Inclusion of Molecular Flexibility in Partition Coefficient Calculations on Oligopeptides using Stochastic Sampling

Nigel G J Richards*, Philip B Williams, and Srinivasan Iyengar

Department of Chemistry, University of Florida, Gainesville, FL 32611, USA.

Fragment-based methods for computing water/1-octanol partition coefficients (log *P*), which employ two-dimensional structural representations, can often yield estimates of variable quality for flexible structures. We now report a novel approach for calculating log *P* which combines continuum solvation potentials and conformational search algorithms. In initial studies employing linear dipeptides, excellent agreement is obtained between calculated and experimental log *P* values using either stochastic sampling or systematic search methods to model the conformations adopted by these dipeptides in both solvent phases. Our results are consistent with proposals that the physical basis of certain correction terms used in empirical methods for log *P* estimation is due to neglect of the redistribution of conformer populations as compounds partition between aqueous and hydrophobic environments.

The transport properties and biological activity of a large number of drugs, pesticides and xenobiotics are often correlated with their molecular hydrophobicity [1]. This property is usually determined by measurement of the water/1-octanol partition coefficient, *P*, and is generally expressed as log *P* [2]. Although several techniques exist for the accurate measurement of this molecular parameter [3], calculation of log *P* values based upon only two- or three-dimensional structural representations has formed the focus of much research effort [4]. At present, the most widely used algorithm for calculating log *P* remains that implemented in the ClogP software package [5]. In this procedure, molecular structures are divided into a series of defined fragments for which parameters have been derived using a database of experimental log *P* measurements [6]. Summation of the relevant parameters together with the use of a series of correction terms then yields an estimate of the partition coefficient for the compound of interest [7], which is often in excellent quantitative agreement with the experimental value. Significantly, for certain classes of compound, even the ClogP algorithm may yield log *P* estimates which are in poor quantitative agreement with experiment. Such failures appear to be most common in the case of structures which can adopt a range of conformations in solution [7], and it is likely that determining the physical basis of these errors may allow additional insight into the molecular interactions underlying solute-solvent interactions [8]. In addition, since the ClogP algorithm is based solely on decomposition of two-dimensional molecular connectivity, it cannot easily predict partitioning differences between diastereoisomeric species or compounds which are simple structural isomers [9].

In the context of drug design, and given recent advances in screening large numbers of peptides for their ability to interact with cellular receptors and other proteins [10], understanding the conformational modulations undergone by linear oligopeptides as they partition from aqueous solution into membranes might also be important in the design of peptidomimetics with defined biological activity [11]. It is likely that amphiphilic, flexible oligopeptides adjust their conformational energies and populations when they move from water into hydrophobic, non-aqueous environments such as 1-octanol [12]. Therefore, as part of efforts to develop novel, conformationally constrained amino acids and peptidomimetics [13], we have begun to investigate alternative computational approaches to the estimation of log *P* for flexible peptides.

A critical problem in calculating log *P* for these amphiphilic molecules, which has usually been ignored in all but a few previous approaches [14], is that such compounds do probably adopt different conformational distributions in water and 1-octanol. Hence, calculations which are based on the analysis of a single molecular conformation are unlikely to reproduce the observed partitioning behavior [15]. Although the theory underlying our approach has been outlined elsewhere [16], the essential elements will be briefly summarized here. The partition coefficient, log *P* is calculated from the difference in two free energy terms as follows:

$$-2.303\ R\ T \log P = \Delta G_{oct} - \Delta G_{aq} \qquad (1)$$

where R is the ideal gas constant, T is the temperature, and ΔG_{aq} and ΔG_{oct} are the free energy changes moving a given compound from the gas-phase into water, and into 1-octanol, respectively. The first of these two free energies can be computed for a compound which adopts a

1060-3425/94 $3.00 © 1994 IEEE

Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences, 1994

series of conformations, using the following expression [17]:

$$\Delta G_{aq} = \Sigma\, w\,(i)\, \Delta G_{aq}\,(i] - \Sigma\, w''\,(i)\, \Delta E_g\,(i) \qquad (2)$$

where the summation is over the i conformations available to the molecule, ΔG_{aq} (i) is the free energy of the i-th conformation in aqueous solution, and ΔE_g (i) is the strain enthalpy of the i-th conformation in the gas-phase. In a similar manner, the energy of transfer from the gas-phase to 1-octanol can be calculated from:

$$\Delta G_{oct} = \Sigma\, w'\,(i)\, \Delta G_{oct}\,(i) - \Sigma\, w''\,(i)\, \Delta E_g\,(i) \qquad (3)$$

where ΔG_{oct} (i) is the free energy of the i-th conformation in 1-octanol. The weighting factors w (i), w' (i) and w'' (i) reflect the relative populations of the i-th conformation in water, 1-octanol and the gas-phase respectively. The inclusion of these functions ensures that any conformational reordering which occurs as the flexible compound partitions from one phase into another is explicitly included in this algorithm. The weighting terms can be easily calculated, as described previously [17], using the following equations:

$$w\,(k) = \exp[\, -\Delta G_{aq}\,(k) / RT\,] \,/\, \Sigma \exp[\, -\Delta G_{aq}\,(i) / RT]$$
$$w'\,(k) = \exp[\, -\Delta G_{oct}\,(k) / RT\,] \,/\, \Sigma \exp[\, -\Delta G_{oct}\,(i) / RT]$$
$$w''\,(k) = \exp[\, -\Delta E_g\,(k) / RT\,] \,/\, \Sigma \exp[\, -\Delta E_g\,(i) / RT\,]$$

Hence, the problem reduces to determining all of the allowed conformational minima for the structure of interest, and to the calculation of the solvation energy of each individual conformation. At first glance this combination of free energy and enthalpy terms appears problematic. However, for the relatively small dipeptide structures discussed in this paper, the contribution of vibrational entropy to the free energy of each conformation was almost identical and since only relative gas-phase energies were used in our approach, this term disappeared in the final calculation of the log *P* estimates. Whether this remains true for larger, more flexible, structures remains the subject of ongoing investigation. If not, then equations (2) and (3) should be modified to include vibrational entropy. On the other hand, the error inherent in ignoring such a term is likely to be negligible in comparison to the problems of determining all relevant conformational minima in our algorithm. Given the computational requirements of this conformational search [18], we also used a continuum solvation approach [19] which employs a set of empirical parameters to relate solvation energy and accessible molecular surface area [20]. In our algorithm, which has been described in detail elsewhere [16], we assume that molecules are composed of a number of substructures, or fragments [21], which possess an intrinsic, and transferrable, contribution to the total free energy change when a conformation is moved from one phase to another. Then, expressions for computing the solvation energies of the k-th conformation in water and 1-octanol can be written as:

$$\Delta G_{aq} = \Sigma\, A\,(m)\, f_s\,(m)\, \mathbf{F}_h\,(m) + \Delta E_g(k) \qquad (4)$$
$$\Delta G_{oct} = \Sigma\, A\,(m)\, f_s\,(m)\, [\mathbf{F}_h\,(m) + \mathbf{F}_p\,(m)] + \Delta E_g(k) \qquad (5)$$

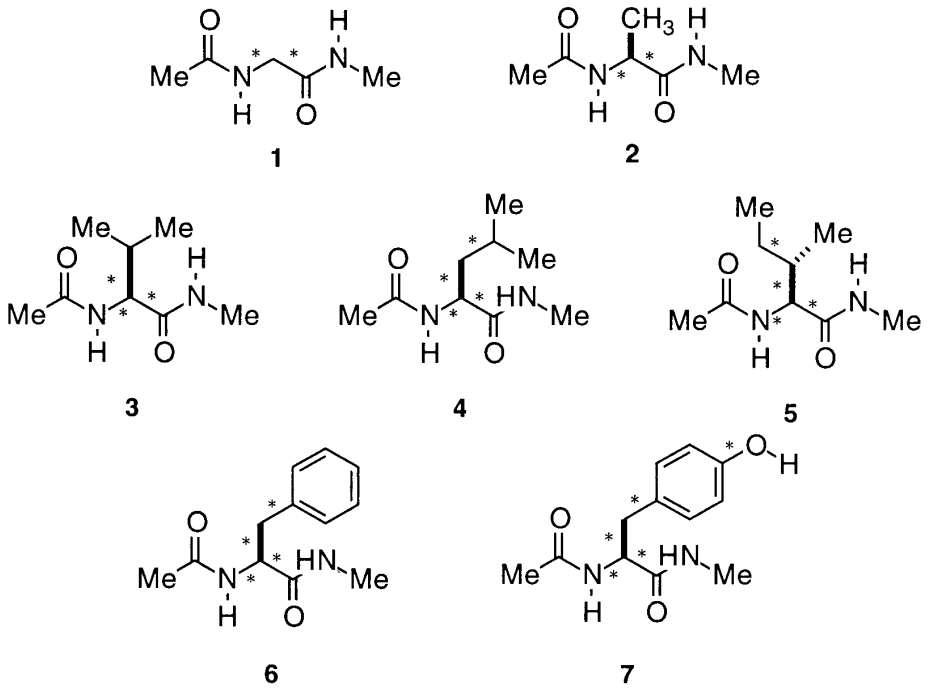
where the structure of interest can be decomposed into m substructures, A (m) is the solvent contactable area of the m-th fragment [16], and $\mathbf{F}_p$ (m) and $\mathbf{F}_h$ (m) are the free energy changes per unit area, associated with transfer of the m-th fragment from the gas-phase into 1-octanol, and from the gas-phase into water, respectively. The sets of parameters $\mathbf{F}_p$ (m) and $\mathbf{F}_h$ (m) are obtained by fitting experimental hydration energy and partitioning data for a limited number of compounds which can only adopt a single conformation [16]. These expressions also contain a 'step' function, f_s (m), which is set to unity for all non-polar fragments such as methyl or methylene groups, but which is used to modulate the solvation energies of fragments able to participate in hydrogen bonding interactions with the solvent. Hence, for an alcohol fragment, we do not assume a linear relationship between solvent accessible area and solvation energy. The physical justification for this function has been discussed elsewhere [16], but this type of term allows our solvation model to reproduce specific hydration effects which have been observed experimentally [22].

With the availability of high quality parameters $\mathbf{F}_p$ and $\mathbf{F}_h$, the problem of computing the partition coefficient for flexible, linear oligopeptides reduces to the determination of their conformational gas-phase energy minima [18]. In this paper, we report the results of partition coefficient calculations for a series of uncharged, flexible dipeptides, **1-7** (Figure 1) in which two methods of sampling the conformations available to these molecules are used in combination with our continuum solvation parameters. We show that the systematic search method, although yielding reasonable log *P* estimates, has a number of intrinsic problems which limit its utility. In addition, we also demonstrate that stochastic sampling to generate estimates of the average energy of these dipeptides in water and 1-octanol can be employed to obtain calculated log *P* values in similar agreement with experiment as those computed using ClogP without the need for inclusion of correction terms based on molecular connectivity.

Methods

A standard set of optimized starting structures for dipeptides **1-7** was generated as described previously [23]. A large number of initial conformations for each dipeptide was then obtained using the MULTICONFORMER systematic search algorithm [24]. In this approach, all

FIGURE 1: **Dipeptides for which log *P* estimates were calculated.** The central bonds of the dihedral angles that were allowed to vary in both the systematic search and stochastic sampling algorithms are marked *.



rotatable dihedral angles were systematically varied in 60° increments to generate new molecular conformations, structures being rejected if any internuclear distance between non-bonded atom pairs was less than 1.5 Å. All amide bonds were maintained in their *trans*-configuration ($\omega = 180^{o}$) in all starting structures. The resulting set of conformers for each dipeptide was then energy minimized using a modified MM2 force field [25], which includes an explicit hydrogen bonding potential, implemented in BATCHMIN V3.0 [26]. A block-diagonal Newton-Raphson algorithm was employed for the minimizations [27], and in all calculations the microscopic dielectric constant was fixed at 1.0. Structures were considered to be optimized when the total RMS gradient was less than 0.1 kJ/mol/Å. No energy cutoffs were employed to select only the lowest energy gas-phase minima as we anticipated that high energy conformations would be significantly stabilized by the solvation potential energy terms. Duplicate conformations were eliminated using standard superimposition methods in which all non-hydrogen atoms in a given structure were overlaid with the equivalent atoms in all of the previously determined, unique conformational minima by least-squares superimposition [28]. If the minimum RMS residuals for the two superimposed structures exceeded 0.25 Å/atom, then the new conformation was defined as a new minimum energy structure and included for use in subsequent partition coefficient calculations. The final set of dipeptide conformations was then used to estimate log *P* through the application of equations (1)-(3). Hence each dipeptide structure was divided into acceptable fragments using a substructure matching algorithm [26]. Solvation energy terms were then evaluated by determining the solvent accessible area for each fragment as previously outlined [16], and combining these areas with the relevant $\mathbf{F}_p$ and $\mathbf{F}_h$ parameters as required by equations (4) and (5). In general, individual fragments were defined so as to correspond to standard functional groups ensuring that electronic effects, such as conjugation, were an intrinsic to the fragment thereby increasing the potential transferrability of parameters. Fragment structures were also defined, as far as possible, so as to be cognate to those employed by ClogP in order to facilitate comparison of our $\mathbf{F}_p$ parameters with those derived using alternative parameterization schemes [6].

The standard set of optimized structures for dipeptides **1**-**7** was employed for the stochastic sampling studies. Two independent simulations were carried out for the water and 1-octanol phases, both using the same standard structure for each dipeptide in the first step. After this step, however, the conformations sampled were determined solely by the steric energies and solvation

parameters used in the simulation, i.e. each sampling trajectory was independent. For a given phase, the average energy of the dipeptide was calculated using a simple Metropolis algorithm [29], the total energy of the initial conformation, E_{old}, being computed using the modified MM2 force field [26] together with the solvation energy term calculated from the fragment areas and the appropriate $\mathbf{F}_h$ and/or $\mathbf{F}_p$ parameters [16]. Hence, for the simulation in water the solvation energy was computed using equation (4), while, in a similar fashion, equation (5) was employed to obtain the solvation term for the dipeptide in the 1-octanol phase. A new conformation was then generated by selecting two dihedral angles, at random, from the set of rotatable torsions and varying each of these angles randomly within an interval of [± 90°]. The total energy of the new structure, E_{new}, was then computed. If $E_{new} \leq E_{old}$, the new conformation was accepted, and its energy stored. On the other hand, if $E_{new} > E_{old}$ then the following expression was evaluated:

$$P_{ac} = \exp[(E_{old} - E_{new}) / RT] \quad (6)$$

where R is the ideal gas constant and T is the temperature. A random number, ζ, was then generated over the interval [0,1] and the conformation accepted only if $P_{ac} > \zeta$. If this conformation was accepted, then E_{new} was stored, for later use in calculating the average energy of the dipeptide, and a new conformation was generated in a similar manner. However, if the new conformation was rejected, then the previous structure was restored and another set of random dihedral variations carried out. In addition, E_{old} was stored again and used in obtaining the appropriate average energy.

The partition coefficient, log *P*, was then calculated from the expression:

$$-2.303\ R\ T \log P = [\Sigma E_{oct}(k) - \Sigma E_{aq}(k)] / N \quad (7)$$

where N is the number of steps in both simulations, E_{oct} (k) and E_{aq} (k) are the total energies of the k-th accepted conformation in the 1-octanol and aqueous phase respectively, R is the ideal gas constant and T is the temperature. Each summation is over the N conformational energies stored in each simulation. In general, since two independent simulations are carried out from the same seed structure, only E_{oct} (1) and E_{aq} (1) have identical values of ΔE_g. After every 100 steps had been completed for both the water and 1-octanol simulations, the value of the partition coefficient was determined using equation (7) and compared with that stored previously. When the two log *P* values were within 0.01, then the two simulations were terminated.

Results and Discussion

Choice of dipeptides Dipeptides **1-7** were selected as model compounds in these studies for several reasons. First, there has been extensive investigation of empirical potential functions for this class of compound and their ability to reproduce observed conformational properties has been widely demonstrated [30]. Second, given recent efforts to model the tertiary structure of proteins using only primary sequence information, methods for searching the conformation space of such compounds are well-proven [31]. More significantly, peptides represent structures which contain a mixture of hydrophilic and hydrophobic functional groups and therefore linear peptides probably exist in very different conformational distributions in water and 1-octanol [32], and should therefore represent structures for which most other methods of partition coefficient estimation fail. The well-defined fragmental nature of peptides was also important in determining the choice of peptides in these initial studies given that our solvation model was parameterized in terms of molecular fragments broadly corresponding to common functional groups [33]. A problem with the use of fragments concerned the transferrability of the sets of $\mathbf{F}_h$ and $\mathbf{F}_p$ parameters which might be poor if the chosen fragments are subject to a high degree of electronic polarization in structures of interest which were not initially part of the parameterization set. This condition was met by the linear peptides used in our studies in that all fragments were connected via either methine or methylene groups, minimizing the likelihood of large differences in intrafragment polarization in the peptide relative to that of the fragment in the original, organic compounds from which the solvation parameters $\mathbf{F}_h$ and $\mathbf{F}_p$ were derived. Uncharged linear peptides were also chosen in order to ensure that complications due to differential ionization in the water and 1-octanol phases could not affect our calculations. In addition, the use of uncharged compounds also reduced the possibility of conformational artifacts in the gas-phase due to charge-charge interactions of large magnitude. Finally, high quality experimental partition coefficient data was also available for structures **1-7** [34].

Solvation energy model Continuum models for estimating molecular solvation energies have formed the subject of intense investigation in recent years [19]. Given that obtaining an accurate representation of the conformational distributions of the peptides in the water and 1-octanol phases appeared to be the time-consuming aspect of our approach, we elected to employ such a model in order to compute the solvation energy of given peptide conformations in solution as rapidly as possible. However, our solvation model, which has been described in detail elsewhere [16], differed in two respects to those usually discussed in the literature [19]. First, we employed the surface area for each fragment which could actually contact solvent molecules [35] rather than the more commonly employed solvent-accessible surface [20].

TABLE 1: Calculated and observed log *P* values for the set of dipeptides 1-7 showing the performance of the two conformation-based methods and the ClogP algorithm.

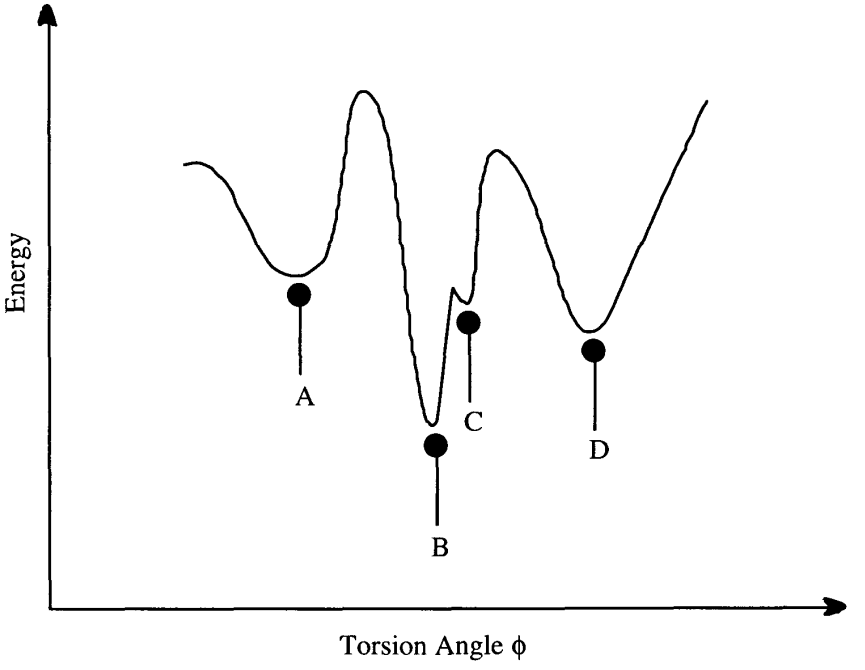
Dipeptide[a]		Partition Coefficent (log *P*)[b]				
		Observed[c]	Systematic Search		Stochastic Sampling	ClogP[d]
Ac-Gly-NMe	**1**	-1.56	-1.34	(12)[e]	-2.42	-1.65
Ac-Ala-NMe	**2**	-1.21	-1.15	(10)[e]	-1.34	-1.33
Ac-Val-NMe	**3**	-0.34	-0.03	(37)[e]	-0.28	-0.41
Ac-Leu-NMe	**4**	0.14	0.40	(133)[e]	0.24	0.12
Ac-Ile-NMe	**5**	0.13	0.18	(123)[e]	0.30	0.12
Ac-Phe-NMe	**6**	0.40	0.60	(79)[e]	0.31	0.23
Ac-Tyr-NMe	**7**	-0.32	-0.88	(84)[e]	-0.23	-0.59
RMS Deviation			0.38		0.34 (0.11)[f]	0.14

(a) Amino acids are abbreviated using standard three letter codes. Structures are as in Figure 1.
(b) Negative values indicate that the dipeptide prefers to partition into the aqueous phase.
(c) All experimental values are taken from [34].
(d) Partition coefficient values computed using ClogP Version 3.54 (1989 release).
(e) Total number of minimum energy conformations used in the log *P* calculation.
(f) RMS deviation of calculated and observed log *P* values. The value in parentheses corresponds to the RMS omitting the result for dipeptide **1**.

Second, the fragment solvation energy for polar fragments was modulated by an additional 'step function (f_S (m) in equations (4) and (5)). While the use of such a function has been justified elsewhere, the role of f_S (m) in our model is to allow the introduction of solvation effects which are due to specific interactions such as hydrogen bonding. Thus, a recent multivariate analysis of the partitioning properties of a series of compounds between water and a range of immiscible solvents, indicated that 95% of the variation in measured log *P* values could be correlated with two structural factors [36]. The first of these was related to the surface area of the cavity introduced into the solvent by the presence of the solute, while the second was directly associated with the hydrogen bonding properties of the solvent. However, the strength of hydrogen bonds between water and proton acceptors such as pyridine nitrogens and ether oxygens has been shown to be relatively insensitive to the solvent accessible area of the proton accepting atom [22], which is inconsistent with the usual assumptions of a linear dependence of solvation energy upon fragment area. More recent studies using linear free energy relationships in partition coefficient calculation have also supported the central importance of specific hydrogen bonding interactions in partitioning [37]. In the case of non-polar fragments such as methine, methylene and methyl, f_S (m) was set equal to unity.

Partition coefficient calculations Initial calculations employed a systematic search procedure with which to generate the gas-phase conformational minima associated with each of the dipeptides **1-7**. For each dipeptide, an initial structure was obtained in which all bond lengths and bond angles were at their equilibrium values and this was used to generate a set of conformations by systematic variation of all of the rotatable torsion angles within the molecule [24]. Each of these was then energy minimized [27] to yield the set of unique gas-phase minima for each dipeptide. Duplicate conformations were eliminated using only superimposition algorithms since it was likely that high-energy gas-phase minima, in which there were usually no internal hydrogen bonds, would possess high solvation energies especially in the aqueous phase. The energy of each of these conformations was then evaluated in water and in 1-octanol using equations (4) and (5) in combination with the solvation parameters. After computing the relative populations in each phase, the energy changes associated with transferring each conformational distribution from the gas-phase into water (ΔG_{aq}) and from the gas-phase into 1-octanol (ΔG_{oct}) were determined using equations (2) and (3). Partition coefficient estimates were then obtained for each dipeptide (Table 1) [38] and compared with those measured experimentally [34]. All of the computed log *P* values were in excellent agreement with experiment. However, a number of potential problems are associated with the use of this systematic search procedure for its general application to calculating partition coefficients of flexible peptides. First, the set of initial conformations used in energy minimization to locate the potential energy minima becomes prohibitively large for compounds

FIGURE 2: **Hypothetical one-dimensional potential energy surface for a flexible molecule such as a peptide.**



possessing more than six rotatable torsion angles [39]. More significantly, the number of conformations which converge to the same energy minimum also increases dramatically, limiting the efficiency of the search. The partition coefficient computation is also very sensitive to the correct elimination of duplicate conformations [40], since counting any minimum twice, especially if it represents a low energy structure in a given phase, will yield an incorrect solvation energy. Therefore, all energy minimizations were highly converged, as judged by the RMS gradient of the potential energy function, so as to ensure that only one structure was used for each potential energy minimum for the dipeptide. Duplicate structures were eliminated by superimposition of structures using all of the non-hydrogens atoms in the molecule, conformations being retained which differed such that the RMS residuals for the two structures exceeded 0.25 Å/atom [26]. Secondly, we were concerned that systematic search methods might fail due to the weighting procedure employed to obtain the final estimate of log *P*. All conformational minima are represented by a single structure in systematic search procedures (Figure 2). However, our weighting scheme ensures that the global minimum energy structure (point B in Figure 2) makes the greatest contribution to the final calculated transfer energy. However, this assumption may not necessarily be correct if the global minimum is deep and narrow, and a higher energy minimum exists which is wide and shallow (point A in Figure 2). In this case, the flexible molecule will presumably sample conformations associated with the higher energy well more often, making their contribution to the overall molecular distribution more important.

We therefore decided to explore the use of stochastic sampling approaches in order to evaluate the average energy of a given dipeptide in both water and in 1-octanol [41]. Although, in principle, molecular dynamics simulations [42] could have been used to sample the important conformational minima for each dipeptide in the two phases, stochastic sampling methods offered two advantages. First, we wanted to ensure that large conformational changes could occur at each simulation step in order that a wide sampling of conformation space could be achieved in a small number of steps. For flexible compounds, molecular dynamics generally requires high temperatures before minima distant from the starting structure in conformation space are located, limiting the rate of convergence in many types of free energy calculation [43]. An excellent discussion of these problems, together with some insights into their solution, has recently been presented as part of computational studies upon the estimation of partition

coefficients for small organic molecules between water and 1-propanol [44]. Second, since our solvation potential employed the solvent-contact area, the solvation energy function was discontinuous and therefore obtaining the derivatives of the function necessary to computing the atomic forces used to determine the molecular dynamics trajectory was likely to fail. Given that log *P* values are measured at constant pressure, any simulations ought to employ an NPT ensemble rather than the NVT ensemble used to sample molecular conformations [45]. In contrast to our acceptance criterion, which was based on the total energy of the conformation, NPT simulations require the inclusion of a volume term such that the acceptance criterion becomes:

$$P_{ac} = \exp[\ \{(\ E_{old} - E_{new}\ ) + \Delta V\} / RT\ ]$$

where $E_{old} \leq E_{new}$ and ΔV is related to the change in volume of the system after the torsional variations to generate the new dipeptide conformation. ΔV is given by the expression [45]:

$$\Delta V = P\ (\ V_{new} - V_{old}\ ) - (1/RT)\ \log_e[\ V_{new} / V_{old}\ ]$$

where P is the pressure of the system, V_{old} and V_{new} are the volumes of the systems comprising the initial and new dipeptide conformations respectively, R is the ideal gas constant and T is the temperature of the simulation. However, given the use of a continuum solvation model in these calculations, it was necessary to assume that the volume changes as the dipeptides adopted new conformations would be small, allowing us to ignore the ΔV term in our stochastic sampling simulation.

Although stochastic sampling was selected to generate the conformational distributions in the two phases because of its ability to sample large regions of conformation space in relatively few steps, the rate of convergence of these types of calculations is limited by the ratio of structures accepted to the total number of steps. Therefore the conformational variation that can be introduced into any starting conformation so as to generate a new structure is limited by the acceptance criterion. Large torsional variations, especially in complex structures, can therefore yield conformations in which there are a number of high-energy nonbonded interactions which result in rejection of the new conformation. Hence, Monte Carlo simulations of proteins often exhibit poor rates of convergence compared to molecular dynamics methods [46]. The use of torsional variation to generate linear peptide conformations, on the other hand, appears well behaved [47]. For example, this approach has been used to generate low energy conformations for Met^5-enkephalin [18b]. In that study, torsion angles were allowed to vary randomly within a range of $\pm 180^o$ given rise to new conformations which were then energy minimized. Similar approaches have been used to generate ensembles of low energy minima for a variety of small organic compounds [48]. Although many groups state that the simulation parameters should be adjusted such that acceptance ratio in any stochastic sampling calculation is approximately 50% [41a], this target appeared purely arbitrary upon examination of the literature. Indeed, acceptance ratios of only 10% have been reported to yield the maximum rate of convergence of certain molecular properties [49]. In order to simulate partitioning at 298 K, where the probability of accepting high energy structures is small, the ability to tunnel through energy barriers is vital if sufficient conformation space is to be sampled in the simulation. A wide torsional window for each angle was therefore selected of $\pm 90^o$. In addition to the size of the torsional window, we also decided to explore the optimum number of torsion angles that should be varied in each simulation step before calculation of the energy of the new conformation. Previous studies have also examined this effect, concluding that variation of a small number of torsion angles tends to give a large proportion of duplicate structures after energy minimization. On the other hand, variation of large numbers of torsion angles lead to a reduction in the number of conformations accepted. In our studies, which were aimed at obtaining converged log *P* estimates in the smallest number of steps and in which large torsional ranges were desired, we determined that variation of two rotatable bonds per simulation step was the optimum choice.

The physical correctness of our stochastic sampling algorithm was tested in two ways. First, the partition coefficient was computed for dipeptide **2** at two different temperatures to ensure that log *P* converged to two different values. Then, in order to determine whether conformation space was being evenly sampled, two simulations of **2** were carried out from the identical initial structure but using different initial seed values for the torsional variation. In the latter study, both simulations converged to an identical value for log *P*. Having established that the approach was behaving reasonably, we proceeded to calculate log *P* estimates for dipeptides **1-7** (Table 1). With the exception of the glycine derivative **1**, all of the computed values were in excellent agreement with experimental values [34]. A number of reasons may explain the poor result for glycine. First, this was the most conformationally flexible dipeptide studied. Therefore, it is possible that conformational space had not been sufficiently sampled to ensure convergence to an accurate log *P* value. However, attempts to increase the number of steps in the simulation to obtain log *P* estimates closer to the experimental value failed to yield significantly different converged values of the partition coefficient. It is possible that torsional variation without modification of bond lengths and/or bond angles might give rise to an inadequate sampling of all relevant conformations in both water and 1-octanol [50], and this would not be expected to affect the rate of convergence of

the overall simulation. Alternate stochastic sampling algorithms are therefore under investigation in whih several types of internal coordinate are varied. Second, any errors in the fragment values used in these studies would have more effect upon the glycine solvation energies given the small number of unique fragments that are present in this structure. The final possibility is that the measured log *P* value for dipeptide **1** is incorrect. Attempts to remeasure this quantity are therefore underway. There is no doubt, however, that estimation of log *P* using the stochastic sampling algorithm worked well in the case of dipeptides **2-7**, confirming that the use of conformation-based methods eliminates the requirement for arbitrary correction terms. In addition, our optimized simulation parameters in which two torsion angles were allowed to vary randomly within $\pm 90^{o}$ of their initial value also gave rise to excellent simulation convergence. Hence, all simulations for dipeptides **2-7** converged within 100 000 steps. Even the highly flexible glycine analog **1** only required 210 000 steps to yield the converged value. We also note that in other studies [51] employing a solvation model which neglected an specific hydrogen bonding interactions, i.e. in which the solvation energy was linearly proportional to area for all fragments, the rates of convergence of the log *P* estimates were extremely slow. For example, in the case of peptide **4**, convergence was not achieved after 400 000 steps.

Conclusions

With the exception of dipeptide **1** we have demonstrated that stochastic sampling in combination with empirical solvation potentials can yield excellent estimates of log *P* for flexible, amphiphilic structures which can adopt very different conformational distributions in water and 1-octanol. Our results also appear to confirm that the observations that ClogP fails to yield accurate log *P* estimates for flexible molecules are due to the neglect of conformational effects [7]. Although many groups have argued that atom-based parameterization schemes are superior [4d], we conclude that fragment-based parameters are viable, and transferrable, provided that the conformational behavior of the structures of interest is included into the calculations. There is no doubt that the computational demands of the approach outlined here far exceed those of ClogP, which clearly remains the best approach to estimating log *P* for many classes of molecule. For example, all of the calculations reported here took at least 48 hours of CPU time on a Silicon Graphics workstation while ClogP yielded log *P* estimates in only a few minutes. On the other hand, our algorithm has several properties that might contribute to its usefulness in rational methods of drug discovery [52].

Two extensions to this work are under development. First, since these calculations employ three-dimensional structural representations, partitioning differences between diasteroisomers should be predicted by this approach [53]. As yet, despite receiving some interest, few methods have been able to treat this problem with any great success [14, 8b]. Second, cyclic peptides remain of great interest in structure-function studies aimed at elucidating the bioactive conformations of naturally-occurring linear peptides [54]. Until recently, there was no clear approach to sampling ring conformations by torsional variation which did not introduce severe bond length or bond angle distortions into the cyclic structure. However, we note that a recent algorithm for sampling ring conformations by dihedral variation has been studied and has proven suitable for use in Metropolis Monte Carlo simulations [55]. A modification of this algorithm has been implemented by our group and its application to the calculation of the partition coefficient for cyclic peptides is under active investigation. Our results will be reported in due course.

Acknowledgements

We thank Dr Michael S. Tute (Pfizer Central Research, U. K.) and Professor H. Weinstein (Mount Sinai School of Medicine) for useful, and stimulating, discussions. We gratefully acknowledge the provision of a CASE studentship from the Science and Engineering Research Council (U.K.) and Pfizer Central Research, U.K. Additional computer resources required for this work were also made available by Pfizer Central Research, U.K.

References

1. (a) Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525. (b) Hansch, C.; Bjorkroth, J. P.; Leo, A. J. *J. Pharm. Sci.* **1987**, *76*, 663. (c) Leo, A. J. *J. C. S. Perkin Trans. 2* **1983**, 825. (d) Testa, B.; Jenner, P.; Kilpatrick, G.; El Tayar, N.; van der Waterbeend, H.; Mardsen, C. D. *Biochem. Pharmacol.* **1987**, *36*, 4041.

2. (a) Tute, M. S. *Adv. Drug. Res.* **1971**, *6*, 1. (b) *Theoretical Drug Design Methods*; Franke, R., Ed.; Elsevier Science Pub.: Amsterdam, 1984. (c) Martin, Y. C. *Quantitative Drug Design*; Marcel Dekker: New York, 1978.

3. (a) *Partition Coefficient: Determination and Estimation*; Dunn, W. J., III, Block, J. H., Pearlman, R. S., Eds.; Pergamon Press: Oxford, 1986. (b) El Tayar, N.; Tsai, R.-S.; Carrupt, P.-A.; Testa, B. *J. C. S. Perkin Trans. 2* **1992**, 79. (c) Mirlees, M.; Moulton, S.; Murphy, C.; Taylor, P. *J. Med. Chem.* **1976**, *19*, 615. (d) Harnisch, M.; Mockel, H.; Schulze, G. *J. Chromatogr.* **1983**, *282*, 315.

4. (a) Leo, A. J. *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P.G., Taylor, J. B., Ramsden, C. A., Eds.; Pergamon Press: Oxford, 1990; Vol. 4, p. 315. (b) Leo, A. J. *Meth. Enzymol.* **1991**, *202*, 544. (c) Suzuki, T. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 149. (d) Bodor, N.; Gabanyi, Z.; Wong, C. *J. Am. Chem. Soc.* **1989**, *111*, 3783. (e) Ghose, A. K.; Pritchett, A.; Crippen, G. M. *J. Comput. Chem.* **1988**, *9*, 80. (f) Klopman, G.; Namboodiri, K.; Schochet, M. *J. Comput. Chem.* **1985**, *6*, 28. (g) Takasuka, M.; Nakai, H.; Shiro, M. *J. C. S. Perkin Trans. 2* **1982**, 1061. (h) Klopman, G.; Wang, S. *J. Comput. Chem.* **1991**, *12*, 1025.

5. Leo, A.; Weininger, D.; Weininger, A.; Pomona College MedChem Project: Claremont, CA 91711.

6. (a) Hansch, C.; Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley Interscience: New York, 1979. (b) Leo, A. J.; Dow, P. Y.; Silipo, C.; Hansch, C. *J. Med. Chem.* **1975**, *18*, 865.

7. Leo, A. *Chem. Rev.* **1993**, *93*, 1281-1306.

8. (a) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431. (b) Jorgensen, W. L.; Ravimohan, C. J. *J. Chem. Phys.* **1985**, *83*, 3050. (c) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184.

9. (a) Pleiss, M. A.; Grunewald, G. L. *J. Med. Chem.* **1983**, *26*, 1760. (b) Bodor, N.; Huang, M.-J. *J. Comput. Chem.* **1991**, *12*, 1182-1186. (c) Williams, P. B.; Hinds, M. G.; Bowen, D. V.; Tute, M. S.; Richards, N. G. J. *Unpublished observations*.

10. (a) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. *Science* **1991**, *251*, 767. (b) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature(Lond.)* **1991**, *351*, 82. (c) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature(Lond.)* **1991**, *351*, 84.

11. (a) Rizo, J.; Gierasch, L. *Annu. Rev. Biochem.* **1992**, *61*, 387. (b) Morgan, B. A.; Gainor, J. A. *Annu. Rep. Med. Chem.* **1989**, *24*, 243.

12. (a) Sargent, D. F.; Schwyzer, R. *Proc. natl. Acad. Sci., U. S. A.* **1986**, *83*, 5774. (b) Schwyzer, R.; Erne, D.; Rolka, K. *Helv. Chim. Acta* **1986**, *69*, 1789. (c)

13. (a) Hinds, M. G.; Welsh, J. H.; Brennand, D. M.; Fisher, J.; Glennie, M. J.; Turner, D. L.; Richards, N. G. J.; Robinson, J. A. *J. Med. Chem.* **1991**, *34*, 1777. (b) Hinds, M. G.; Richards, N. G. J.; Robinson, J. A. *J. C. S. Chem. Commun.* **1988**, 1447. (c) Parr, I. B.; Boelein, S. K.; Schuster, S. M.; Richards, N. G. J. *Submitted for publication.*

14. (a) Kantola, A.; Villar, H. O.; Loew, G. H. *J. Comput. Chem.* **1991**, *12*, 681. (b) Alkorta, I.; Villar, H. O. *Int. J. Quant. Chem.* **1992**, *44*, 203.

15. (a) Camilleri, P.; Watts, S. A.; Boraston, J. A. *J. C. S. Perkin Trans. 2* **1988**, 1699. (b) Hopfinger, A.; Battershell, R. *J. Med. Chem.* **1976**, *19*, 569.

16. Richards, N. G. J.; Williams, P. B.; Tute, M. S. *Int. J. Quantum. Chem., Quantum. Biol. Symp.* **1991**, *18*, 299.

17. (a) Kang, Y. K.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4109. (b) Kang, Y. K.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4118.

18. (a) Howard, A. E.; Kollman, P. A. *J. Med. Chem.* **1988**, *31*, 1669. (b) Li, Z.; Scheraga, H. A. *Proc. natl. Acad. Sci., U. S. A.* **1987**, *84*, 6611. (c) Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1987**, *8*, 826.

19. (a) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. F. *J. Am. Chem. Soc.* **1990**, *112*, 6127. (b) Ooi, T.; Ootabake, M.; Némethy, G.; Scheraga, H. A. *Proc. natl. Acad. Sci., U. S. A.* **1987**, *84*, 3086. (c) Eisenberg, D.; McLachlan, A. D. *Nature(Lond.)* **1986**, *319*, 199. (d) Jean-Charles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 1454. (e) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305. (f) Wong, M. W.; Wiberg, K. B.; Frisch, M. J. *J. Am. Chem. Soc.* **1992**, *114*, 1645.

20. (a) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151. (b) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379. (c) Richmond, T. J. *J. Mol. Biol.* **1984**, *178*, 63.

21. (a) Rekker, R.; Mannhold, R. *Calculation of Drug Lipophilicity*; VCH: Weinheim, 1992. (b) Rekker, R. *The Hydrophobic Fragmental Constant*; Elsevier Scientific Pub: Amsterdam, 1977.

22. Taylor, P. *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P.G., Taylor, J. B., Ramsden, C. A., Eds.; Pergamon Press: Oxford, 1990; Vol. 4, p. 241.

23. Richards, N. G. J.; Williams, P. B.; Tute, M. S. *Int. J. Quantum Chem.* **1992**, *44*, 219.

24. Lipton, M. A.; Still, W. C. *J. Comput. Chem.* **1988**, *9*, 343.

25. Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.

26. Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R. M. J.; Lipton, M. A.; Caufield, C. E.; Chang, G.; Hendrickson, T. F.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.

27. (a) White, D. N. J. *Comput. Chem.* **1977**, *1*, 225. (b) Burkert, U.; Allinger, N. L. *Molecular Mechanics. ACS Monograph 177*; American Chemical Society: Washington, 1982.

28. (a) Kabsch, W. *Acta Crystallogr. Sect. A.* **1978**, *A34*, 827. (b) Kabsch, W. *Acta Crystallogr. Sect. A.* **1976**, *A32*, 922.

29. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

30. Hall, D.; Pavitt, N. *J. Comput. Chem.* **1984**, *5*, 441.

31. (a) Moult, J.; James, M. N. G. *Proteins, Struct. Funct. Genet.* **1986**, *1*, 146. (b) White, D. N. J.; Kitson, D. H. *J. Mol. Graph.* **1986**, *4*, 112.

32. (a) Grathwohl, C.; Wüthrich, K. *Biopolymers* **1976**, *15*, 2043. (b) Dungan, J. M.; Hooker, T. M. *Macromolecules* **1981**, *14*, 1812.

33. Abraham, D. J.; Leo, A. J. *Proteins, Struct. Funct. Genet.* **1987**, 2, 130.

34. (a) Wynne, W. J.; van Buuren, H. J.; Wakelkamp, W. *Experientia* **1982**, *38*, 655. (b) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664.

35. De Bruijn, J.; Hermens, J. *Quant. Struct.-Act. Relat.* **1990**, *9*, 11.

36.(a) Koehler, M.; Grigoras, S.; Dunn, W. J., III. *Quant. Struct.-Act. Relat.* **1988**, *7*, 150. (b) Dunn, W. J., III.; Koehler, M. G.; Grigoras, S. *J. Med. Chem.* **1987**, *30*, 113.

37. Leahy, D. E.; Morris, J. J.; Taylor, P. J.; Wait, A. R. *J. C. S. Perkin Trans. 2* **1992**, 705.

38. Some values reported in this paper are different compared to those reported in a preliminary description of this work [23]. The current values have been obtained after a recent re-examination of the step values associated with specific polar fragments, and the correction of the fragment parameter employed for hydroxyl groups attached to aromatic rings.

39. Marshall, G. R. *Computer-Aided Molecular Design*; Richards, W. G., Ed.; IBC: London, 1989.

40. (a) Burt, C.; Richards, W. G. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 231. (b) Burt, C.; Richards, W. G.; Huxley, P. *J. Comput. Chem.* **1991**, *12*, 1139. (c) Bowen-Jenkins, P. E.; Cooper, D. L.; Richards, W. G. *J. Phys. Chem.* **1985**, *89*, 2195.

41. (a) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987. (b) Dunn, W. J., III.; Nagy, P. I. *J. Comput. Chem.* **1992**, *13*, 468. (c) Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M. A.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419. (d) Jorgensen, W. J.; Briggs, J. M.; Contreras, M. L. *J. Phys. Chem.* **1990**, *94*, 1683.

42. (a) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992. (b) Brooks, C. L., III.; Pettit, B. M.; Karplus, M. *Adv. Chem. Phys.* **1988**, *71*, 1. (c) Karplus, M.; Petsko, G. A. *Nature(Lond.)* **1990**, *347*, 631.

43. (a) Mitchell, M. J.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 271. (b) Mazor, M.; Pettit, B. M. *Mol. Simulations* **1991**, *6*, 1. (c) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532. (d) Mark, A. E.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Chem. Phys.* **1991**, *94*, 3808.

44. Essex, J. W.; Reynolds, C. A.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 3634.

45. McDonald, I. R.; Singer, K. *Mol. Phys.* **1972**, *23*, 29.

46. (a) Northrup, S. H.; McCammon, J. A. *Biopolymers* **1980**, *19*, 1001. (b) Rossky, R. J.; Doll, J. D.; Friedman, L. H. *J. Chem. Phys.* **1978**, *69*, 4628.

47. Noguti, T.; Go, N. *Biopolymers* **1985**, *24*, 527.

48. Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379.

49. (a) Valleau, J. P., Whittington, S. G. *Statistical Mechanics A. Modern Theoretical Chemistry*; Berne, B., Ed.; Plenum Press: New York, 1977; Vol. 5, p. 137. (b) Wood, W. W.; Jacobson, J. D. *J. Chem. Phys.* **1957**, *27*, 1207.

50. Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, M. J.; Naider, F. *J. Am. Chem. Soc.* **1979**, *101*, 6842.

51. Williams, P. B. *Ph. D. Thesis* **1991**, University of Southampton, U. K.

52. (a) Böhm, H.-J. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61. (b) Cohen, N. C.; Blaney, J. M.; Humblet, C.; Gund, P.; Barry, D. C. *J. Med. Chem.* **1990**, *33*, 883. (c) DesJarlais, R. J.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. *J. Med. Chem.* **1988**, *31*, 722. (d) Goodford, P. *J. Med. Chem.* **1985**, *28*, 849.

53. Preliminary calculations for the pair of diastereoisomeric tripeptides Ac-**S**-Pro-**S**-Tyr-NMe and Ac-**R**-Pro-**S**-Tyr-NMe have indicated that this conformation-based approach can reproduce the observed partitioning differences for these two peptides.

54. (a) Taub, R.; Greene, M. I. *Biochemistry* **1992**, *31*, 7431. (b) Wüthrich, K.; Freyberg, B. V.; Weber, C.; Wider, G.; Traber, R.; Widmer, H.; Braun, W. *Science* **1991**, *254*, 953. (c) Altschuh, D.; Vix, O.; Rees, B.; Thierry, J.-C. *Science* **1991**, *256*, 92.

55. Guarneri, F.; Wilson, S. R. *Tetrahedron* **1992**, *48*, 4271.