# On the Parabolic Relationship between Drug Potency and Hydrophobicity 

JAMER W. MCl゙aR1.ANい<br><br>


#### Abstract

A new derivation of the Hansch-Fnjita relationship between drag potency and hydrophobirity is premted. The argument is based on probability concepts and from this vantage shows that the relationship is not, sinfety speaking, parabolic, but it may be considered to be so for all practical purposes. This upproach to the problem allows one to consider mantiple and branched ronkes of ding mavel simmlaneonsly: no change in the basie rebatonship is incured thereby. Further analysis of these remalts show that practirally all ding moleralo that reach their receptors travel by equivalent most-probable rontes: moleones traveling by other mates nanally do not result in drng-receptor complexes.


In 1964 Hansch and F'ujita suggested that the probability of a drug penetrating a cell membrane by a passive transport mechanism is a parabolic function of the drug's hydrophobic bonding capacity. ${ }^{1}$ These workers also proposed the new extrathermodynamic substituent constant $\pi$ as a measure of hydrophobicity. The definition of $\pi$ parallels that for $\sigma$ in the Hammett relationship:

$$
\begin{equation*}
\log (k / l)_{X}-\log (h l)_{11}=\pi \tag{1}
\end{equation*}
$$

In eq $1(k / l)_{\mathrm{X}}$ is the partition coefficient ${ }^{2}$ for a member of a drug series which bears the substituent X while ( $k / /)_{15}$ is the partition coefficient of the unsubstituted parent compound; a standard solvent pair is used to determine these values. The Hansch-Fíujita relationship is given either by eq 2 or by eq 3. The potency

$$
\begin{equation*}
\log 1 / C_{X}=a \pi_{X}^{2}+b \pi_{X}+c \tag{-1}
\end{equation*}
$$

$$
\begin{equation*}
\log 1 / c_{\mathrm{X}}=a\left[\left.\log (k /)_{\mathrm{X}}\right|^{2}+b \log (k /)_{\mathrm{X}}+c\right. \tag{3}
\end{equation*}
$$

index, $\log 1 / C_{X}$, is considered to be directly proportional to the logarithm of the probability that the drug will cross one or more cell membranes in a given system. The concentration $C_{X}$ is that result ing in some standard biological response, e.g.. EDD ${ }_{g n}$ and $\mathrm{LD}_{\text {io }}$. The constants $a . b$, and $c$ are characteristics of the biological test $s y$ stem. The initial argument for of 12 and $3^{3}$ was intuitive,' but recently the Hansch school has advanced a kinetic argument based on a model of altemating aqueous and lipid (membrane) phases." Using a computer they found solutions to a set of complex differentiat equations for selected values of $k / l$. Values for $\log C$ at an whitrarily chosen receptor region were obtained. A plot of these data against $\log t / l$ gave a set of points which were fitted to a parabolic curve by the method of least scuares. The fit wits not perfect. and the set of points was skewed slightly away from the maximum of the regression curve, but for all pactical purposes it could be said that the parabolic relationship under discussion was confirmed by the mathematieal analysis of a suitable model syst ens. White working on a different approach to this problem, I independently struck upon a new derivation of this parabolic relationship which to my mind is easier to grasp since only an under-

[^0]standing of elementary probability conceptrand algebra is required. This approach has other adrantages which will become apparent further on.

For most drugs, getting to the recepter site is a chance affair. A molecule given at the site of administration normally wanders somdomly about in some initial an fuous phase. and in the conse of time is aboubed by the lipid phase of some cell membrate. At this point it may either be reabsorbed by the initial aqueots phase or to may penctrate the membrame by being extracted by the aqueous phase on the other side. This proceses is repeated a number of times matil the drug molecone finally roaches the aquerns phase from which the drug receptor complex can be fermed directly in one step. Figure 1 illustrates this andition. The figure in not

 branes in a byoothetienl bologieal syone The me:mangs of the varions symbols ate exphaned in the tex
meant to represent actual structure of cell membrames. but rather designates the physical situation important to the passive transport of drags, i.e.. the relationship between the penetration of a drug to its partition coefficient which has bern meognized in various forms since the pioneering work of Meyer and Orertom.3

In ligure 1. ato repmesuts the agueche phase where the drug is intially applacd; hap reparents the firet lipid phase the drug cuters. It this point the droge can either return to arn or anter an. The intemenes are
 such that agueons phases are designated bre even mombered subseripts white the lipid phases are designatiod by odd: than the final aguerous phase from which the drug-receptor comphex com he formed will be ath, wher $n$ is an eren mmbre. The rate constant which determines how readily an mocharged molecular specirs at :a

[^1]given concentration will enter a lipid phase is $k$ while the rate constant for the reverse process is $l$. The partition coefficient is, of course, $k / l$. It is assumed here that the nature of the intervening aqueous phases are essentially the same as aq, and correspondingly the nature of the intervening lipid phases are essentially the same as lip. Thus, regardiess where the drug is in the maze of inter- and intracellular fluids and membranes, $k$ and $l$ remain essentially unaltered.

If the rate at which the drug-receptor complex forms is slow with respect to the speed that an equilibrum condition is approached in these intervening phases, then the various lipid barriers will have little influence upon the potency of a drug because the concentration of drug in all aqueous phases will be essentially the same. Thus, under equilibrium conditions, the potency of a drug should be independent of its partition coefficient insofar as getting the drug to the receptor site is the only role the partition coefficient plays. Where the partition coefficient is a direct reflection of a drug's hydrophobic bonding capacity in the drug-receptor complex, then it will once again have an important influence on the drug's potency.

More often than not., however, the drug does not. have sufficient time to establish an equilibrium condition in a biological system. Besides getting to its intended target, it also wanders to other corners of the system where it is metabolized or excreted. Only a small fraction of the total drug administered actually reaches the receptor site, and it must do so within a limited time or eise the reservoir of drug in aqo will be depleted through other losses to the point where it will no longer be able to maintain an adequate concentration at the more distant receptor site $\mathrm{aq}_{n}$. Under the conditions of determining the potency of a drug, just enough compound is administered to effect a submaximal response. From the foregoing discussion one would expect and frequently observes that the response appears soon after the administration of the drug, i.e., when the drug in the reservoir phase is at its maximum concentration, and then the response fades as the concentration is reduced through various losses. In this situation the standard biological response should occur before a significant amount of drug has left aq $0_{0}$. Therefore, it can be assumed that the concentration of drug molecules at the receptor site is quite small with respect to the concentration in the reservoir phase, and it can be calculated by multiplying the concentration in aqo times the probability of a drug molecule reaching the receptor site. The concentration in aq is, of course, that which gives the standard response and forms the basis of the potency index, $\log 1 / C$. It remains for us to estimate the probability factor. Naturally one should not expect to do this in an absolute sense, but there should be some means to estimate the relative probabilities of the various members of a drug series from the model given in I'igure 1. The following is offered as such a method.

The probability of a molecule reaching $\mathrm{aq}_{n}$ from $\mathrm{aq}_{0}$ is given by eq 4. Eq 4 assumes that there is only one

$$
\begin{equation*}
P_{0, n}=P_{0,1} \cdot P_{1,2} \cdot P_{2,3} \cdot P_{3,4} \cdot \ldots P_{n-2, n-1} \cdot P_{n-1, n} \tag{4}
\end{equation*}
$$

path to the receptor and that there is an exact (although unknown) number of aqueous-lipid interfaces to cross. While it is highly improbable that there is only one
pathway available to a drug, it is convenient to develop this equation at this point. Branched and alternate pathways will be discussed later on. As a simplifying assumption we will consider that the probability of a neutral molecule moving from an aqueous phase into a lipid phase is the same throughout the system, such that eq 5 is true.

$$
\begin{equation*}
P_{0,1}=P_{0,3}=P_{n-2, n-1} \tag{5}
\end{equation*}
$$

Similarly, the reverse situation is given by eq 6 . By

$$
\begin{equation*}
P_{1,2}=P_{3,4}=P_{n-1, n} \tag{6}
\end{equation*}
$$

combining eq 5 and 6 with eq 4 , eq 7 is obtained. It

$$
\begin{equation*}
P_{0, n}=\left(P_{0,1}\right)^{n / 2} \cdot\left(P_{1,2}\right)^{n / 2} \tag{7}
\end{equation*}
$$

will be assumed here that the probability of a drug transferring from $a q_{0}$ to $l i p_{1}$ is determined at the interface of the phases. Not all drug molecules reaching the interface will succeed in penetrating it; the preference of entering lip $p_{1}$ or being reflected to $\mathrm{aq}_{0}$ is determined by the relative values of $k$ and $l$. When $k$ is larger than $l$ most drug molecules will readily enter lip ${ }_{1}$ and a lesser fraction will return to aq; when $l$ is greater than $k$ the reverse situation obtains. Thus, within a particular time span the number of molecules entering lip $p_{1}$ will be proportional to $k$, while the total number of molecules presented the opporiunity to enter lip ${ }_{1}$ will be proportional to the sum of $k$ and $l$. Hence, $P_{0,1}$ is given by eq 8 . Division of the upper and lower parts of

$$
\begin{equation*}
P_{0,1}=\frac{k}{k+l} \tag{8}
\end{equation*}
$$

the fraction in eq 8 by $l$ gives eq 9 . In the reverse

$$
\begin{equation*}
P_{0,1}=\frac{k / l}{k / l+1} \tag{9}
\end{equation*}
$$

situation, $P_{1.0}$ is the probability of a molecule in lip getting into $a q_{0}$, and it is given by eq 10 . However,

$$
\begin{equation*}
P_{1.0}=1-P_{0,1} \tag{10}
\end{equation*}
$$

according to the model given in Figure 1, $P_{1,0}$ is not different from $P_{1, \ldots}$, hence:

$$
\begin{equation*}
P_{1,2}=1-P_{0,1} \tag{11}
\end{equation*}
$$

Substituting this new expression into eq 7 we have:

$$
\begin{gather*}
P_{0, n}=\left(P_{0,1}\right)^{n / 2}\left(1-P_{0.1}\right)^{n / 2}  \tag{12}\\
P_{0, n}=\left(\frac{k / l}{k / l+1}\right)^{n / 2}\left(1-\frac{k / l}{k / l+1}\right)^{n / 2}  \tag{13}\\
P_{0, n}=\left(\frac{k / l}{k / l+1}\right)^{n / 2}\left(\frac{1}{k / l+1}\right)^{n / 2}  \tag{14}\\
P_{0, n}=\frac{(k / l)^{n / 2}}{(k / l+1)^{n}} \tag{15}
\end{gather*}
$$

Thus, according to eq 15 , the probability of a drug reaching a receptor site is a function of its partition coefficient and the number of intervening aqueouslipid interfaces between $\mathrm{aq}_{0}$ and $\mathrm{aq}_{n}$.

As eq 15 stands it is difficult to envision even qualitativery how the probability is changing as $k / l$ and $n$ vary. However, some insight is gained by examining Table I which gives solutions to eq 15 for selected values of $k / l$ and $n$. In situations where the drug must pass through
T.abas: 1

Probability of Dreg Cronsing; $n$ Intmpates: Po, a

a Foxept for the case where $n=1$, all probabilities were calculated from eq 15 ( wee text). ${ }^{n}$ Catrulated from eq 9 tsee text).
at least one lipid phase, then increasing difficulty is met as (i) the partition coefficient becomes very large or very small, and (ii) the number of lipid barriers increases. Where $n>1$, a maximum probability is found when $k / l$ is equal to unity $(1.00)$. Where $n=2$ or more, the probability of a drug reaching the receptor region drops to nil when the partition coefficient is infinite or is zero. This mathematical conclusion is intuitively satisfying, since if a drug has an infinite partition coefficient, we can see how the drug would be readily taken up by lipu, but it would never be able to leave that phase; on the other hand, a drug which as a parition coefficient of zero can never get into lip in the first place, and hence is blocked from reaching aqu.

When the probabilities from Table I are plotted against $\log k / l$ the family of curves in Figure 2 is ob-


Fignre :- - The probability of a drag reaching a reeptor region (aqn) as a function of the $\log$ partition coefficient ( $\log k / l$ ) and of $n$, the mamber of interfaces separating the site of drig application from the receptor region.
tained. Each member of this fanily has its maximum value at $\log k / l=0$, and is reminiscent of a normal distribution eurve.

It is of interest to see if the probability argument of the present work gives conclusions similar to those of the Hansch school of thought. As a first step one can convert the probabilities of Table I into their logenthmic form as ham been done in Table II. A plot of these numbers against the corresponding values for log $k /$ results in the family of emves shown in Figure 3. ( ${ }^{3}$

 ordinate me both in log mats.
inspection one is tempted to believe that the curves are parabolas. But closer analysis reveals that this is not the case. For any particular set of data in Table II where $n>1$, a parabola can be fitted to the set of point* $\log P_{0, n}, \log k / 1$. Regardless of the value of $n$ chosen, the correlation coefficient of the regression curve is always $r=0.98 \%$, i.e. $97.0 \%$ of the variance is accounted for by regression. The "unexplained" $3.0 \%$ of the variance is too large to be result of round-off error. However, when one examines the differences between the input values of $\log P_{0, n}$ and the values calculated from the regression equation it becomes obvious that the differences themselves are a function of $\log k / l$ and $n$ (see l"igure 4). Because these differences are not random it must be concluded that the

Table II
Logarithm of Probability of Drug Crossing $n$ Interfaces: Log $P_{0, n}$

| $k / l$ | $n=0$ | $n=1$ | $n=2$ | $n=3$ | $n=4$ | $n=5$ | $n=6$ | $n=8$ | $n=10$ | $\log k / l$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0.000 | 0.000 | $-\operatorname{Inf}$ | - Inf | - Inf | - Inf | - Inf | - Inf | - Inf | - Inf | - Inf |
| 0.001 | 0.000 | -3.00 | $-3.00$ | -4.50 | -6.00 | -7.50 | -9.00 | $-12.00$ | $-15.00$ | -3.00 |
| 0.003 | 0.000 | $-2.52$ | $-2.52$ | -3.79 | $-5.05$ | -6.31 | $-7.57$ | $-10.10$ | $-12.63$ | -2.52 |
| 0.01 | 0.000 | -2.00 | -2.00 | -3.01 | -4.00 | -5.02 | $-6.03$ | -8.03 | $-10.04$ | -2.00 |
| 0.03 | 0.000 | -1.54 | $-1.55$ | $-2.32$ | -3.10 | -3.87 | $-4.65$ | -6.19 | -7.74 | $-1.52$ |
| 0.1 | 0.000 | $-1.04$ | $-1.08$ | -1.62 | -2.17 | -2.70 | $-3.25$ | -4.34 | -5.41 | $-1.00$ |
| 0.3 | 0.000 | -0.64 | $-0.75$ | $-1.13$ | $-1.51$ | -1.89 | $-2.22$ | $-3.00$ | $-3.75$ | -0.22 |
| 1 | 0.000 | $-0.30$ | $-0.60$ | -0.90 | -1.20 | -1.51 | $-1.81$ | -2.41 | -3.01 | 0.00 |
| 3 | 0.000 | -0.13 | -0.73 | -1.09 | $-1.45$ | -1.82 | $-2.10$ | -2.91 | -3.63 | 0.48 |
| 10 | 0.000 | -0.04 | $-1.08$ | -1.62 | -2.17 | -2.70 | $-3.25$ | -4.34 | -5.41 | 1.00 |
| 30 | 0.000 | -0.014 | $-1.51$ | -2.26 | $-3.00$ | $-3.76$ | $-4.52$ | -6.02 | $-7.53$ | 1.48 |
| 100 | 0.000 | -0.0044 | $-2.01$ | $-3.01$ | -4.00 | -5.02 | $-6.03$ | -8.03 | $-10.04$ | 2.00 |
| 300 | 0.000 | -0.0017 | -2.48 | $-3.72$ | -4.95 | $-6.20$ | $-7.46$ | -9.92 | $-12.40$ | 2.48 |
| 1000 | 0.000 | -0.0004 | -3.00 | -4.50 | -6.00 | -7.50 | $-9.00$ | -12.00 | $-15.00$ | 3.00 |
| Inf | 0.000 | 0.000 | - Inf | $-\operatorname{Inf}$ | $-\operatorname{Inf}$ | - Inf | - Inf | - Inf | - Inf | Inf |

a Values of Table I converted into logarithmic form.


Figure 4.-The difference between values of $\log P$ in Table II and those calculated from a regression parabola used to correlate the data.
relationship between $\log P_{0, n}$ and $\log k / l$ is not, strictly speaking, parabolic, however, for all practical situations it may as well be.

At this point the question of branched and alternate pathways can be taken up again. When more thàn one route of travel is open to a drug molecule then the probability of a drug molecule reaching its target will be the sum of the probabilities of the various routes available (see eq 16). In eq 16 the left hand superscript of each

$$
\begin{equation*}
P_{0, n}=\left({ }^{1} P_{0, n}+{ }^{2} P_{0, n}+{ }^{3} P_{0, n} \ldots{ }^{j} P_{0, n}\right) \tag{16}
\end{equation*}
$$

probability term within the parenthesis designates one identifiable route. However, each ${ }^{j} P_{0 . n}$ term in eq 16 is a function of $\log k / l$ as illustrated by the family of curves in Figure 2. Each member of this family is
similar to a normal distribution curve with its maximum at $\log k / l=0.00$. The sum of such curves is still another curve of the same form, ${ }^{4}$ and hence the presence of branched and alternate routes to the receptor site does not alter the basic relationship of $P_{0, n}$ to $\log k / l$. The normal distribution curve therefore approximates this relationship for the probability considering all routes as well as an individual route.

In real situations there may be several independent routes involving the same number of interfaces, $n$. These routes of equal probability can be considered degenerate in analogy to quantum mechanics' degenerate states, i.e., states of equal probability. Thus, for molecules traveling in a set of degenerate routes, the probability of reaching a receptor is independent of the path taken. More important, however, routes involving the fewest interfaces are the routes of highest probability. Hence, these minimum pathways will bear the greatest traffic as far as drug molecules reaching the receptors are concerned. Obviously, the minimum pathway could be comprised of several degenerate routes, and this set of routes would be the primary access a drug molecule would have to the receptors. Only a negligible number of molecules will reach the receptors by routes having
(4) It would be difficult to prove this assertion mathematically, but it can be readily demonstrated to be true for the case at hand. Atypical regression equation is that for $n=2$ (see eq i).

```
log}\mp@subsup{P}{0,2}{}=-(0.261\pm0.013)(\operatorname{log}k/l\mp@subsup{)}{}{2}
                                    (0.002 士 0.0221) (logg k/l) - 0.804 (i)
```

Let us now consider the probabilities for three routes consisting of different values of $n$, say $n=2, n=4$, and $n=6$, By summing these prohabilities for different values of $\log k / l$, and taking the logarithms thereof, we can eventually arrive at the regression equation eq ii.
$\log \left(P_{0.2}+P_{0,4}+P_{0,6}\right)=-(0.271 \pm 0.016)(\log k /)^{2}-$

$$
(0.002 \pm 0.026)(\log k / l)-0.739 \quad \text { (ii) }
$$

Thus the data leading to eq ii fit a parabola only slightly less better than the data leading to eq $i$. In fact the difference is probably dne to the greater round-off error inherent in eq ii. Most interesting are the differences between the "observed" and calculated values given by hoth equations. The differences associated with eq ii show the same regular fonrth-order dependence on $\log k / l$ as those associated with eq $i$. These differences are observed for all values of $n$ (see Figure 4). The addition of higher valued terms in $n$ (e.g. $P_{0,8}$ ) to eq ii will have very little effect on the outcome since such terms make a negligible contribution to the overall probability relative to the lower ordered terms.
six or more interfaces above those of the minimum pathway.

Until now the time element has not been seriously considered in these deliberations. Since the main interest in work of this kind lies in the relative potencies of drugs as measured in a standard biological test, it is fair to limit our discussion to only those tests in which the observation period is the same in each instance. Thus, in the measurement of potency, time is considered to be held constant. In looking at the time variable one guickly realizes that the observation period is normally set by previous experience in dealing with drug- of the type being studied. So to begin with it is known that the test system allows sufficient time for at least some drug molecules to reach the receptors. Xaw if there are no membranes in the biological system. and a drug is administered as a concentrated dose in a small region at one end of the sestem, the time it takes for one of these added molecules to reach the other end is defusion controlled. and is a function of the molecules' average velocity. At a constant temperature, molocules of the same mass but of different chemical structures will have the same average velocitios (there may be exceptions of course) : hut even if one molecule has twice the mase of mother it will rake it onty 1.414 times as long to travel the same distance. Hence. other things being equal, the lighter molecules will have a slightly higher probability of reaching a particular distant point than the heavier ones in a given period.

However. these differences are negligible compared to those introduced when membranes and partition coeffieients are considered as is well known. The introduction of membranes to the sestem does not really alter a drug's late of travel, what it doe is to effeetively increase the distance it must cover. It may be repelled at an interface: in order to penetrate it, a molecule must wander abont random? matih, by chanes. it
strikes the interface again. The molecule many have to repeat this process many times. If it is not very suceessful, it will never reach the receptor to help effect the biological response; after all, there is a time limit.

At this point one may appreciate the above argument on equivalent most-probable router. Those molecules which travel by the router of greatest probability will ba owowhelmingly represented at the receptor. Ahmost all others are wasted. What is important in determining the final unterome is the particular pathway the drug molecule travels, :md the probability that : pare tienlar molecule penctrates an interface upon collision.

Suppose two molecule with different partition coofficients happen to follow exactly the same path. and are repelled at various interfaces exactly the same momber of times: each motecule arriver at the receptor region at approximately the same time (we make a small correction for the different masses). An interesting evont. but surely the molecule whose partition coefficient is least faromble has overome qreater odds to achere itstatas. In a particular tost system pactically alf molecules which reach the receptor travel by equivalent most-probable routes and arrive at approximately the same time. The problem therefore reduces to determining the probability that particular molecnles will follow the most favorable routes, and this of comse is done by the arguments leading to ea 15. Higher doses of drugs with unfavomble partition coefficients must be administered in order to offer more chances for enotgh molecules to get through to the reecptor in time to effect a response. The foregoing argument may not apply to all possible typer of drugs, but it is probably ratid for many common cases.

Acknowledgments.-Thanks are due to Dr. L. H. Connver of these labomtories for his conemagennmt in preparing this work for publication.

# Mass Spectral Analysis in the Identification of Human Metabolites of Warfarin ${ }^{1}$ 

<br> 

Recomed April 13, 16\%


#### Abstract

The mass spectra of warfarin [ 3 -( $\alpha$-acetonylbenzyl)-4-hydroxycommarin], a denterimm labeled analog, in hdroxylated analogs, and a recheed side cham analog ale discused and mechanistic rationalizations are given for the major fragmentation processes observed. These data together with nv, the, and synthetic data are nsed to identify and establish 6-and 7-hydroxywarfarin and the two diastereoisoners of 3 - $[\alpha$-( 2 -hydroxypropyl)benzyl|-4hydroxyconmarin as metabolites of warfarin in notmal man.


Warfarin [3-( $\alpha$-acetonylbenzyl)-4-hydroxycoumarin, 1a] is an oral anticongulant commonly employed in this country. Among the problems encountered in its clinical use in the occasionally inordinate difficulty in the maintenance of a stable degree of anticoagulation; as a

[^2]consequence hemorrhage can result from the same drug dose that previously produced acceptable hypoprothrombinemia. In addition the magnitude of the anticoagulant effect appears to be extremely sensitive to the influence of other drugs. These effects can usually be correlated with changes in the plasma clearance rate of the coumarin anticongulants. For example? several

[^3]
[^0]:    
    (2) J. T. Penniston, L. Beckelt. 1). L. Hentles, and C. Hanscta, 1fol. Phurmucol.. 5, 333 (1969). A number of other symbols have been proposed for the partitiun coefficient; Hansch for exanulle uses $P$. However, becarse the preseat discussion leans heavily on probability arguments. $P$ shall he reserved as a symbol for prolphility. The Hansel school bis recently. proposed $k ?$ as an alternate symlol, and it will be thed in the presebt disenssion.

[^1]:    
    

[^2]:    * To whon correspondence should he addressed.
    (1) This investigation was supported in part by The University of California Academic Senate Grant 10, San Francisco Division, and in part under a Grant-in-Aid of the American Heart Ascociation, supported thy be Alameda, San Francisco, San Mateo, und Santa Clara Heart Associations. The authors are grateful for the tednical aisistance of Mr. Ken chan and to Endo and thoot Laboatorie for why ying warfarin.

[^3]:    
     W/d, Aw. 207, 203: •1

