of $\mathrm{BH}_{2}$ are deduced to be -11.2 and $4.1 \mathrm{kcal} \mathrm{mol}^{-1}$, respectively. However, based on the relative energy of $5 \mathrm{e}_{\perp}$, the inductive apicophilicity is indicated to be only $-5.8 \mathrm{kcal} \mathrm{mol}^{-1}$. This discrepancy is probably due to the only partial optimization of structure $5 \mathbf{e}_{\perp}$ (the $\mathrm{H}_{\mathrm{e}} \mathrm{PB}$ angle was fixed at $90^{\circ}$ ) which raises its energy.

## Conclusions

The relative energies, calculated at MP4SDTQ/6-31G* + ZPE, for both TP and SP conformations reveal inherent substitution effects. Some of the energy differences are quite close to empirical apicophilicity values or experimental energy differences of more highly substituted compounds. The SP structures for $\mathrm{PH}_{4} \mathrm{X}$ (X $=\mathrm{Li}, \mathrm{Na}, \mathrm{BeH}$, and MgH ) are the most stable conformations. The apicophilicities of $\mathrm{CH}_{3}, \mathrm{OH}$, and SH are small. The preferential $\mathrm{Y}-\mathrm{H}$ orientation in equatorial $\mathrm{PH}_{4} \mathrm{OH}$ and $\mathrm{PH}_{4} \mathrm{SH}$ is orthogonal to the equatorial plane. This preference is an important factor in determining the conformations of cyclic phosphorane systems. Our apicophilicity scale (in kcal mol ${ }^{-1}$ ) is $\mathrm{OH}(0.4)>$
$\mathrm{SH}(-0.1)>\mathrm{CH}_{3}(-0.9)>\mathrm{PH}_{2}(-3.3)>\mathrm{NH}_{2}(-7.2)>\mathrm{SiH}_{3}$ (-8.6). $\mathrm{PH}_{4} \mathrm{~F}$ and $\mathrm{PH}_{4} \mathrm{Cl}$ are not suitable for apicophilicity evaluation because of their ionic character.

Substituent effects on pentacoordinated phosphoranes can be divided into inductive (or $\sigma$ ) and $\pi$-bonding effects. Inductive effects dominate in the SP and apically substituted TP conformations, and the relative energies correlate with the electronegativities. $\pi$-Interactions play an important role in the equatorially substituted TP conformations of $\mathrm{NH}_{2}$ and OH . After dissection and delection of these $\pi$-bonding effects, the inductive contribution remaining also correlates with electronegativity.

Acknowledgment. This work was supported by NATO and by USPH NIH Grant GM30369, the Deutsche Forschungsgemeinschaft, Fonds der Chemischen Industrie, and the Convex Computer Corporation. Y. Zhang was supported by an Alexander von Humboldt Foundation fellowship. We thank H. Xie for technical assistance.

# Assessing Molecular Similarity from Results of ab Initio Electronic Structure Calculations 

Jerzy Cioslowski* and Eugene D. Fleischmann<br>Contribution from the Department of Chemistry and Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida 32306-3006. Received March 2, 1990


#### Abstract

A new molecular similarity index, called the number of overlapping electrons (NOEL), is proposed. This similarity index can be computed very rapidly from the natural orbitals and their occupation numbers of the molecules under comparison. The low computational cost makes it possible to optimize the mutual orientation of molecules by maximizing NOEL. The magnitude of NOEL is related to the number of electrons in the molecular fragment common to both molecules. The new approach is illustrated on the examples of benzene, aniline, nitrobenzene, and 4-nitroaniline molecules and the acetate, isoxazole 3 -oxide and isoxazole 5 -oxide anions.


## Introduction

In its early stages, chemistry has been mainly the science of comparing and classifying molecules and chemical reactions. Only in the last 50 years has it become possible to rationalize the observed properties of molecules by using the first principles of quantum mechanics. The properties that have rigorous theoretical definitions, such as energies and multipole moments, are easily amenable to theoretical approaches. On the other hand, properties of more intuitive nature, such as reactivity, aromaticity, or similarity, are more difficult to quantify.

Pharmacologists, toxicologists, and medicinal chemists find it particularly convenient to discuss and classify the physiological action of molecules by using instinctive measures of molecular similarity. For example, the taste of various substances is usually believed to be related to their molecular shape. ${ }^{1}$ Organic chemists constantly use the concept of molecular similarity by invoking the notions of functional groups and synthons. Yet. the attempts to rigorously define the molecular similarity are quite scarce in the chemical literature. One of the obvious reasons is that certain arbitrary assumptions have to be necessarily made in order to judge the similarity of molecules from the results of quantum-mechanical calculations. The first assumption is that of molecular rigidity because calculations involve the Born-Oppenheimer approximation, which freezes the positions of the nuclei. The second assumption is of a more philosophical nature. One has to postulate

[^0]that similar molecules have similar electron distributions. Third, one has to adopt a particular form of the functional which yields the numerical magnitude of the molecular similarity. As in the case of the definition of atomic charges, a plethora of choices are obviously possible; we insist, however, that any acceptable measure of the molecular similarity should conform to the following rules: (1) It should be derivable from the wave functions of the molecules in question alone. It should not depend upon either explicit or implicit assumptions about the basis sets used in calculations or the level of theory employed. (2) It should have some clearly recognizable physical or mathematical interpretation. (3) It should be computationally feasible. In particular, it should allow for optimization of the mutual orientation of the molecules in question by means of maximization of the similarity measure.

Although substantial progress have been recently achieved in quantification of the three-dimensional shape and similarity of molecules, ${ }^{2}$ the only reported practical calculations involve a similarity index based on the electron density that has been put forward by Carbo, Leyda, and Arnau. ${ }^{3}$ If the molecules under comparison, A and B , have the electron densities $\rho_{\mathrm{A}}(\overrightarrow{\mathrm{r}})$ and $\rho_{\mathrm{B}}(\overrightarrow{\mathrm{r}})$, respectively, then the similarity index reads

$$
\begin{equation*}
R_{\mathrm{AB}}=\int \rho_{\mathrm{A}}(\overrightarrow{\mathrm{r}}) \rho_{\mathrm{B}}(\overrightarrow{\mathrm{r}}) \mathrm{d} \overrightarrow{\mathrm{r}} /\left[\int \rho_{\mathrm{B}}^{2}(\overrightarrow{\mathrm{r}}) \mathrm{d} \overrightarrow{\mathrm{r}} \int \rho_{\mathrm{B}}^{2}(\overrightarrow{\mathrm{r}}) \mathrm{d} \overrightarrow{\mathrm{r}}\right]^{1 / 2} \tag{1}
\end{equation*}
$$

[^1] Chem. 1988. 2, 299. 325.
(3) Carbo. R.: Leyda, L.: Arnau, M. Int. J. Quantum Chem. 1980. 17, 1185.

The similarity index, $R_{\mathrm{AB}}$, complies with points 1 and 2 of our proposition. However, calculations of $R_{\mathrm{AB}}$ are computationally very demanding. To reduce the amount of necessary computations, the authors of ref 3 proposed CNDO-like approximations in calculations of the necessary integrals. Once these approximations are implemented, however, the (approximate) index does depend explicitly on the basis sets used in computing the electron densities. On the other hand, the computational cost of implementing eq 1 allows only very small molecules to be compared. Only sin-gle-point (no optimization of the orientation of molecules) calculations of $R_{\mathrm{AB}}$ were attempted and were found to be time consuming, although only a minimal basis set (STO-3G) and a powerful parallel computer system were employed. ${ }^{4}$

In this paper we report a new definition of the molecular similarity index which makes it possible to compare the electronic structure of molecules without excessive computational cost. Organization of the paper is as follows: First, we outline the theoretical approach. Second, we explain the computational details of our calculations. Sample calculations of the similarity index are reported and followed by a discussion of the results. Finally, the Conclusions section addresses advantages of the new approach.

## Theory

Let the two molecules under comparison, $\mathbf{A}$ and B , be described by the respective first-order density matrices ${ }^{5} \Gamma_{A}\left(\overrightarrow{\mathrm{x}}, \overrightarrow{\mathrm{x}}^{\prime}\right)$ and $\Gamma_{\mathrm{B}}\left(\overrightarrow{\mathrm{x}}, \overrightarrow{\mathrm{x}}^{\prime}\right)$. These matrices are conveniently expanded in terms of the natural spin-orbitals ${ }^{6}$

$$
\begin{equation*}
\Gamma_{\mathrm{A}}\left(\overrightarrow{\mathrm{x}}, \overrightarrow{\mathrm{x}}^{\prime}\right)=\sum_{\mathrm{i}} n_{\mathrm{Ai}} \phi_{\mathrm{Ai}}^{*}(\overrightarrow{\mathrm{x}}) \phi_{\mathrm{Ai}}\left(\overrightarrow{\mathrm{x}}^{\prime}\right) \tag{2}
\end{equation*}
$$

and

$$
\begin{equation*}
\Gamma_{\mathrm{B}}\left(\overrightarrow{\mathrm{x}}, \overrightarrow{\mathrm{x}}^{\prime}\right)=\sum_{\mathrm{j}} n_{\mathrm{Bj}} \phi_{\mathrm{Bj}}^{*}(\overrightarrow{\mathrm{x}}) \phi_{\mathrm{Bj}}\left(\overrightarrow{\mathrm{x}}^{\prime}\right) \tag{3}
\end{equation*}
$$

One can assess the similarity between $A$ and $B$ by calculating the distance between $\Gamma_{A}$ and $\Gamma_{B}$ :

$$
\begin{align*}
D_{\mathrm{AB}} & =\iiint_{\mathrm{A}}\left|\Gamma_{\mathrm{A}}\left(\overrightarrow{\mathrm{x}}, \overrightarrow{\mathrm{x}}^{\prime}\right)-\Gamma_{\mathrm{B}}\left(\overrightarrow{\mathrm{x}}, \overrightarrow{\mathrm{x}}^{\prime}\right)\right|^{2} \mathrm{~d} \overrightarrow{\mathrm{x}} \mathrm{~d} \overrightarrow{\mathrm{x}}^{\prime} \\
& =\sum_{\mathrm{i}} n_{\mathrm{A}}{ }^{2}+\sum_{\mathrm{i}} n_{\mathrm{Bi}}{ }^{2}-2 \sum_{\mathrm{ij}} n_{\mathrm{Ai}} n_{\mathrm{Bj}}\left|\left\langle\phi_{\mathrm{Ai}} \mid \phi_{\mathrm{Bj}}\right\rangle\right|^{2} \tag{4}
\end{align*}
$$

The case of $D_{\mathrm{AB}}=0$ would correspond to a perfect matching between molecules $A$ and $B$, whereas any dissimilarity between A and B would manifest itself in a positive value of $D_{\mathrm{AB}}$. Since the first two terms in eq 4 are constant for a given pair of molecules, we postulate that the quantity

$$
\begin{equation*}
(\mathrm{A}, \mathrm{~B})=\sum_{\mathrm{ij}} n_{\mathrm{Ai}} n_{\mathrm{Bj}}\left|\left\langle\phi_{\mathrm{Ai}} \mid \phi_{\mathrm{Bj}}\right\rangle\right|^{2} \tag{5}
\end{equation*}
$$

be regarded as a measure of the molecular similarity between $A$ and $B$.

The similarity index, eq 5 , which we propose to be termed "the number of overlapping electrons" (NOEL), bears some resemblance to the "pars-orbital character" introduced some time ago by Polansky and Derflinger' and subsequently elaborated by Polansky et al. ${ }^{8}$ and Golebiewski. ${ }^{9}$ However, our definition is more general because NOEL can be calculated from wave functions of any form. A closer look at eq 5 reveals several desirable properties of our approach. The similarity index is a well-defined quantity derived from the overlap of the respective first-order density matrices. It does not invoke any explicit references to the basis set or the actual form of the wave function. For the SCF wave function, it assumes the value equal to the
(4) Bowen-Jenkins. P. E.: Richards. W. G. Int. J. Quantum Chem. 1986. 30. 763.
(5) Davidson, E. R. Reduced Density Matrices in Quantum Chemistry: Academic Press: New York. 1976.
(6) Löwdin. P. O. Phys. Rev. 1955, 97, 1474.
(7) Polansky. O. E.: Derflinger. G. Int.J. Quantum Chem. 1967. I. 379.
(8) Fratev. F.: Polansky, O. E.: Mehlhorn. A.: Monev. V. J. Mol. Struct. 1979. 56, 245.
(9) Golebicwski, A. Acta Phys. Polonica 1974. A46. 719.

Table I. HF/6-31G** Energies and Dipole Moments at the Optimized Geometries of the Benzene, Aniline, Nitrobenzene, and 4-Nitroaniline Molecules and the Acetate, Isoxazole 3-Oxide, and Isoxazole 5-Oxide Anions

| molecule | symmetry | energy, au | dipole <br> moment, D |
| :--- | :---: | :---: | :---: |
| benzene | $D_{6 h}$ | -230.7138600 | 0.0 |
| aniline | $C_{s}$ | -285.7475068 | 1.5212 |
| nitrobenzene | $C_{2}$ | -434.1842678 | 5.0670 |
| $p$-nitroaniline | $C_{2 v}$ | -489.2224214 | 7.4675 |
| acetate anion | $C_{s}$ | -227.2298676 | $\mathrm{n} / \mathrm{a}$ |
| isoxazole 3-oxide anion | $C_{s}$ | -318.8762987 | $\mathrm{n} / \mathrm{a}$ |
| isoxazole 5-oxide anion | $C_{s}$ | -318.9020512 | $\mathrm{n} / \mathrm{a}$ |

Table II. Similarity between the Benzene Molecule and the Molecules of Benzene, Aniline, Nitrobenzene, and 4-Nitroaniline

| molecule | NOEL |
| :--- | :--- |
| benzene | 42.000 |
| aniline | 41.825 |
| nitrobenzene | 41.730 |
| 4-nitroaniline | 41.619 |

Table III. Similarity between the Aniline, Nitrobenzene, and 4-Nitroaniline Molecules

|  | comparison $^{a}$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| orientation | aniline-aniline | nitrobenzene- <br> nitrobenzene |  |  |  |
| ipso | 50.000 | 49.398 | 64.000 | 64.000 | 72.000 |
| ortho | 41.610 | 41.622 | 41.838 | 41.838 | 41.653 |
| meta | 41.607 | 41.621 | 41.435 | 41.435 | 41.416 |
| para | 41.754 | 41.726 | 41.517 | 41.517 | 55.288 |

${ }^{a} 4-\mathrm{NA}=4$-nitroaniline.
number of electrons in molecule A , if molecules A and B are identical. If both molecules $A$ and $B$ share an identical fragment, then ( $\mathrm{A}, \mathrm{B}$ ) equals the number of the electrons in this molecular fragment plus the magnitude of the residual overlap between the remaining parts of A and B . This observation justifies the proposed name for ( $\mathrm{A}, \mathrm{B}$ ).

## Details of Calculations

All of the molecular orbital calculations were carried out at the $\mathrm{HF} / 6-31 \mathrm{G}^{* *}$ level of theory with full geometry optimization. The relevant total energies and dipole moments are reported in Table I. The Gamess package ${ }^{10}$ was used to optimize the geometries, and the final wave functions were calculated with the GAUSSIAN 88 suite of programs. ${ }^{11}$

Computations of the NOEL similarity index proceed as follows: First, the molecular orbitals are read from the checkpoint files of the GAUSSIAN 88 package. The necessary overlap integrals, eq 5 , are computed with the aid of a three-point Gaussian quadrature which is exact for the $s$, $p$, and $d$ basis functions. The derivatives of NOEL with respect to the translations and rotations of molecule B are computed analytically at the same time. One should point out that. in general, the magnitude of NOEL is a function of three translations ( $x, y$, and $z$ ) and three rotations (three Euler angles). In addition to this, there are two distinct positions of molecule $B$, approximately related to each other by a plane reflection. The starting molecular orientation is fixed by specifying a triad of atoms from each molecule. The starting orientation is then found by minimizing the distances between the atoms belonging to the two triads. The optimization of the mutual orientation of molecules by maximizing the value of NOEL is initiated with calculation of the Hessian matrix by finite differences. The Hessian matrix is updated throughout the optimization process. It usually takes three to five iterations to maximize NOEL within the accuracy of $10^{-10}$. The calculations presented in the subse-

[^2]Table IV. Similarity between the Aniline, Nitrobenzene, and 4-Nitroaniline Molecules

|  | comparison $^{a}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| orientation | aniline- <br> nitrobenzene | aniline- <br> 4-NA | nitrobenzene- <br> 4-NA |  |
| ipso | 48.716 | 48.716 | 49.343 | 63.802 |
| ortho | 41.674 | 41.675 | 41.452 | 41.755 |
| meta | 41.591 | 41.591 | 41.532 | 41.440 |
| para | 41.475 | 41.475 | 48.674 | 48.309 |

${ }^{a} 4$-NA $=4$-nitroaniline.
quent section of this paper require ca. $2-3 \mathrm{~h}$ of CPU time on the VAX 3100 workstation per molecular orientation. The computer program, which is fully interfaced with the GAUSSIAN 88 package, is available from one of the authors (J.C.) upon request.

## Sample Calculations and Discussion

The major features of our new approach are clearly demonstrated by ab initio electronic structure calculations on the similarity between the benzene, aniline, nitrobenzene, and 4 -nitroaniline molecules. Interpretation of the calculated magnitudes of the NOEL index, which are reported in Tables II-IV, is aided by the observation that the numerical value of NOEL is expected to be close to the number of electrons in the molecular fragment shared by both molecules. This number is decreased by any instance of an imperfect overlap between the nuclei in the matching fragment. On the other hand, the residual overlap between the remaining parts of the molecules under comparison gives rise to an increased value of NOEL.

While comparing the substituted benzenes with the benzene molecule itself (Table II), one can expect the values of NOEL to be close to 41 , for the $\mathrm{C}_{6} \mathrm{H}_{5}$ fragment contributes $6 \times 6+5$ $=41$ electrons. The actual numbers for the aniline and nitrobenzene molecules are higher by about 0.8 , which demonstrates a significant residual similarity between the hydrogen atom and the amino and nitro groups. For the 4 -nitroaniline molecule one could expect the value of 40 (for the $\mathrm{C}_{6} \mathrm{H}_{4}$ fragment). The actual number of 41.6 arises from an additional residual overlap, which contributes the remaining 1.6 electrons.

As mentioned in the previous section, for each starting molecular orientation there are in general two distinct final orientations approximately related to each other by a plane reflection. This is the case for the aniline-aniline and nitrobenzene-nitrobenzene pairs (Table III). For these pairs, as well as for the system of two 4-nitroaniline molecules, there are four starting orientations resulting in either four or eight local maxima of NOEL and the same number of final orientations. The values of NOEL for the ipso orientation correspond to either perfect matching of molecules or a slight dissimilarity between the molecules and their mirror images that result from the nonplanarity. This dissimilarity is quite substantial for the aniline molecule, due to the pyramidal arrangement of the hydrogens bonded to the nitrogen atom. The small twisting of the nitro group in the nitrobenzene molecule does not affect the value of NOEL significantly.

The magnitudes of NOEL for the ortho, meta, and para orientations of the aniline and nitrobenzene molecules reflect sharing of the $\mathrm{C}_{6} \mathrm{H}_{5}$ fragment. The ortho and meta orientations in the 4 -nitroaniline molecule give rise to a common $\mathrm{C}_{6} \mathrm{H}_{4}$ fragment, whereas the value of NOEL of 55.3 for the para arrangement describes sharing of the $\mathrm{C}_{6} \mathrm{H}_{4} \mathrm{~N}_{2}$ fragment $(6 \times 6+4+2 \times 7$ $=54)$ together with some residual overlap.

Similar trends are observed in the comparisons between different molecules (Table IV). The aniline-nitrobenzene pair has a common $\mathrm{C}_{6} \mathrm{H}_{5}$ fragment for the ortho, meta, and para orientations. One expects sharing of the $\mathrm{C}_{6} \mathrm{H}_{5} \mathrm{~N}$ fragment with $6 \times 6+5+$ $7=48$ electrons for the ipso arrangement. The actual value of NOEL (48.7) confirms that this is the case. The agreement between the expected and actual magnitudes of NOEL is also substantial in the comparisons with the 4 -nitroaniline molecule. Some relevant numbers read as follows: for aniline vs 4 -nitroaniline ipso expected $49\left(\mathrm{C}_{6} \mathrm{H}_{4} \mathrm{NH}_{2}\right)$ and actual 49.3 , and para
expected $47\left(\mathrm{C}_{6} \mathrm{H}_{4} \mathrm{~N}\right)$ and actual 48.7 ; for nitrobenzene vs 4 nitroaniline ipso expected $63\left(\mathrm{C}_{6} \mathrm{H}_{4} \mathrm{NO}_{2}\right)$ and actual 63.8 , and para expected $47\left(\mathrm{C}_{6} \mathrm{H}_{4} \mathrm{~N}\right)$ and actual 48.3.

The comparisons among substituted benzenes illustrate basic features of our approach. The computed values of NOEL follow the pattern expected from the topological characteristics (structural formulas) of the molecules under comparison. In contrast, the degrees of similarity among the acetate, isoxazole 3 -oxide (1) and isoxazole 5 -oxide (2) anions cannot be directly assessed from their


1


2
structures. Both $\mathbf{1}$ and $\mathbf{2}$ contain a molecular fragment reminescent of the carboxyl anion. However, while in 1 one of the oxygens is replaced by the nitrogen atom, in 2 the relevant carbon-carbon bond is expected to be shorter than in the acetate anion. Our calculations show that $\mathbf{1}$ is more similar to the acetate anion than 2, as reflected by the values of NOEL ( 28.480 vs 27.786). The difference comes primarily from the mismatch of the molecular orbitals, which is more pronounced in the isoxazole 5-oxide-acetate pair. The similarity index for the isoxazole 3 -oxide-isoxazole 5 -oxide pair equals 41.149 (to be compared with 44 electrons present in each of the molecules). One should point out that the fact that $\mathbf{1}$ is more similar to the acetate anion than $\mathbf{2}$ may easily explain a 5000 -fold difference in the physiological activities of the two powerful GABA agonists: muscimol and isomuscimol. ${ }^{12}$ Both GABA and its agonists are believed to bind to the GABA $A_{A}$ cellular receptor through their (pseudo)carboxyl anionic termini. ${ }^{13}$

## Conclusions

We believe that the above sample calculations demonstrate adequately the advantages of our definition of molecular similarity. The proposed similarity index is rigorously defined and can be used in connection with any form of electronic wave functions. Interpretation of the computed similarity index is straightforward since the magnitude of NOEL depends to the first-order approximation on the similarities in the atomic connectivity (topology) of the molecules under comparison. Subtle differences in similarity between like chemical systems can be recognized easily since small additional variations in NOEL are induced by both geometrical and electron distribution factors.

Our test calculations indicate that computation of the NOEL index is very rapid. The reason is that the computational cost of calculating NOEL scales with only the third power of the number of electrons. This property of our index makes it possible to perform similarity calculations for systems large enough to be of interest in pharmaceutical research. The cost of computing the index $R_{\mathrm{AB}}$, proposed by Carbo et al. ${ }^{3}$ increases proportionally to the fourth power of the size of the systems under comparison. The same is true about the modified index, $H_{\mathrm{AB}}$, introduced by Hodgkin and Richards. ${ }^{14}$

One should stress that finding the optimum alignment of molecules constitutes a multiple-maximum optimization of the similarity index. Depending on a particular application, the molecular similarity can be defined as the value of the index corresponding to a global maximum or, alternatively, to a specific local maximum. In either case, a rapid evaluation of the similarity index is crucial for locating these maxima. Our definition of the molecular similarity allows for extensive optimizations without excessive computational cost.

We believe that, when properly interpreted, the NOEL simi-

[^3]larity index can be a powerful tool for systematizing the properties of chemical systems in a rigorous manner. Besides applications in drug design, the similarity index can be useful in correlating reactivities of functional groups in organic systems, in comparing similar molecules, in computerized interpretation of the IR spectra, and in pattern recognition.

Acknowledgment. This research was partially supported by the U.S. D.O.E. under contract DE-FC05-85ER250000 and by the Florida State University through time granted on its VAX and ETA10-G digital computers. One of the authors (J.C.) also acknowledges support by the Camille and Henry Dreyfus Foundation New Faculty Award Program.

# Simulation of the Structure and Dynamics of the Bis(penicillamine) Enkephalin Zwitterion 

Paul E. Smith, Liem X. Dang, ${ }^{\dagger}$ and B. Montgomery Pettitt*<br>Contribution from the Department of Chemistry, University of Houston. 4800 Calhoun Road, Houston, Texas 77204-5641. Received April 16, 1990


#### Abstract

We have performed a molecular dynamics simulation of the enkephalin derivative Tyr-(D)Pen-Gly-Phe-(D)Pen (DPDPE) in aqueous solvent. Electrostatic interactions were calculated with the Ewald method so as to accurately model the large interactions between the DPDPE zwitterion and the solvent and to avoid the use of electrostatic cutoff's or neutral chemical blocking groups. DPDPE is found to be extremely constrained with very little variation in the main-chain dihedral angles. Flexibility found in the region of the central glycine is greatly restricted compared with glycines in straight-chain peptides. Several conformational transitions are observed for the two aromatic side chains, indicating a high degree of flexibility in the side chains and that DPDPE does not have a single conformation in solution. This suggests that the arrangement of the tyrosine and phenylalanine aromatic residues in the bound conformation may be selected by the receptor environment. The main-chain conformation of DPDPE in solution has a parallel arrangement of peptide groups. This electrostatically unfavorable structure is stabilized by interactions of the carbonyls with the solvent. Solvent structure around the $\mathbf{N}$ terminus is found to be considerably localized, involving short ammonium cation to water contacts with lifetimes of the order of 50 ps or more. In comparison, water structure around the $C$ terminus is much more mobile with lifetimes of the order of 20 ps or less.


## 1. Introduction

When studying the interactions between receptors and their ligands it is usual to have a reliable structure for both the receptor and the ligand. Without prior knowledge of at least one of these structures the possibility of isolating the major stabilizing interactions is considerably reduced if not impossible. Many systems of biochemical importance for which the structure or conformation of the receptor is unknown can be approached by using synthetically constrained ligands. One such system is that of the neurotransmitters and their interactions with the opioid receptors. ${ }^{1}$

Neurotransmitters are small peptides that facilitate the transfer of nerve impulses across the synaptic cleft. The initial nerve impulse along the neuron triggers the release of the neurotransmitters into the cleft where they diffuse toward a series of receptors on the opposite side of the cleft. The interaction between the neurotransmitter and the receptor enables the continuation of the nerve impulse. A wealth of information has been obtained on the action of different neurotransmitters with their respective receptors, ${ }^{1.2}$ but very little structural detail is known. For opioids the problem is also compounded by the presence of more than one receptor type ( $\delta, \mu$, and $\kappa$ are known), each of which is specific for a different neurotransmitter or different conformations of the same neurotransmitter.

As the structure of the receptor sites is unknown the traditional approach to investigate the interactions between a receptor and a ligand is to search for common, and therefore important, structural features (pharmacophores) between known neurotransmitters. ${ }^{3-5}$ This was attempted by starting with the two endogenous neuromodulators Tyr-Gly-Gly-Phe-Leu and Tyr-Gly-Gly-Phe-Met, known as Leu and Met enkephalin, respectively. By adding, deleting, and modifying various residues and then testing for biological activity, valuable information has been gained concerning the important features necessary for activity. ${ }^{3}$ Most

[^4]significantly, synthetic constraints were used to severely restrict the number of conformations available to the peptide at normal temperatures.
This type of mapping was used with some success and yielded the most specific and one of the most potent neurotransmitters known for the $\delta$-opioid receptor. ${ }^{3}$ This neurotransmitter is a pentapeptide enkephalin derivative with sequence Tyr -(D)Pen$\overline{\text { Gly-Phe-(D)Pen (DPDPE). The penicillamine residue (an iso- }}$ propylcysteine derivative) was used to constrain the ring structure via a disulfide bond.
DPDPE has been studied by NMR, model building, molecular mechanics, and quenched high-temperature molecular dynamics. ${ }^{6}$ One major class of conformations was found to satisfy the experimental data but did not correspond to the lowest energy conformation using continuum solvent models. The search for a representative population distribution of conformers was hindered by the simplified treatment. Simulations under vacuum always gave structures with a strong interaction, or salt bridge, between the ammonium and carboxylate termini (unpublished results). This contradicted the available NMR evidence. On the other hand simulations using the dielectric constant of bulk water severely diminished all the electrostatic interactions, especially those with the ionic groups. Hence, the correct modeling of the solvent and

[^5]
[^0]:    (1) Amoore, J. E. Molecular Basis of Odor: Thomas: Springfield. IL. 1970.

[^1]:    (2) Mezey, P. G. J. Comp. Chem. 1987, 8, 462. Mezey, P. G. J. Math.

[^2]:    (10) Original version by: Dupuis, M.: Spangler, D.; Wendoloski, J. J. NRCC Program QG01 (1980). ETA10-G conversion by Moncrieff, D.: Fleischmann. E. D.
    (11) GAUSSIAN 88: Frisch. M. J.: Head-Gordon. M.: Schlegel. H. B.: Raghavachari. K.: Binkley. J. S.: Gonzalez. C.: Defrees. D. J.: Fox. D. J.: Whiteside, R. A.: Seeger, R.: Melius, C. F.: Baker. J.: Martin. R. L.: Kahn. L. R.: Stewart, J. J. P.: Fluder. E. M.: Topiol, S.: Pople. J. A. GAUSSIAN. Inc.: Pittsburgh PA.

[^3]:    (12) Cioslowski. J.: Fleischmann, E. D. Proc. Natl. Acad. Sci. U.S.A. Submitted for publication.
    (13) Boulander, T.; Vercauteren. D. P.: Durand, F.; Andre, J. M. Int. J. Quantum Chem. Biol. Symp. 1988. 15. 149.
    (14) Hodgkin. E. E.: Richards, W. G. Int. J. Quantum Chem. Biol. Symp. 1987. 14, 105.

[^4]:    ${ }^{\dagger}$ Present address: 1BM Research Labs, Almaden Research Center, 650 Harry Road. San Jose. CA 95120-6099.

[^5]:    (1) Schiller, P. W. In Pepiides: Analysis, Synthesis, Biology, Volume 6, Opioid Peptides: Biology, Chemistry, and Genetics; Udenfriend, S.: Meienhofer, J., Eds.; Academic Press: New York, 1984; Vol. 6, pp 219-268.
    (2) Hruby, V. J.: Krstenansky, J. L.: Cody, W. L. Annu. Rep. Med. Chem. 1984, 19, 303-312.
    (3) Hruby, V. J.; Pettitt, B. M. In Computer-Aided Drug Design: Methods and Applications; Perun, T. J.; Propst, C. L., Eds.; Marcel Dekker: New York, 1989, pp 405-460.
    (4) Mierke. D. F.; Said-Nejad, O. E.: Schiller, P. W.; Goodman, M. Biopolymers 1990, 29, 179-196.
    (5) Aubry. A.; Birlirakis, N.: Sakarellos-Daitsiotis, M.: Sakarellos, C.; Marraud, M. Biopolymers 1989, 28, 27-40.
    (6) Hruby, V. J.: Kao, L.; Pettitt. B. M.; Karplus, M. J. Am. Chem. Soc. 1988, 110, $3351-3359$.

