions between an organic molecule and an error-free observer outfitted with a tape recorder. The end product is a digital rendering of the possible messages.

The core ideas of Brownian processing are illustrated in Figure 3. In the serial processing (upper portion) of 3-one-5-methyl-cyclohexene, each atom–bond–atom unit is a randomly selected nearest neighbor of the previous one—the most probable event following collision and recognition at one molecular site entails a nearest-neighbor site. In parallel processing (lower portion), a random collision involves a cluster of nearest-neighbor units of the structure graph. Brownian methods tabulate the possible clusters of two, three, and so forth code units.

Equation III-1 addresses the *alphabetic message* contained via a Brownian tape. Higher order I manifest larger-size messages. Thus, $I_{N+1} \geq I_N$ irrespective of the processing mode: a hexamer of code units offers greater instruction capacity than a pentamer. Importantly, I quantities can be related to equivalent fuel terms: 1.00 bit corresponds to a fuel measure of $k_B T \log_e 2$. As in our previous work, it is useful to address I values in units of $k_B T$, whereby the reduced (i.e., dimensionless) variables are denoted using bold italics, for example, $I_6^{(S)}$.¹⁶ Importantly, the reduced terms for alphabetic code delineate one facet of a chemical message.

The second follows from a subtle but critical feature of Brownian processing. When registering different sites of a molecule or aggregate, a working processor must distinguish (C–H) from (C–O), (C=O) from (C–O), (C=C) from (C–C), and so forth. A faithful message recording obtains only by a processor's ability to sense precisely the electric potential of an atom–bond–atom state. Recognition of a state entails a comparison that reduces uncertainty along *both* code unit and energy dimensions. Viewed another way, the sites of a molecule or aggregate communicate effectively with a Brownian processor. In so doing, they affect how work is accomplished by the processor, the extent to which depends on the potential differences between the recognition sites.

In Brownian processing, the alphabetic uncertainty is addressed via the weighted surprisals and summation in eq III-1. By contrast, it proves most straightforward to handle the processor energy uncertainty via covalent bond energies compiled in standard tables.²³ During Brownian processing, the degree of energy uncertainty depends on the dispersion about the average, of which the standard deviation σ serves as an unbiased estimate. As with the surprisal quantities, σ estimates can be established to several orders of tape analysis. For our purpose, the dispersion quantities will be denoted $\sigma_{En}^{(S)}$ and $\sigma_{En}^{(P)}$. When reported in units of $k_B T$, the reduced-variable terms will be distinguished using bold italics, for example, $\sigma_{En}^{(S)}$ and $\sigma_{En}^{(P)}$. Importantly, the energy dispersions for a Brownian tape render the second facet of a chemical message.

As an example, Figure 5 illustrates Brownian data for 3-one-5-methyl-cyclohexene/ethanol. The upper panel shows $I_n^{(S)}$ (filled squares) and $I_n^{(P)}$ (open squares) as a function of analysis orders 1–8. Included are the results for 3-one-5-methyl-cyclohexene and ethanol molecules under isolated conditions. One finds the serial and parallel quantities to be closely aligned, differing by only a few percent. By contrast, combining ethanol with 3-one-5-methyl-cyclohexene enhances the alphabetic information by 15–18% at all orders. The lower panel features $\sigma_{En}^{(S)}$ and $\sigma_{En}^{(P)}$ data. Again, one finds the serial and parallel results to differ by only a few percent. By contrast, combining ethanol with 3-one-5-methyl-cyclohexene diminishes the molecule's energy dispersion by more than 20%.

Brownian methods can be applied to any combination of molecules via their structure graphs. The results chart the chemical message space along alphabetic and energy dimensions. Data such as in Figure 5 demonstrate that combining one molecule with another tunes the space in a nontrivial way. These effects are reminiscent of solvent-induced perturbations of electronic energy levels.²⁴ The degree of

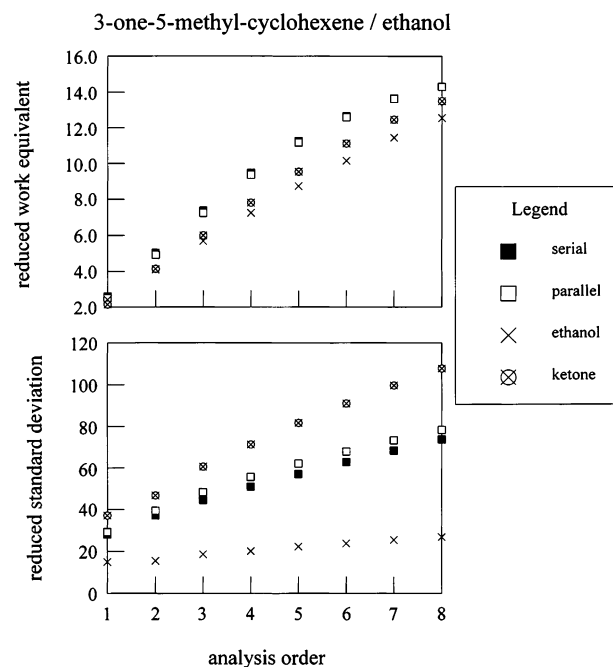


Figure 5. Sample data for 3-one-5-methyl-cyclohexene/ethanol. Upper panel and lower panels show, respectively, the reduced work and energy dispersion as a function of analysis order. Data marked by crosses and crossed-circles pertain, respectively, to ethanol and 3-one-5-methyl-cyclohexene under isolated conditions. The error bars associated with different Brownian runs are smaller than the symbol widths.

perturbation depends on many things that include the structure and diversity of the interacting compounds.

Brownian processing offers two different modes, serial and parallel, while the tape analysis extends to multiple orders—any system can be mined to yield a ponderous amount of data! For conciseness, then, the results in the remaining sections will concern serial processing only; parallel data as in Figure 5 prove very similar. Moreover, the informatics will be presented using a compressed format. This follows from computing the *average values* of $I_n^{(S)}$ and $\sigma_{En}^{(S)}$ over multiple orders of the aggregate message space. Via the averaging, a point coordinate ($\langle I_n^{(S)} \rangle$, $\langle \sigma_{En}^{(S)} \rangle$) so obtained defines a dimensionless length L in an information-dispersion message plane. We will refer to L as the *dispersion length* for an aggregate system of interest; L complements the length quantities defined in the energy-information plane in ref 16.

Figure 6 illustrates the data compression workings. The upper panel shows $\sigma_{En}^{(S)}$ versus $I_n^{(S)}$ for 3-one-5-methyl-cyclohexene combined with five different solvent compounds: cyclohexane, ethyl ether, acetonitrile, dioxane, and ethanol. There is one data point for each analysis order as in Figure 5. The lower panel shows the corresponding L quantities that follow from averaging $I_n^{(S)}$ and $\sigma_{En}^{(S)}$ over eight orders. For the sample data, one finds 3-one-5-methyl-cyclohexene/acetonitrile to demonstrate the greatest L . The five systems, moreover, offer two distinct L groupings. In particular, the dispersion length data for 3-one-5-methyl-cyclohexene/cyclohexane, /ethyl ether, /dioxane, and /ethanol are closely aligned, whereas L for 3-one-5-methyl-cyclohexene/acetonitrile is markedly greater. We note that it is exceedingly difficult to make intuitive sense of the analogue messages for these systems (cf. Figure 2). By contrast,

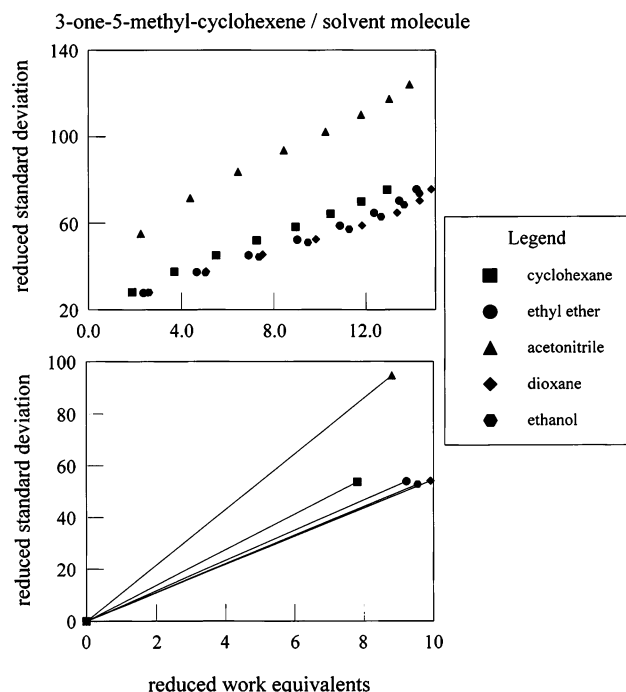


Figure 6. Data for 3-one-5-methyl-cyclohexene combined with different solvent compounds. The upper panel shows reduced energy dispersion versus reduced work equivalents for analysis orders 1–8. By averaging the quantities, dispersion lengths L are obtained as shown in the lower panel.

Brownian L values are readily discriminated and indeed assist in relating angstrom-scale structure with chemical action.

IV. DISPERSION LENGTH, MOLECULAR ACTIVITY, AND COMMUNICATION

A system's chemistry depends on the details of structure and composition. These can be organized along multiple dimensions (HOMO–LUMO energy, lipophilicity, etc.) so as to correlate structure and function.²⁵ In the approach elected here, chemical function can be investigated via the L quantities obtained by Brownian processing. A correlation framework along molecular information and communication lines is thus contained in this section.

The central premise is that chemical activity is contingent upon the effective transmission and reception of digital information. As in Figure 1, systems for accomplishing both tasks entail a source, encoder, channel, and decoder—the role of noise is addressed elsewhere.²⁶ Systems devoid of information pose no activity because they have nothing to communicate. By contrast, a source's activity is enhanced if its impact at a decoding (i.e., reagent or chemical receptor) site is somehow enhanced. Under special circumstances, different sources can register equal impact; in an analogous way, disparate charge distributions can exert identical forces on a point charge, given suitable repositioning and medium dielectric tuning. In the framework, molecular activity is viewed as a natural byproduct of information transfer in a solution environment. As things turn out, the criterion for equal impact forms a basis for quantitative structure–function correlations.

To initiate the framework, one defines a series of standard or reference states through which all activity comparisons are made. A standard state arguably could be any combina-

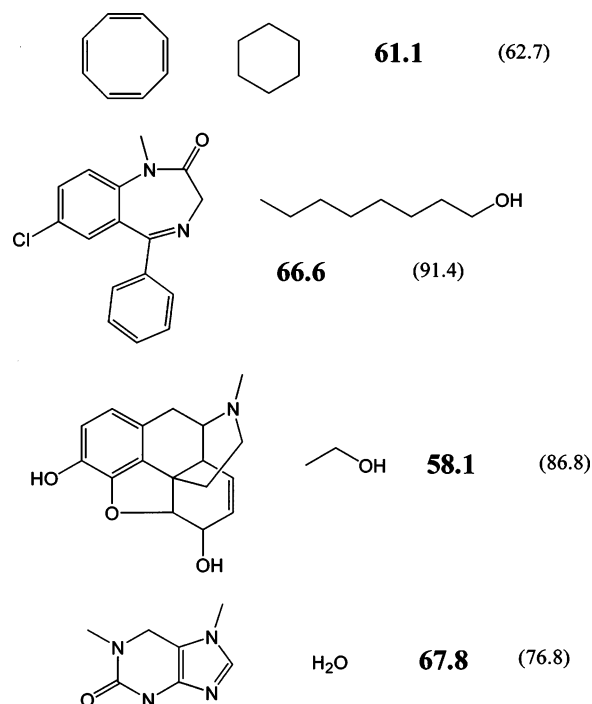


Figure 7. Binary systems and dispersion lengths. The structure graphs for cyclooctatetraene/cyclohexane, diazepam/octanol, morphine/ethanol, and caffeine/water are represented. L values obtained by Brownian processing are shown in bold-face type. The dispersion lengths for the left-hand-side compounds under isolated conditions are listed in parentheses.

tion of solvent and solute compounds: 1:1, 2:1, ..., 5:3, and so forth. However, molecule–solvent interactions via two-body collisions are anticipated to weight far more importantly than higher-order (three-molecule, four-molecule, etc.) collisions. Further, it is most straightforward to encode states via a single solvent and solute compound, as represented by their structure graphs. Thus, Figures 5 and 6 present data for binary systems; in the correlation framework, these will pertain to the standard states of 3-one-5-methyl-cyclohexene combined with different solvent compounds. The dispersion lengths for standard states will be distinguished via 0 superscripts; thus, the lower panel of Figure 6 offers L^0 data for five different systems. Carbon chemistry poses an infinite number of molecular pair combinations; the same statement holds true for informatic standard states.

Values of L^0 are case-specific and, thus, difficult to gauge in advance. To emphasize this, Figure 7 depicts the standard states of four diverse systems: cyclooctatetraene/cyclohexane, diazepam/octanol, morphine/ethanol, and caffeine/water. For each system, L^0 is reported in bold print while the parenthetic values refer to the dispersion lengths for the left-hand-side compounds under isolated conditions. Note how length-based rankings differ for the two sets of conditions. The caffeine molecule, for example, evinces a third-place ranking when in isolated form. L^0 for caffeine/water combinations proves highest in Figure 7, owing to code unit and energy diversity.

L^0 can be viewed as the informatic analogue of a standard chemical potential μ^0 . The latter quantifies the free energy density and diffusive power for a system under defined accessible circumstances.²⁷ Accordingly, L^0 also reflects two interrelated properties: an aggregate's capacity for controlling work performance and for communicating with a

reagent/receptor site under special conditions. “Special conditions” refers to a single aggregate in the absence of all other information sources. In the framework, the standard state of caffeine/water ($L^\circ = 67.8$, Figure 7) offers greater potential for work control than does cyclooctatetraene/cyclohexane ($L^\circ = 61.1$). Further, a standard state of diazepam/octanol ($L^\circ = 66.6$) exerts greater impact at a reagent/receptor site than does morphine/ethanol ($L^\circ = 58.1$). Interestingly, the reagent/receptor site details do not have to be specified when correlating structure and function along informatic and communication lines.

Nature fortunately expresses states that deviate from standard. For nonstandard states, a system’s chemical potential scales as follows:

$$\mu = \mu^\circ + RT \log_e \gamma c \quad (\text{IV-1})$$

where R and T have their usual meaning and γ and c denote activity and concentration coefficients, respectively.²⁸ In our approach, the dispersion lengths for nonstandard states will be taken to scale in a parallel way:

$$L = L^\circ + \tau \log_e \kappa \Theta \quad (\text{IV-2})$$

where τ is a temperature parameter. κ will be interpreted by analogy with c of eq IV-1: κ pertains to the information available from and transmitted by an aggregate system of interest. As with c , κ equals unity for a standard state yet is modified when additional sources must be taken into account. For example, $\kappa \neq 1$ for caffeine/water if acetone/ethanol is added to the communication network.

The log argument in eq IV-2 includes the parameter Θ . For a standard state, a reagent/receptor site can register information in an unperturbed fashion—the site is in contact with only one source at a time. Matters are different, however, if new sources are put into play. These affect not only the transmission attributes (weighed by κ) of existing sources but also the reagent/receptor properties. For example, the addition of cyclooctatetraene/hexane would alter the interactions between a reagent/receptor and an information source already in place such as caffeine/water. As with parameter κ , Θ equals unity for a standard state and is modified otherwise in a case-specific way. In the correlation framework, Θ will be referred to as a receptor bias or distribution parameter. The information for states with $\Theta > 1$ is registered more effectively by a reagent/receptor, compared with standard states, and states with $\Theta < 1$. By contrast, sources with identical Θ values affect a receiver of information in the same way. One says they belong to the same informatic distribution—a type of class distinction—of which Θ provides a convenient label accessible through Brownian processing.

Figure 8 offers a schematic for the correlation framework. The upper portion illustrates two hypothetical aggregate sources of information communicating with a reagent/receptor under standard state conditions. The lower half represents a nonstandard state, along with the tuning effects of κ and Θ . For this hypothetical case, Aggregate 2 affects the information transmitted by Aggregate 1—and vice versa. Thus, $\kappa \neq 1$ for each party, their transmissions represented by arrows of different intensity. Further, Aggregate 2 affects how information is registered by a reagent/receptor in contact

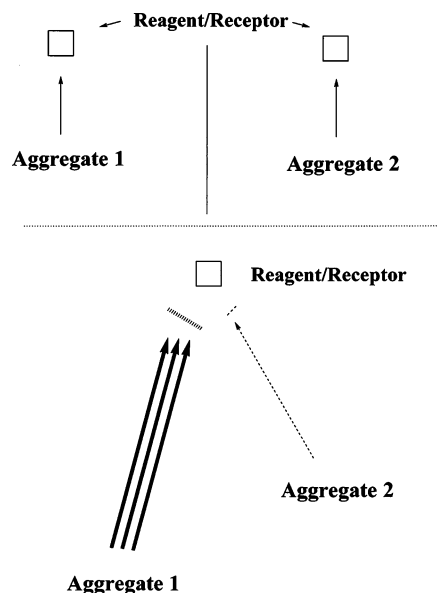


Figure 8. Molecular communication for standard versus nonstandard states. The upper portion depicts two hypothetical aggregates in communication contact with a reagent/receptor under standard-state conditions. The lower portion pertains to a nonstandard state. Here, each aggregate system affects the information transmitted by the other as represented by arrows of different intensities. How the reagent/receptor site registers the information is also affected. The registration effects are depicted by the apertures placed between the reagent/receptor and information sources.

with Aggregate 1—and vice versa. Here, $\Theta \neq 1$ for both parties where their registration effects are denoted by apertures of different widths.

Insights follow from viewing molecules under exceptional circumstances, namely, when their chemical potentials prove identical.^{27,28} Structure–function correlations can be established on informatic/communication grounds by applying an analogous strategy.

For two systems such as those in Figure 8, one has via eq IV-2

$$\begin{aligned} L_1 &= L_1^\circ + \tau \log_e \kappa_1 \Theta_1 \\ L_2 &= L_2^\circ + \tau \log_e \kappa_2 \Theta_2 \end{aligned} \quad (\text{IV-3})$$

A unique condition is then in place if the dispersion length of one system is matched by the other. In this case, $L_1 = L_2$ and

$$\log_e(\kappa_1/\kappa_2) = \tau^{-1}(L_2^\circ - L_1^\circ) + \log_e(\Theta_2/\Theta_1) \quad (\text{IV-4})$$

The above follows from the combination and rearrangement of eq IV-3.

Equation IV-4 asserts the criterion for different sources of information to exert equal impact at a reagent/receptor site. The analogous μ criterion defines the state of chemical equilibrium, where the logarithm of an activity ratio scales with the difference between standard μ values.^{27,28} The correlation framework of this paper has a parallel form. Thus, the logarithm of a κ ratio is proportional to a difference in L° values. In applying eq IV-4, κ ratios are taken to scale with molecular activity as can be measured in a variety of experimental contexts.

Structure–function correlations are identified via libraries of molecules and corresponding activity data. The informatic

Table 1. Ionization Constants of Acetic Acid and Derivative Compounds in Water at 25 °C³⁰

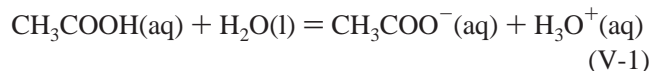
informatic group	molecule	ionization constant
Group I	trimethylacetic acid	9.4×10^{-6}
	acetic acid	1.76×10^{-5}
	vinylacetic acid	4.57×10^{-5}
	chloroacetic acid	1.40×10^{-3}
	bromoacetic acid	2.05×10^{-3}
	dichloroacetic acid	3.32×10^{-2}
Group II	trichloroacetic acid	2×10^{-1}
	phenylacetic acid	5.2×10^{-5}
	acetoacetic acid	2.62×10^{-4}
Group III	iodoacetic acid	7.5×10^{-4}
	cyanoacetic acid	3.65×10^{-3}
	thioacetic acid	4.7×10^{-4}
Group IV/outlier test compounds	propionic acid	1.34×10^{-5}
	<i>n</i> -butyric acid	1.54×10^{-5}
	dibromoacetic acid	4.1×10^{-2} (ref 31)

and communication route presented above utilizes the same data along with L° differences quantified by Brownian processing. This route is demonstrated in the next section.

V. STRUCTURE–FUNCTION CORRELATIONS VIA BROWNIAN PROCESSING

We demonstrate Brownian applications in four different contexts. As with all structure–function and molecular information studies, the results are significant for their organizational and inferential power.^{7–13,25} They enable predictions of chemical activity on the basis of library statistical trends.

A. Acid Ionization in Aqueous Solution. Acetic acid and derivative compounds are readily solvated by water; they further demonstrate a propensity for ionization.²⁹ When acetic acid dissociates in water, the products are hydrogen and acetate ions in the aqueous phase, namely



with the dissociation constant K_A determined by the following ratio at equilibrium:

$$K_A = \frac{[\text{CH}_3\text{COO}^-(\text{aq})][\text{H}_3\text{O}^+(\text{aq})]}{[\text{CH}_3\text{COOH}(\text{aq})]} \quad (\text{V-2})$$

The dissociation activities for a library of acid/water solutions are readily compared via K_A values measured at constant temperatures. Table 1 lists 25 °C data as reported in the Handbook of Chemistry and Physics.³⁰ As would be expected, the K_A values depend on the number and identity of the substituents. The precise structure–function correlations are nontrivial, however. For example, $K_A^{(\text{bromo})}$ is comparable to $K_A^{(\text{chloro})}$, but $K_A^{(\text{iodo})}$ is significantly less than $K_A^{(\text{chloro})}$; $K_A^{(\text{dichloro})} \approx 24 \times K_A^{(\text{chloro})}$, whereas $K_A^{(\text{trichloro})} \approx 140 \times K_A^{(\text{chloro})}$.

Brownian processing was applied to acetic acid and derivative graphs combined with water according to the Section II.C and III procedures. The results led to the group sorting exercised in Table 1 and the log-linear plot of Figure 9. In placing the Figure 9 data points, K_A and L° for acetic acid/H₂O served as the reference values; the point for acetic acid/H₂O, thus, appears at (0,0). Note that the sorting algorithm employed maximized the value of the linear

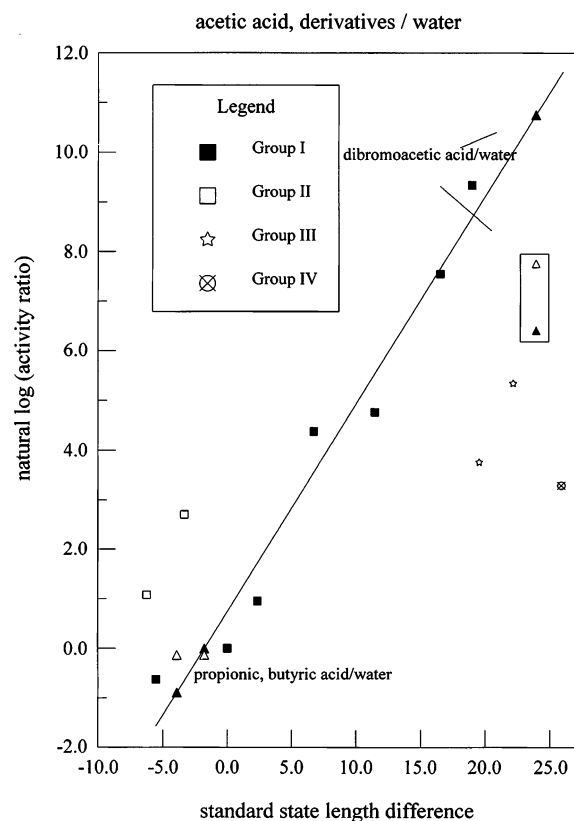


Figure 9. Data for organic acid/water systems. The natural log of K_A has been plotted versus L° differences computed for the Table 1 systems. The reference point for acetic acid/water appears at (0,0). Filled triangles located activity predicted for propionic, butyric, and dibromoacetic acid; open triangles reflect activities reported in refs 23 and 24. The linear correlation coefficient for the filled square isocline is 0.978.

correlation coefficients (>0.950 typically) and the number of standard deviations σ separating each informatic group ($>3\sigma$ typically); minimized were the differences between the slopes affiliated with each group (ca. 10%).

Figure 9, thus, depicts a series of isoclines, the central one accentuated by a best-fit regression line. As described by eq IV-4, each isocline expresses a common slope that is proportional to L° differences and a unique ordinate intercept. Thus, the organic acid/water systems affiliated with a given isocline demonstrate the same informatic bias/distribution parameter (Θ value).

Isoclines furnish a natural way to sort different systems on a message space/molecular communication basis. In our experience, the final sorting details are difficult to predict in advance. Yet, they enable estimates of activity for test compounds on the basis of their informatic structure. Interestingly, sorting along informatic lines helps account for inconsistencies in chemical trends. For example, K_A for iodoacetic acid is markedly disparate from those of bromo- and chloroacetic acid. Brownian methods contribute that iodoacetic acid/water belongs to an entirely different communication set, one shared by cyanoacetic acid/water. Evidently, iodoacetic acid/water and cyanoacetic acid/water process information in a way that cannot be inferred from the Group I behavior.

The results of three predictive exercises are included in Figure 9. These concern propionic, butyric, and dibromoacetic acid combined with water. The activities for the first

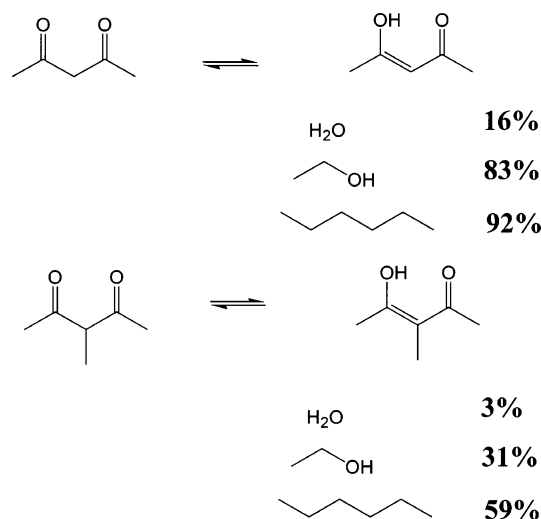


Figure 10. Enolization chemistry of diketones in solution. Percentages refer to the enol content reported in ref 36 for 2,4-pentanedione and 3-methyl-2,4-pentanedione in water, ethanol, and *n*-hexane.

two systems are projected to be affiliated with the acetic acid/water isocline owing to chemical similarities. There is uncertainty, however, surrounding dibromoacetic acid/water; one would predict, in advance, this system to be affiliated with either the Group I or III isoclines. The regression behavior, thus, locates the filled triangles in Figure 9, two possible for dibromoacetic acid/water. For comparison, the open triangles represent the experimental (literature) activities whereby decent agreement is observed for propionic/water and butyric acid/water.

The Handbook of Chemistry and Physics does not report K_A for dibromoacetic acid, but a value is listed in Langes Handbook of Chemistry.³¹ The activity inferred by Brownian processing (filled triangles) is consistent with the effects of electron-withdrawing substituents, although the agreement is only fair with the literature value for dibromoacetic acid/water. Table 1 and Figure 9 address solution chemistry that is elementary. The significant point, however, is new structure–function correlations can be established on informatic and molecular communication grounds.

B. Ketone/Enol Tautomerizations. The degree of enolization depends on a ketone's structure and solvent environment.³² This chemistry was investigated in considerable detail by Meyer and co-workers in the early 1900s.³³ Among other things, Meyer developed a quantitative assay for ketone/enol mixtures that involved rapid titrations with bromine. Extensive use—and improvement—have been made of this procedure over the years, and a variety of spectroscopic techniques have also been applied.^{32,34,35} The significance of this chemical genre cannot be understated. The tautomeric forms of a molecule predicate entirely different responses to halogen reagents. The responses are directly correlated with the structures conveyed by elementary graphs. Tautomerization chemistry testifies as to the utility of line/vertex graphs and digital code in describing small-scale electronic phenomena.

The enolization of 2,4-pentanedione (acetylacetone) and 3-methyl-2,4-pentanedione is represented in Figure 10. The numerical values refer to the percent enol content measured in water, ethanol, and *n*-hexane solutions.³⁶ By and large, a nonpolar aprotic solvent favors the enol side of the reaction,

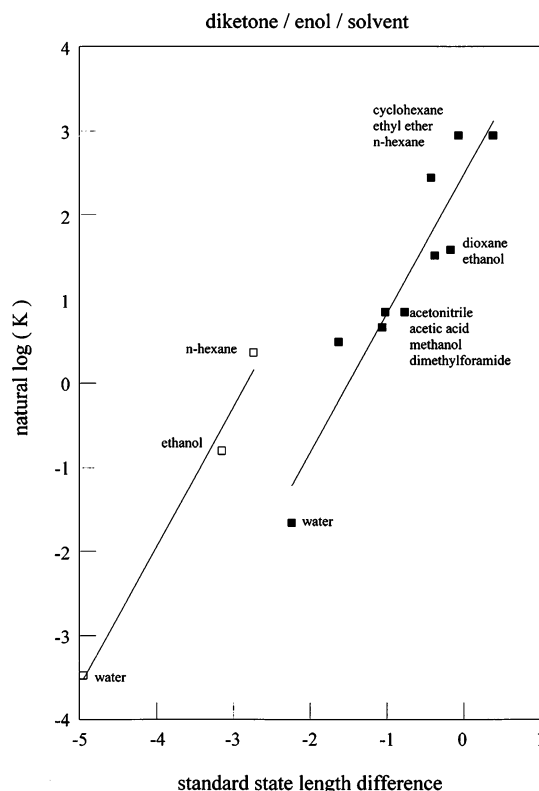


Figure 11. Informatic data for tautomeric systems. The natural log of the tautomerization equilibrium constants K has been plotted versus L° differences. Filled and open squares pertain, respectively, to binary systems of 2,4-pentanedione and 3-methyl-2,4-pentanedione combined with different solvent molecules. The solvent identities are listed for each data point. The correlation coefficients for the filled and open square isoclines are 0.935 and 0.992, respectively. The difference between the best-fit intercept values (Θ parameters) is statistically significant at above the 99% level of confidence.

although the behavior is complicated in quantitative terms. For example, the percent enolization of 2,4-pentanedione in ethanol is comparable to that measured in *n*-hexane. By contrast, there is a factor-of-two disparity for these solvents containing the 3-methyl derivative.

Brownian processing can address chemistry such as in Figure 10. In Figure 11, the logarithms of the tautomerization equilibrium constants K have been plotted versus the L° differences between the enol and diketone reaction sides. Data for 2,4-pentanedione (filled squares) in cyclohexane, ethyl ether, acetonitrile, dioxane, dimethyl foramide, methanol, and acetic acid have been included.³⁷ The data for 3-methyl-2,4-pentanedione is marked by the open squares.³⁶

As shown in Figure 11, Brownian methods identify an isocline and, thus, distribution for each data set. As with Figure 9, the group sorting is that which maximized the correlation coefficients and degree of separation; the difference between the regression line slopes turns out to be about 1%. Unlike in Figure 9, the data point positions within each group depend entirely on the particular solvent identity. The two isoclines demonstrate significantly different intercept values. This means that 2,4-pentanedione in water ($K_{\text{taut}} \approx 0.19$) expresses the same Θ ratio as 2,4-pentanedione in *n*-hexane ($K_{\text{taut}} \approx 11.5$)—and all other binary systems in the informatic distribution. The Θ ratios are distinctly different, however, for solutions containing 3-methyl-2,4-pentanedione. This is the case, despite the structural similarities, of the

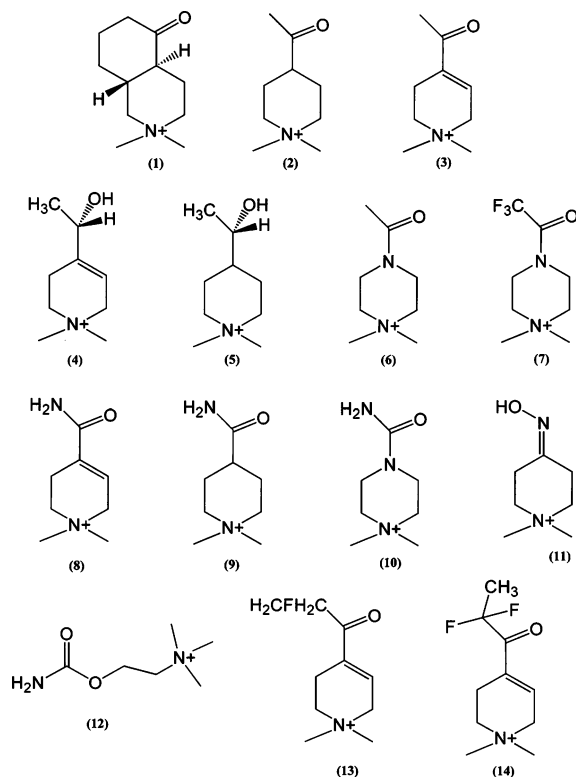


Figure 12. Nicotinic agonist library. The activity of compounds 1–11 has been measured relative to carbamylcholine, compound 12. Structures 13 and 14 refer to test compounds discussed in the text.

Figure 10 diketones and enols. Brownian processing, thus, identifies new structure–function correlations for the solvent-dependent chemistry.

C. Potency of Nicotinic Agonists in Aqueous Solution.

The nicotinic receptor has been studied for several decades.³⁸ The chemical transmitter for the receptor is acetylcholine, mimics of which have been the subject of extensive searches.^{38,39} This research has been multipronged, combining molecular modeling, organic synthesis, and biological activity measurements.^{38–40} The goal has been to understand how specific molecular templates and functionalities tune activity at the receptor site. The problems are formidable given that the receptor structure has not been completely characterized.

Significant attention has been directed at the nicotinic agonists illustrated in Figure 12.^{39,40} The activity of compounds 1–11 relative to carbamylcholine 12 have been measured. The heterocyclic compounds 3 and 7 demonstrate the first and second highest activities, respectively; compound 1 turns out to be the least potent library member. Using a variety of computational tools, several structure–function correlations have been identified, on the basis of hydrogen bonding and receptor Coulombic interactions.^{39,40}

Brownian methods offer an approach on informatic and molecular communication grounds. For the results illustrated here, water was used as the biologically relevant solvent combined with each of the Figure 12 compounds. The informatic data appear in Figure 13, where the logarithms of activity ratios have been plotted versus L_o differences. In constructing the plot, the activity and L_o for carbamylcholine/water served as the reference values.

Figure 13 asserts three informatic distributions via isoclines while compound 11 appears as an outlier. As in Figure 9

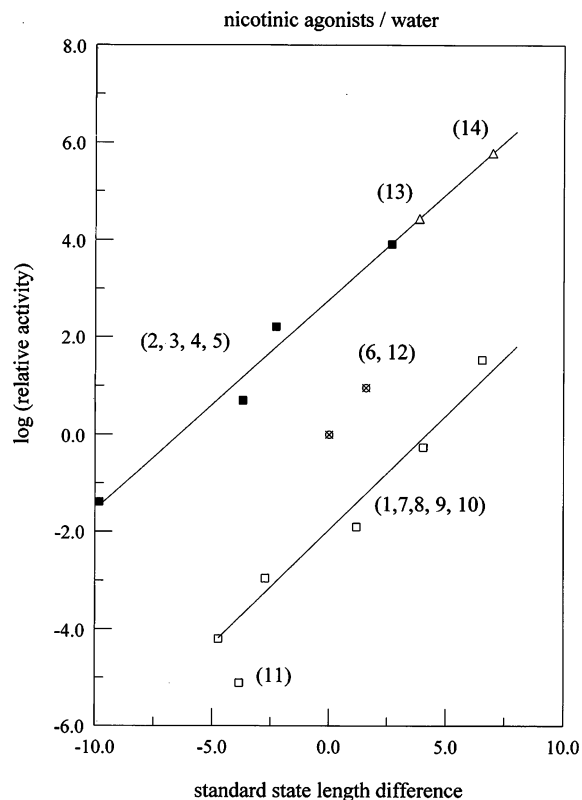


Figure 13. Informatic data for nicotinic agonist/water systems. Numerical labels refer to Figure 12 compounds. The correlation coefficients for the filled and open square isoclines are 0.986 and 0.987, respectively. The difference between the best-fit intercept values (Θ parameters) is statistically significant at the 99% level of confidence.

(acid chemistry), the distributions promote a group sorting of compounds along informatic lines. In so doing, activity predictions for test compounds are facilitated. For example, Brownian methods were extended to 13 and 14 of Figure 12 combined with water. The structures of these molecules are advocated via the most potent library members, 3 and 7. In essence, 13 is a monofluoro derivative of 3, while 14 is a combination methyl and difluoro derivative. Note that the use of fluorine atoms, as opposed to other halogens, is recommended via the structure and measured activity of 7.

The structural similarities between 13 and 3 and between 14 and 3 facilitate the distribution assignments. In turn, Brownian methods provide the L_o computations and activity location within the distribution. The predicted activities for the test compounds are marked by the open triangles in Figure 13. One observes that the test compounds offer a potential for greater activity than all of the other library members. In turn, Brownian methods furnish molecular design suggestions on informatic grounds.

D. Inhibitory Efficacy of Clavulanic Acid and Derivatives. Research on the β -lactam antibiotics has included studies of materials that accelerate the deactivation of these compounds.⁴¹ It was recognized early on that bacteria produce β -lactamases that act on penicillin and cephalosporin by catalyzing the hydrolysis of the β -lactam ring. However simple, this chemistry obliterates the therapeutic value of the antibiotics.^{41, 42}

Interestingly, clavulanic acid and derivatives were discovered to inhibit a variety of β -lactamases.⁴³ As a result, these molecules could be combined with antibiotics so as to

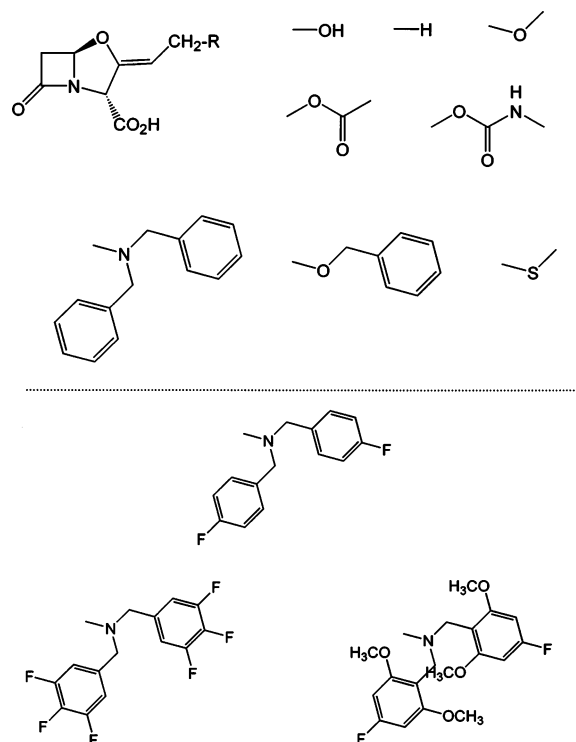


Figure 14. Clavulanic acid and derivatives. The upper portion shows the template of clavulanic acid along with several R substituents. The lower portion shows three test substituents proposed on the basis of the library informatic structure.

counter bacteria resistance. In exploratory work, the activities of various clavulanic acid compounds were quantified via IC_{50} values, the concentrations required to effect a 50% inhibition of the bacteria-manufactured enzymes. The principal structure for clavulanic acid is illustrated in the upper left-hand portion of Figure 14. The upper half of the figure includes several R substituents specific to derivative compounds; these have been synthesized and investigated for their inhibitory activity.⁴⁴ For our part, Brownian processing was applied to all of these molecules combined with water, the biologically relevant solvent.

In Figure 15, the logarithm of IC_{50} for *Citrobacter freundii* β -lactamases has been plotted versus L^o . Only one distribution (isocline) is evident as marked by the filled squares corresponding to five library members: the hydroxyl, hydrogen, methyl ether, benzyl ether, and phenylamine derivatives of clavulanic acid combined with water. The points for the other compounds—acetate, carbamate, and thioether derivatives—appear far removed (several standard deviations) from the regression line.

One learns the following from Figure 15. The acetate and carbamate derivatives of clavulanic acid pose the lowest IC_{50} and, thus, the greatest efficacy at thwarting *Citrobacter freundii* resistance via β -lactamases. From an informatic/communication standpoint, however, the actions of these inhibitors do not adhere to any apparent distribution. It is, then, difficult to propose, on a rational basis, test compounds that resemble the acetate and carbamate derivatives and improve upon their efficacy. This is where the filled square data assist.

Three test substituents (R groups) for clavulanic acid are illustrated in the lower half of Figure 14. These are inspired by the phenylamine substituent—that which promotes in

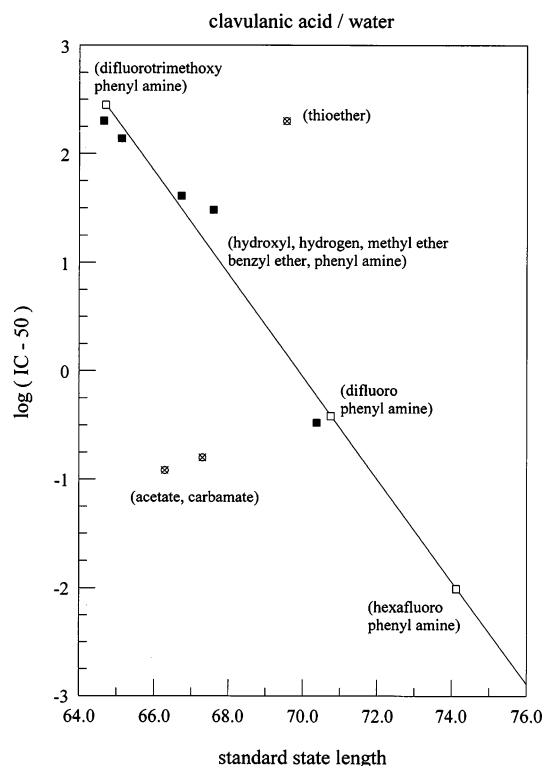


Figure 15. Data for clavulanic acid/water systems. Labels refer to Figure 14 compounds. The correlation coefficient for the filled square regression line is 0.971.

clavulanic acid the lowest IC_{50} in one subset of library compounds. This subset is identified via Brownian processing, the activity dependence on L^o described by eq IV-4. The results for the test compounds are represented by the open squares in Figure 15. One finds that IC_{50} for the difluorophenylamine compound is predicted to be comparable to that of the unsubstituted phenylamine: the derivatization, thus, offers no significant improvement. By contrast, the inhibitory activity of the hexafluorophenylamine compound is projected to exceed that of *all* the library members, including the acetate and carbamate derivatives. Other compounds can be proposed, for example, using the difluorotrimethoxyphenylamine substituent appearing in Figure 14. The inhibitory effect for this derivative is predicted to be lowest of all the library members. On informatic and molecular communication grounds, its synthesis and biological investigation should not be given high priority, for example, in a discovery and design program. Brownian methods offer ways to infer rationally how chemical activity can be diminished.

VI. DISCUSSION

For chemists, the communication of angstrom-scale events relies on digital code and two-dimensional (2D) graphics. However simple and idealized, the 100-plus-year-old system enjoys astonishing precision and versatility. A naturally occurring protein with 100 chiral centers offers $2^{100} \approx 1.3 \times 10^{30}$ optical isomers. There is a unique graph and code assembly that can be constructed for each one of these molecules. Moreover, abundant substrates for each compound can be depicted via resonance graphs. For chemists, the combination of digital and graphic communication holds singular advantages over analogue.

This paper focused on the informatics of aggregation states and attendant solvent effects. Brownian processing was adapted to these kinds of states so that the message spaces for a variety of real-life systems could be assembled in digital terms. The aim was to develop practical methods for correlating chemical structure with function on informatic grounds. Among other things, the task necessitated a deeper probe of a digital basis for communication among molecules (Section IV) in solvent environments. The results and their applications are, thus, significant on three accounts.

The first is that code and graphics such as in Figure 10 (diketones and enols) convey the key covalent properties of carbon-based molecules. The line, letter, and integer assemblies for 2,4-pentanedione ably assert which atoms are most strongly coupled to one another and what spatial regions pose the greatest charge density. Brownian processing, however, takes matters a step further. When the random walk techniques are applied to an aggregate state of 2,4-pentanedione plus ethanol, contact is made with noncovalent traits. This is brought about by treating graph-and-code combinations as zipped files containing random variables, that is, as information sources. The contact is worth making because noncovalent properties tie to chemical activity in a critical but difficult-to-intuit way—the extent to which 2,4-pentanedione enolizes hinges on whether the noncovalently attached solvent is ethanol or dioxane or acetic acid. At the molecular level, solvent–solute interactions are notoriously complex to unravel.⁴⁵ Brownian methods sidestep much of the complexity by considering the possible digital messages subsidiary to the interactions.

Second, the results are significant on thermochemical grounds. The code and graphs such as in Figure 7 (cyclooctatetraene/cyclohexane, etc.) connect effectively with enthalpy, heat capacity, and other extensive quantities. Brownian methods, however, introduce a new measure L° for organic materials. Thus, systems with relatively low L° values, for example, $\text{CH}_4/\text{CH}_3\text{D}$, offer few internal markers that can be discriminated by a Brownian processor, compared with $\text{CHBrCl}/\text{CH}_2\text{O}$. While the former system contains energy and, thus, a capacity for performing work, it poses little facility for controlling how work gets done. Brownian methods can affiliate any combination of molecules with an informatic dispersion length. In so doing, the relations are extended between chemical graphs and their underlying thermodynamic functionality.

The third point of significance concerns structure–function relations. Many chemical problems entail identifying what set(s) a particular molecule belongs to. One assigns, for example, 2,4-pentanedione and acetic acid to different sets (textbook chapters, tables, etc.) on the basis of their functional group structure and chemical reactivity. In turn, 3-methyl-2,4-pentanedione is placed in the same set as 2,4-pentanedione; iodoacetic acid is delegated to the same collection as acetic acid. The molecules of different sets can still be compared; for example, for both acids and ketones, their reaction behavior involves carbon–oxygen functionalities. But such comparisons yield insight only to a point because of the gross chemical disparities.

Set assignment issues pose sorting problems whereby Brownian processing serves as an additional tool. By assembling digital message strings based on possible collisions in solution, Brownian methods connect with the

operational subtleties distinguishing 3-methyl-2,4-pentanedione from 2,4-pentanedione and iodoacetic acid from acetic acid. The information relayed via chemical graphs and code facilitates high-level sorting. The examples of Section V demonstrate how a lower-level sorting can be exercised on informatic and molecular communication grounds. Such is both useful and accessible given the precision of the chemical code in graph form.

VII. CLOSING

In the work for this paper, Brownian processing was developed further as a means of probing molecular information and its communication, with aggregation states taken into account. Examples were offered relating structure and function, with the methodology extendable to other systems and contexts.

Two additional problems were implicit, however. Figure 1 asserts that noise is an inescapable element of information communication and reception. Thus, an imperfect processor—or one operating in real time—will sometimes read (C–H) as (C–C), (C–C) as (C=C), and so forth. The effects of noise, thus, represent another critical issue in the informatics of molecules and structure–function correlations. A second issue arises because information quantities are necessarily layered. A packet of 32 ($= 2^5$) manila folders offers five bits of information at one level of random queries. If the folders contain printed materials, new levels are posed by the pages, paragraphs, words, and so forth. In an analogous way, organic molecules express high-level information via their atom/covalent bond networks; electronic orbitals, however, pose information capacity at a still deeper level. The latter type of information must be taken into account for a more comprehensive understanding of chemical processes. Our current efforts focus on these two areas, and the results will be reported in a forthcoming paper.

ACKNOWLEDGMENT

The author is grateful to Professors James Babler, Daniel Becker, and David Crumrine for helpful discussions about molecular structure and function. The author is similarly appreciative of two referees for comments and constructive criticism. Thanks are also extended to Matthew Schulmerich (NSF/REU student) for assistance during the early stages of the Brownian processing research.

REFERENCES AND NOTES

- (1) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, 27, 379. The italics are Shannon's. Excerpts from Shannon's paper are prominent in at least one textbook. See McEliece, R. J. *The Theory of Information and Coding*; *Encyclopedia of Mathematics and Its Applications*; Rota, G. C., Ed.; Addison-Wesley: Reading, Massachusetts, 1977; Vol. III.
- (2) Nyquist, H. Certain Factors Affecting Telegraph Speed. *Bell Syst. Tech. J.* **1924**, April, 324.
- (3) Hartley, R. V. L. Transmission of Information. *Bell Syst. Tech. J.* **1928**, July, 535.
- (4) See, for example, Fisher, R. A. Theory of Statistical Estimation. *Proc. Cambridge Philos. Soc.* **1925**, 22, 700.
- (5) Ash, R. B. *Information Theory*; Dover Publications: New York, 1990; Chapter 1.
- (6) Exceptions involve gas-phase reactions amenable to high-resolution spectroscopy. See Levine, R. D. The Information Theoretic Approach to Molecular Reaction Dynamics. *Annu. Rev. Phys. Chem.* **1978**, 29, 59.

- (7) Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix, and Molecular Branching. *J. Chem. Phys.* **1978**, *67*, 4517.
- (8) Bertz, S. H. The First General Index of Molecular Complexity. *J. Am. Chem. Soc.* **1981**, *103*, 3599.
- (9) Basak, S. C.; Mekenyan, O. Topological Indices and Chemical Reactivity. In *Graph Theoretical Approaches to Chemical Reactivity*; Bonchev, D., Mekenyan, O., Eds.; Kluwer Academic: Dordrecht, The Netherlands, 1994; Chapter VIII, p 221.
- (10) Godden, J. W.; Stahura, F. L.; Bajorath, J. Variability of Molecular Descriptors in Compound Databases Revealed by Shannon Entropy Calculations. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 796. Stahura, F. L.; Godden, J. W.; Xue, L.; Bajorath, J. Distinguishing Between Natural Products and Synthetic Molecules by Descriptor Shannon Entropy Analysis and Binary QSAR Calculations. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1245. Godden, J. W.; Bajorath, J. Differential Shannon Entropy as a Sensitive Measure of Differences in Database Variability of Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1060. Godden, J. W.; Bajorath, J. Chemical Descriptors of Distinct Levels of Information Content and Varying Sensitivity to Differences Between Selected Compound Databases Identified by SE-DSE Analysis. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 87.
- (11) Basak, S. C.; Grunwald, G. D.; Niemi, G. J. Use of Graph-Theoretic and Gemoetrical Molecular Descriptors in Structure–Activity Relationships. In *From Chemical Topology to Three-Dimensional Geometry*; Balaban, A. T., Ed.; Plenum Press: New York, 1997; Chapter IV, p 73.
- (12) Bajorath, J. Selected Concepts and Investigations in Compound Classification, Molecular Descriptor Analysis, and Virtual Screening. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 233.
- (13) Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structure*. Research Studies Press-Wiley: Chichester, U. K., 1993.
- (14) Graham, D. J.; Schacht, D. Base Information Content in Organic Molecular Formulae. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 942.
- (15) Graham, D. J. Information Content in Organic Molecules: Structure Considerations Based On Integer Statistics. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 215.
- (16) Graham, D. J.; Malarkey, C.; Schulmerich, M. V. Information Content in Organic Molecules: Quantification and Statistical Structure Via Brownian Processing. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1601.
- (17) Graham, D. J.; Schulmerich, M. V. Information Content in Organic Molecules: Reaction Pathway Analysis Via Brownian Processing. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1612.
- (18) Bennett, C. H. Thermodynamics of Computation – A Review. *Int. J. Theor. Phys.* **1982**, *21*, 905. See also Feynman, R. P. *Feynman Lectures on Computation*; Hey, A. J. G., Allen, R. W., Eds.; Addison-Wesley: Reading, Massachusetts, 1996.
- (19) Tribus, M.; McIrvine, E. C. Energy and Information. *Sci. Am.* **1971**, *225*, 179.
- (20) 3-one-5-methyl-cyclohexene was the featured molecule for Brownian processing demonstrations in ref 16.
- (21) See, for example, Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.
- (22) For the field calculations, a molecular mechanics routine obtained from CambridgeSoft Corporation (Cambridge, MA 02139) was used in conjunction with several valence bond approximations.
- (23) Cox, J. D.; Pilcher, G. *Thermochemistry of Organic and Organometallic Compounds*; Academic Press: New York, 1970.
- (24) Robinson, G. W. Intensity Enhancement of Electronic Transitions by Weak Intermolecular Interactions. *J. Chem. Phys.* **1967**, *46*, 572.
- (25) Kuntz, I. D.; Meng, E. C.; Shiochet, B. K. Structure-Based Molecular Design. *Acc. Chem. Res.* **1994**, *27*, 117.
- (26) Graham, D. J. Manuscript in preparation.
- (27) Lewis, G. N.; Randall, M. (Revised by Pitzer, K. S.; Brewer, L.); *Thermodynamics*; McGraw-Hill: New York, 1991; Chapter 14.
- (28) Denbigh, K. *The Principles of Chemical Equilibrium*; Cambridge University Press: Cambridge, MA, 1971; Chapter 2.
- (29) Roberts, J. D.; Caserio, M. C. *Basic Principles of Organic Chemistry*; W. A. Benjamin: New York, 1965; Chapter 14.
- (30) *Handbook of Chemistry and Physics*; Weast, R. C., Ed.; Chemical Rubber Co.: Cleveland, OH, 1972; pp D120–D121. One notes K_A values for acetic acid derivatives reported in other texts to be somewhat different. See Morrison, R. T.; Boyd, R. N. *Organic Chemistry*, 6th ed.; Prentice Hall: Englewood Cliffs, NJ, 1992; Chapter 19, p 735. March, J. *Advanced Organic Chemistry*, 2nd ed.; McGraw-Hill: New York, 1977; Chapter 8, p 240.
- (31) *Langes Handbook of Chemistry*, 13th ed.; Bean, J. A., Ed.; McGraw-Hill: New York, 1985.
- (32) Toullec, J. Keto–Enol Equilibrium Constants. In *The Chemistry of Enols*; Rappoport, Z., Ed.; Wiley: New York, 1990; Chapter 6.
- (33) Meyer, K. H.; Kappelmeier, P. Über die Tautomerie des Acetessigesters. (Über Keto–Enol-Tautomerie III). *Chem. Ber.* **1911**, *44*, 2718. Meyer, K. H. Über den Zusammenhang zwischen Konstitution und Gleichgewicht bei keto–enol-desmotropen Verbindungen. (Über Keto–Enol Tautomerie. VI). *Chem. Ber.* **1912**, *45*, 2846.
- (34) Conant, J. B.; Thompson, A. F., Jr. The Free Energy of Enolization in the Gaseous Phase of Substituted Acetoacetic Esters. *J. Am. Chem. Soc.* **1932**, *54*, 4039.
- (35) See, for example, Folkendt, M. M.; Weiss-Lopez, B. E.; Chauvel, J. P., Jr.; True, N. S. Gas-Phase ¹H NMR Studies of Keto–Enol Tautomerism of Acetylacetone, Methyl Acetoacetate, and Ethyl Acetoacetate. *J. Phys. Chem.* **1985**, *89*, 3347.
- (36) House, H. O. *Modern Synthetic Reactions*. W. A. Benjamin: Menlo Park, CA, 1972; Chapter 9.
- (37) Pocker, Y.; Spyridis, G. Modulation of Tautomeric Equilibria by Ionic Clusters. Acetylacetone in Solutions of Lithium Perchlorate–Dethyl Ether. *J. Am. Chem. Soc.* **2002**, *124*, 10373.
- (38) Beers, W. H.; Reich, E. Structure and Activity of Acetylcholine. *Nature* **1970**, *228*, 917.
- (39) Waters, J. A.; Spivak, M.; Hermsmeier, M.; Yadav, J. S.; Liang, R. F.; Gund, T. M. Synthesis, Pharmacology, and Molecular Modeling Studies of Semirigid, Nicotinic agonists. *J. Med. Chem.* **1988**, *31*, 545. Spivak, C. E.; Yadav, J. S.; Shang, W. C.; Hermsmeier, M.; Gund, T. M. Carbamyl Analogues of Potent, Nicotinic Agonists: Pharmacology and Computer Assisted Molecular Modeling Study. *J. Med. Chem.* **1989**, *32*, 305.
- (40) Gund, T. Molecular Modeling of Small Molecules. In *Guidebook on Molecular Modeling in Drug Design*; Cohen, N. C., Ed.; Academic: New York, 1996.
- (41) Demarco, P. V.; Nagarajan, R. In *Cephalosporins and Penicillins*; Flynn, E. H., Ed.; Academic: New York, 1972; Chapter 8.
- (42) Pratt, R.; Dufrenoy, J. Cytochemical Interpretation of the Mechanism of Penicillin Action. *Bacteriol. Rev.* **1948**, *12*, 79.
- (43) Howarth, T. T.; Brown, A. G. Clavulanic Acid, a Novel β -Lactam Isolated from *Streptomyces Clavuligerus*; X-ray Structure Analysis. *J. Chem. Soc., Commun.* **1976**, 266.
- (44) Brown, A. G.; Francois, I. Clavulanic Acid and Related Compounds: Inhibitors of β -Lactamase Enzymes. In *Medicinal Chemistry*, 2nd ed.; Ganellin, C. R.; Roberts, S. M., Eds.; Academic: London, 1993.
- (45) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An overview of Methods based on Continuous Distributions of Solvent. *Chem. Rev.* **1994**, *94*, 2027.

CI050101M