Deriving a quantitative chirality measure from molecular similarity indices.

ARTICLE *in* JOURNAL OF MEDICINAL CHEMISTRY · NOVEMBER 2000

Impact Factor: 5.45 · Source: PubMed

Deriving a Quantitative Chirality Measure from Molecular Similarity Indices

Romualdo Benigni,[§] Marina Cotta-Ramusino,*[,†] Grazia Gallo,[#] Fabrizio Giorgi,[#] Alessandro Giuliani,[§] and Maria Rosaria Varì[†]

Laboratorio di Tossicologia Comparata ed Ecotossicologia and Laboratorio di Chimica del Farmaco, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy, and Laboratori di Chimica, Sigma-Tau S.p.A., Via Pontina Km. 30.400, 00040 Pomezia, Rome, Italy

Received January 31, 2000

A versatile new method has been developed as a continuous symmetry measure for chiral compounds. The application of principal component analysis (PCA) to the complete $N \times N$ pairwise similarity matrices (electrostatic potential and shape indices) of a series of dihydropyridine calcium channel antagonists allowed to single out a chirality component and to compute a chirality score in terms of the between-enantiomers difference on the component value. The possibility to have chirality defined continuously at the series level could be of importance in eudismic analyses where the relative potency of two enantiomers is studied as well as in QSAR studies dealing with chiral molecules in order to improve the power of the generated models.

Introduction

Chirality, despite its importance in medicinal chemistry applications, remains quite an elusive concept for QSAR parametrization. In recent years, a number of articles have focused on the possibility of describing chirality as a "continuous" property instead of a "black and white" property.

Chirality metrics have been developed during the past decade which mainly made use of mathematical and computational methods.[1−4] Zabrodsky et al. proposed a "continuous symmetry measure" (CSM)[5] method which quantifies the minimal distance movement for points of an object in order to transform it into a shape of the desired symmetry, while Lipkowitz et al.[6] have assessed the orbital desymmetrization induced by common chiral auxiliaries. On the other hand, Harris et al.[3] demonstrated the possibility to generate chiral measures of a geometric object in terms of pseudoscalars (invariant under proper rotations but changing sign under improper rotations) vanishing at different points as a molecule is continuously deformed into its mirror image. From this it is concluded that handedness is not an absolute concept but depends on the property being observed.

At the very basis of the idea of chirality is the concept of superposition of two molecular graphs and that of impossibility of a perfect matching by means of rotation and translation operators for an enantiomer couple (absence of reflection symmetry). It should be noted that also the computation of the similarity indices between two molecules is based on the superposition of the two molecular graphs. The molecules are embedded into 3D grids, and the correlation between the value of different properties (electrostatic potentials and fields, shapes, lipophilic potentials, etc.) at the same position in space is computed at the orientation corresponding to the maximal superposition between the two molecules.

This common dependence from superposition prompted us to investigate the possibility of deriving a chirality measure from molecular similarity computations. Our line of reasoning was similar to the one previously adopted by Seri-Levy et al.[7] who tried to establish a correlation between the eudismic ratio (i.e. the relative potency of two drug enantiomers) and their chiral coefficient defined as 1 − similarity. Our approach is different in that we used principal component analysis (PCA) to extract, from the overall similarity, the specific contribution of chirality.

It is worth noting that, on a pure topological basis in 2D,[4] it is possible to measure chirality in terms of graph invariants; however, we followed a completely different approach, investigating the possibility to make chirality "emerge" as a 3D feature directly by its basic definition of nonsuperimposability of two mirror images.

The application of PCA to the complete $N \times N$ pairwise similarity matrix of a congeneric series of chiral structures allowed us to single out a "chirality component" which orders the elements of each enantiomer couple in symmetrical positions with respect to the mean, thus allowing the direct computation of a chirality score in terms of the between-isomers difference on the component value. This generated an explicit quantitative chirality measure at the congeneric series level directly usable for an eudismic QSAR analysis.

Materials and Methods

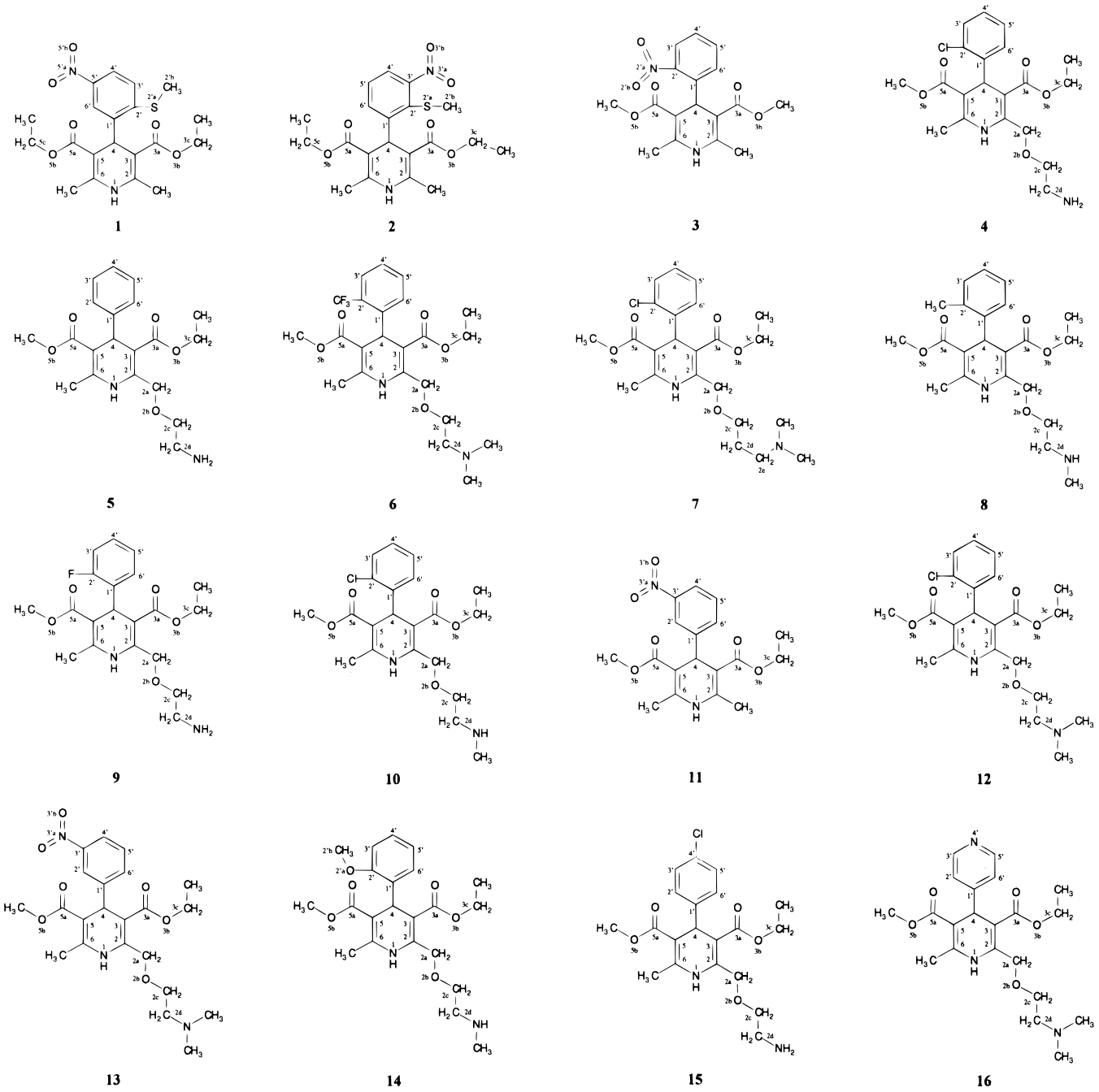
Data Set. The selected compounds were a subset of the dihydropyridine calcium channel inhibitors studied by Arrowsmith et al.[8] together with two sulfur-containing dihydropiridine derivatives[9] and the reference compounds nifedipine and nitrendipine (see Chart 1, structures **3** and **11**, respectively). The presence of nonchiral molecules (compounds **1**−**3**) allowed us to perform a consistency test: a reliable "chirality score" must give an average value to nonchiral structures.

Computational Procedures. The molecular structures of the dihydropyridine derivatives were constructed with the molecular modeling package Sybyl[10] and fully optimized with the Tripos force field method.[11] The conformational space of

* Author for correspondence. Tel: +390649902627. Fax: +390649387100. E-mail: m.cotta@iss.it.
[§] Laboratorio di Tossicologia Comparata ed Ecotossicologia, Istituto Superiore di Sanità.
[†] Laboratorio di Chimica del Farmaco, Istituto Superiore di Sanità.
[#] Sigma-Tau S.p.A.

10.1021/jm0009134 CCC: $19.00
Published on Web 09/09/2000

Chart 1. Molecular Structures of the Studied Molecules



the considered structures was first explored by a systematic search, varying the relevant dihedral angles (inter-ring and side chain ones) in either 15° or 30° (for compounds with a bigger number of dihedral angles) steps. The molecular structures corresponding to the relative minima on the conformational hypersurface and whose energy was not greater than 4 kcal·mol^{-1} with respect to the most stable conformer of each compound were submitted to semiempirical MO calculations. The AM1 Hamiltonian implemented in Mopac 6.0[12] was used with the BSGF procedure for SCF convergence.

Partial atomic (Coulson) charges were also computed for each molecular structure. The most stable conformer of each dihydropyridine derivative was then selected for the computation of pairwise similarity indices. For chiral compounds only one enantiomer (the one displaying the *R* configuration) was submitted to the semiempirical MO calculations; the corresponding *S* enantiomer was generated by reflection through the plane defined by the atoms of the dihydropyridine ring.

Computation of Pairwise Similarities. Since molecular similarity indices are very sensitive to the relative orientation of the considered compounds and depend on pairwise molecular indices, all the molecular structures were appropriately aligned using the "ALIGN DATABASE" command in Sybyl by which an optimal rigid alignment of each structure is sought with respect to a common template (the dihydropyridine ring). The similarity indices based on electrostatic potential and shape were calculated using the ASP software.[13] The electrostatic potential similarity index was obtained using the Carbo equation[14] and the three Gaussian function approximation to reproduce the inverse distance dependence of this property.[15] Shape indices were calculated as suggested by Meyer and Richards,[16] also using the Gaussian function approximation. Combined similarity indices were also computed with Asp.

Statistical Methodology. The $N \times N$ pairwise similarity matrix constitutes an implicit representation of the data set; to make it explicit and thus having the possibility of attaching numerical coordinates to each single molecular structure, we submitted the original similarity matrix to a PCA. The result of PCA is an eigenvector expansion of the original matrix, where the eigenvectors (principal components) with the largest

Table 1. PCA Solutions in Terms of Normalized Eigenvalues (percent of explained variance) for the Electrostatic Potential, Shape, and Combined Metrics[a]

Electrostatic Potential Similarity					
component	PCCH1	PCCH2*	PCCH3	PCCH4	PCCH5
% explained variance	53.3	19.02	11.0	6.73	2.74
cumulative	53.3	72.35	83.3	90.0	92.8
Shape Similarity					
component	PCSH1*	PCSH2	PCSH3	PCSH4	PCSH5
% explained variance	31.31	22.44	10.11	8.42	7.93
cumulative	31.31	53.75	63.86	72.27	80.20
Combined Similarity					
component	PCCOM1	PCCOM2*	PCCOM3	PCCOM4	PCCOM5
% explained variance	39.24	26.33	9.04	6.66	4.55
cumulative	39.24	65.57	74.61	81.27	85.82

[a] The asterisk marks the chiral component of each similarity matrix.

eigenvalues explain most of the variability in the original matrix. Thus the space spanned by the first principal components constitutes a Cartesian orthogonal space assigning to each single molecule a position consistent with its distances with all the other molecules.[17−19] Given that chirality contributes to the actual value of the distances between two molecules, we expect it to emerge as one of the principal components of the similarity matrix. The requirement for a principal component to be considered a chirality measure at the congeneric series level is that the component must exactly discriminate *R* and *S* isomers and place nonchiral molecules between the two *R* and *S* populations. If this is the case, it can be used as a measure of the amount of chirality of a molecular species by the computation of the difference between the component values relative to each isomer.

Results and Discussion

The three different metrics (electrostatic potential, shape, combined) generated the PCA solution reported in Table 1. Each of the three similarity matrices can be projected into a five-components space, explaining 93%, 80%, and 86% of total variability, respectively. Each of the three metrics gave rise to a chiral component: the first principal component of the shape metrics and the second component of both the electrostatic potential (EPS) and combined metrics. On the other components the two elements of each enantiomer couple have practically coincident values; this corresponds to the well-known fact that two enantiomers have identical chemicophysical features. This means that all the information on chirality present in the $N \times N$ distance matrices is embedded into the chiral components and that the relative importance of chirality in each data set roughly corresponds to the percentage of explained variance of the relative chiral component. From Table 1 we observe that chirality accounts for 31% of shape similarity variability and for 19% and 22% of EPS and combined similarity variability.

Figures 1 and 2 report the spaces spanned by the two major components of electrostatic potential and shape metrics: looking at the figures, one can immediately recognize the two basic requirements of the chirality score (between-isomers symmetry and average position for nonchiral structures). This recognition is made easier by the fact that principal components have, by construction, a zero mean and a unitary standard deviation.

All three chirality components (PCCH2, PCSH1, PCCOM2) are strongly correlated (Table 2), thus pointing to a reliability of the extracted chirality measure and, indirectly, to the estimate we did of the relative importance of chirality in the different data sets: as a

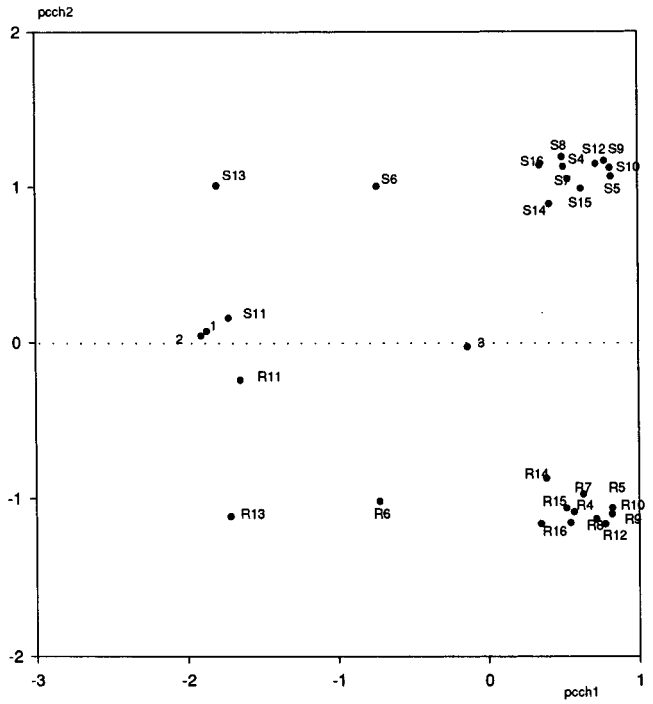


Figure 1. Molecules in the space spanned by the first two components of the electrostatic potential metrics. The dotted line corresponds to the mean value of the chiral component; the nonchiral molecules (**1**−**3**) are positioned near the line; the elements of each enantiomer pair occupy approximately symmetrical positions with respect to the average line.

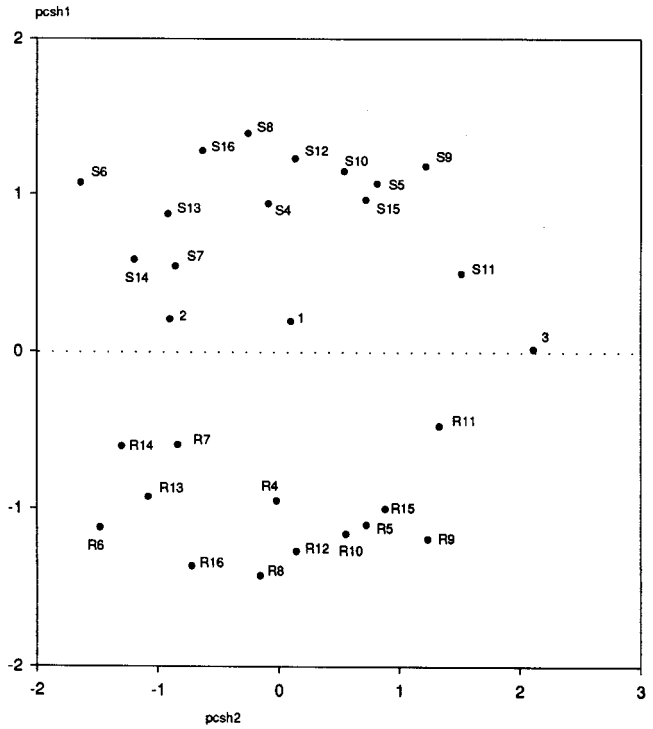


Figure 2. Molecules in the space spanned by the first two components of the electrostatic potential metrics. The dotted line corresponds to the mean value of the chiral component; the nonchiral molecules (**1**−**3**) are positioned near the line; the elements of each enantiomer pair occupy approximately symmetrical positions with respect to the average line.

Table 2. Pearson Product Moment Correlation Coefficients between Chirality Components of the Three Data Sets[a]

	PCCH2	PCSH1	PCCOM2
PCCH2	1.00	0.981	0.997
PCSH1		1.00	0.991
PCCOM2			1.00

[a] The extremely high level of correlation points to a substantial concordance of the three metrics as for chirality ordering of molecules.

Table 3. Enantiomer Pairs in Decreasing Order of Chirality[a]

enantiomer pair	chirality index	enantiomer pair	chirality index
8	2.578	**4**	2.088
16	2.462	**15**	2.054
12	2.411	**13**	1.989
9	2.326	**7**	1.595
10	2.267	**14**	1.492
5	2.152	**11**	0.661
6	2.117		

[a] The chirality index corresponds to the difference between the elements of each couple in terms of CHIR (synthethic chirality score).

matter of fact all three chiral components report the same information. This strong correlation allows us to cumulate all three chirality scores into a general synthetic chirality index (CHIR) by means of PCA. The computation of the difference in terms of CHIR between the elements of each enantiomer pair (Table 3) allowed us to put all the considered molecular species along a continuous chirality axis.

Our results confirm the possibility to obtain a continuous measure of chirality for organic molecules. This measure is not absolute but relative to a given congeneric series. This stems from the fact that the superposition of the compared molecules is an unescapable prerequisite for the computation of similarity indices, and only in the case of congeneric series does the search for the maximal superposition between two molecular structures give rise to unambiguous results. Nevertheless, for the great majority of applications, this is not a major drawback.

Moreover our results indicate that the molecular similarity indices, at odds with classical chemicophysical descriptors, retain the chirality information. This can have a major importance in QSAR studies. Let us imagine a QSAR equation expressed in the usual form: $\log(1/C) = a(\text{partition}) + b(\text{steric}) + c(\text{electronic})$. In this equation the two elements of an enantiomer couple have the same values of the regressors but not necessarily the same value of the dependent variable $\log(1/C)$, given that the modeled activity could involve the interaction with a receptor discriminating the two isomers. If this is the case, adding to the above equation the chirality term could ameliorate the obtained fit.

The generality of our approach, i.e., the possibility of generating a "pure chiral" component for each application of the proposed method, remains an open question. As a matter of fact, the results of PCA heavily depend on the considered data set, with the extracted components being the "order parameters" which organize the variability of the data set.[20] We expect that also in other situations, with a well-balanced series of chiral congeners, such a chiral component can be found. This means that the selection of the congeners series is the crucial step for a pure chiral component to be singled out.

It is worth noting the difference between our approach and the one proposed by Seri-Levy and colleagues. The Oxford group[7] considered as the "chirality coefficient" of a molecular species the distance (1 − similarity) between the two enantiomers; on the contrary we computed the chirality score in terms of a chirality component derived from the entire $N \times N$ similarity matrix. This means that our chirality index is a global feature of the series, which in turn is projected at the local level of the single enantiomer couples so allowing for a consistent chirality ordering along the series to be made. As a matter of fact, we demonstrated that chirality per se explains only a comparatively minor part of the global distance information of our series (31% of shape similarity, 22% and 19% of EPS and combined metrics, respectively, see Table 1). This is particularly cogent, since molecular similarity carries a large amount of information, including that coded in the usual physicochemical descriptors of molecules.[17,21] In our approach this information was filtered out from the pure chirality by the principal components algorithm.[22] PCA allows us to consider chirality as an autonomous concept out of the particular molecular features used to compare the molecules.

This implies that the extraction of a chirality component can allow for a direct estimation of the role of chirality in a given pharmacological activity, without the confounding of other factors so putting on firm bases the eudismic analysis of drug action.

References

(1) Buda, A. B.; Auf der Heyde, T.; Mislow, K. On Quantifying Chirality. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 989−1007.
(2) Randic, M.; Razinger, M. Molecular Shapes and Chirality. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 429−441.
(3) Harris, A. B.; Kamien, R. D.; Lubienski, T. C. Molecular Chirality and Chiral parameters. *Rev. Modern Phys.* **1999**, *71*, 1745−1756.
(4) Randic, M. On the Characterization of Molecular Attributes. *Acta Chim. Slov.* **1998**, *45*, 239−252.
(5) Zabrodsky, H.; Peleg, S.; Avnir, D. Continuous Symmetry Measures. *J. Am. Chem. Soc.* **1992**, *114*, 7843−7851.
(6) Lipkowitz, B. K.; Gao, D.; Katzenelson, O. Computation of Physical Chirality: An Assessment of Orbital Desymmetrisation Induced by Common Chiral Auxiliaries. *J. Am. Chem. Soc.* **1999**, *121*, 5559−5564.
(7) Seri-Levy, A.; West, S.; Richards, W. G. Molecular Similarity, Quantitative Chirality, and QSAR for Chiral Drugs. *J. Med. Chem.* **1994**, *37*, 1727−1732.
(8) Arrowsmith, J. E.; Campbell, S. F.; Cross, P. E.; Stubbs, J. K.; Burges, R. A.; Gardiner, D. G.; Blackburn, K. J. Long-Acting Dihydropyridine Calcium Antagonists. 1. 2-Alkoxymethyl Derivatives Incorporating Basic Substituents. *J. Med. Chem.* **1986**, *29*, 1696−1702.
(9) Baldwin, J. J.; Claremon, D. A.; Lumma, P. K.; McClure, D. E.; Rosenthal, S. A.; Winquist, R. J.; Faison, E. P.; Kaczorowski, G. J.; Trumble, M. J.; Smith, G. M. Diethyl 3,6-Dihydro-2,4-dimethyl-2,6-methano-1,3-benzothiazocine-5,11-dicarboxylates as Calcium Entry Antagonists: New Conformationally Restrained Analogues of Hansch 1,4-Dihydropyridines Related to Nitrendipine as Probes for Receptor-Site Conformation. *J. Med. Chem.* **1987**, *30*, 690−695.
(10) SYBYL 6.4; Tripos Associates, Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144-2913.
(11) Clark, M.; Cramer III, R. D.; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982−1012.
(12) Mopac 6.0; Frank J. Sailer Res. Lab., U.S. Force Academy, Colorado Springs, CO 80840.
(13) Automated Similarity Package; Oxford Molecular Ltd., The Magdalen Centre, Oxford Science Park, Sandford on Thames, Oxford OX4 4GA, United Kindom.
(14) Carbò, R.; Leyda, L.; Arnau, M. An electron density measure of the similarity between two compounds. *Int. J. Quantum Chem.* **1980**, *17*, 1185−1189.
(15) Good, A. C.; Hodkin, E. E.; Richards, W. G. Utilisation of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188−191.

(16) Meyer, A. Y.; Richards, W. G. Similarity of molecular shape. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 426−439.
(17) Benigni, R.; Cotta-Ramusino, M.; Giorgi, F.; Gallo, G. Molecular similarity matrices and quantitative structure−activity relationships: a case study with methodological implications. *J. Med. Chem.* **1995**, *38*, 629−635.
(18) Benigni, R.; Giuliani, A. Analysis of Distance Matrices for Studying Data Structures and Separating Classes. *Quant. Struct.-Act. Relat.* **1993**, *12*, 397−401.
(19) Blomberg, N.; Gadboulline R. R.; Niges, M.; Wade R. C. Classification of Protein Sequences by Homology Modeling and Quantitative Analysis of Electrostatic Similarity. *PROTEINS: Struct. Funct. Genet.* **1999**, *37*, 379−387.
(20) Broomhead, D. S.; King G. P. Extracting Qualitative Dynamics from Experimental Data. *Physica D* **1986**, *20*, 217−236.
(21) Amat, L.; Carbo', R.; Ponec, R. Simple Linear QSAR Models Based on Quantum Similarity Measures. *J. Med. Chem.* **1999**, *42*, 5169−5180.
(22) Giuliani, A.; Colosimo, A.; Benigni, R.; Zbilut, J. P. On the constructive role of noise in spatial systems. *Phys. Lett. A* **1998**, *247*, 47−52.

JM0009134