ELSEVIER

A molecular field-based similarity approach to pharmacophoric pattern recognition

Jordi Mestres,* Douglas C. Rohrer, and Gerald M. Maggiora

Computer-Aided Drug Discovery, Pharmacia & Upjohn, Inc., Kalamazoo, Michigan

The use of molecular field-based similarity approaches for obtaining quality molecular alignments and for identifying field-based patterns in bioactive molecules is described. In addition to pairwise similarities, computation of multimolecule similarities affords a means for determining consensus multimolecule alignments. These multimolecule alignments constitute the basis for developing models for the relative binding of bioactive molecules to common protein-binding sites and for the graphical portrayal of molecular field similarity surface plots that identify, visually, molecular regions possessing similar molecular field characteristics. The latter information can then be exploited in the design of molecules that mimic appropriate characteristics of these highly similar steric and electrostatic domains. Regions with low steric and electrostatic similarity in suitably aligned sets of bioactive molecules represent tolerant domains where new structural motifs can be incorporated without significant reductions in activity. To illustrate the potential applicability of the actual molecular field-based similarity approaches to the design of bioactive molecules, a study on a set of HIV-1 protease inhibitors is presented.

Keywords: molecular field similarity, molecular alignment, graphical similarity approaches, field-based pharmacophoric patterns

INTRODUCTION

The search for common features in a series of known active molecules is an important component in the development and optimization of lead compounds in the drug design process. Proper identification of these features characterizes the key molecular structural and electronic patterns required for biological activity, and their particular three-dimensional (3D) arrangement defines a *3D pharmacophore model* for the family of molecules under study[1,2]. Structural information contained in this model can then be used for 3D database searching to identify other molecules containing the hypothesized pharmacophore. In addition, the 3D patterns can serve as guides for structural modifications designed to improve the bioactivity or other appropriate characteristics of the potential drug molecule[3,4].

A variety of computational techniques based on the concept of *molecular similarity* have been developed to identify possible similarities among molecules[5–9]. The main advantage of similarity-based methods is that they deal only with the structures of the effector molecules (e.g., agonists, antagonists, inhibitors) and therefore, in principle, detailed 3D structural information about the receptor site, although desirable, is not required. Thus, similarity-based methods are particularly useful in the initial stages of drug design projects when experimentally determined structural data on receptor sites are scarce. However, similarity-based methods do require that effector molecules with similar features bind in the same region of the receptor site and involve a consistent set of binding interactions. Under this assumption, the relationship between particular features of a molecule and its biological activity can, in many instances, be inferred in terms of its similarity to other molecules possessing biological activity for the same receptor[10–13]. The value of this powerful paradigm has yet to be fully exploited in drug design.

To perform 3D comparisons among molecules, molecular structures need to be aligned by some means[14]. Moreover, as molecules can be superimposed in many different ways, a reliable *molecular alignment* technique must be able to explore a wide range of overlay possibilities and to locate the set of best possible alignments for further analysis. Several atom-based[15,16], feature-based[17–21], shape-based[22–24], and field-based[12,25–35] approaches to molecular similarity have been developed to accomplish this task. They differ mainly in the kind

Color Plates for this article are on pages 103–106.

Address reprint requests to: Jordi Mestres, Institut de Quimica Computacional, Universitat de Girona, 17071 Girona, Catalonia, Spain.

Received 3 December 1996; revised 31 January 1997; accepted 3 February 1997.

* Permanent address: Institut de Quimica Computacional, Universitat de Girona, 17071 Girona, Catalonia, Spain.

655 Avenue of the Americas, New York, NY 10010

1093-3263/97/$17.00
SSDI S0263-7855(97)00003-9

of features considered for quantifying the similarity between molecules. Among the methods, molecular field-based similarity (MFS) approaches appear to be well suited for obtaining relevant molecular alignments, as they are less directly associated with the spatial positions of individual atoms or other molecular features[14]. This is the type of approach utilized in the present work[34,35].

Although a great deal of effort has been focused on the development of quantitative methods for evaluating MFS for drug design purposes, little attention has been given to the development of graphical tools for the visual analysis of field-based similarities and dissimilarities. Along this line, one of the most widely used graphical tools consists of comparative molecular field analysis (CoMFA) coefficient contour maps[12]. Other graphical approaches for identifying common molecular field regions include root mean square (RMS) difference maps[32], surface property mapping onto a sphere by gnomonic projection[36,37], self-organizing maps generated by Kohonen neural networks[38,39], and largest deviation surfaces in electrostatic potentials[40]. A discrete approach to locate maximum MFS loci has also been proposed[34].

An additional benefit from the use of MFS approaches is that, once an internally consistent molecular alignment has been obtained, they provide an ideal basis for constructing graphical images that clarify the nature of the composite MFS (e.g., steric, electrostatic, lipophilic) among molecules. The present work aims to illustrate the potential of *molecular field similarity surfaces* as a useful graphical tool for elucidating, visually, those common molecular field regions contributing highly to the overall similarity and to differentiate these regions from those contributing poorly. Such visual information can then be exploited in the design of novel molecules that mimic the common structural and electronic pharmacophoric field patterns.

METHODOLOGICAL ASPECTS

All calculations reported in this work were performed using the program MIMIC, a recently developed MFS program[34]. MIMIC permits fast and accurate MFS calculations based on a molecular steric volume (MSV) field, a molecular electrostatic potential (MEP) field, or a weighted contribution of the similarities from both fields. An atomic-centered single-Gaussian approximation is used to represent the MSV field[41]. The MEP field is evaluated using the atomic point-charge approximation[42], and the nuclear-smoothing approach proposed by Good et al.[27] is followed to eliminate divergences at nuclear positions, using a three-Gaussian approximation to the 1/r term. Although these are rather crude approximations, they are generally adequate to reproduce qualitatively most of the important steric and electrostatic features of a molecule.

Several alternative ways of evaluating MFSs have been proposed[43–45]. Following the original form defined by Carbó et al.[43], an MFS index between two molecules A and B, S_{AB}, is given by

$$S_{AB} = \frac{\int F_A(\mathbf{r})F_B(\mathbf{r})\,d\mathbf{r}}{[\int F_A(\mathbf{r})F_A(\mathbf{r})\,d\mathbf{r}]^{1/2}[\int F_B(\mathbf{r})F_B(\mathbf{r})\,d\mathbf{r}]^{1/2}} \quad (1)$$

where $F(\mathbf{r})$ is the molecular field used to evaluate the similarity. The nature of the molecular fields considered determines the range of possible S_{AB} values: when positive-definite molecular fields (such as MSV) are used, the range is [0, 1]; when nonpositive-definite molecular fields (such as MEP) are used, the range is [−1, 1]. A value of 1 is achieved only in the limit of molecular identity, whereas any dissimilarity between the two molecules will be reflected in a value less than 1. Because S_{AB} depends on the mutual orientation between the two molecules A and B, an optimization procedure is required to locate molecular alignments that maximize S_{AB}. Furthermore, a variety of molecular alignments are possible, leading to different local similarity maxima[34]. Thus, the exploration of the similarity space produced by comparison of the molecular fields of molecules A and B in the search for multiple alignment solutions requires a large number of S_{AB} evaluations. The use of the analytical form of S_{AB} [Eq. (1)] to evaluate the similarity between molecules together with the molecular field approximations employed significantly accelerate MFS calculations and, ultimately, optimizations.

Molecular field-based similarity indices can be evaluated for specific points in a grid embedding the molecules[45–47]. This allows an assessment of not only the overall MFS between molecules, but also the MFS of particular regions surrounding the superimposed molecules. Thus, the Carbó form of a grid-based MFS index can be calculated by accumulating the similarity contributions at every grid point, $S_{AB}(\mathbf{r}_i)$, to the total molecular field similarity as

$$S_{AB} = \sum_i S_{AB}(\mathbf{r}_i) \quad (2)$$

where

$$S_{AB}(\mathbf{r}_i) = \frac{F_A(\mathbf{r}_i)F_B(\mathbf{r}_i)}{\left[\sum_j F_A(\mathbf{r}_j)F_A(\mathbf{r}_j)\right]^{1/2}\left[\sum_j F_B(\mathbf{r}_j)F_B(\mathbf{r}_j)\right]^{1/2}} \quad (3)$$

and $F(\mathbf{r}_i)$ symbolizes the value of the molecular field at a particular grid point $\mathbf{r}_i$, with summations extending over all grid points. The same boundary conditions mentioned above apply for grid-based cumulative indices.

From Eq. (3) the basic idea of constructing molecular field similarity surfaces follows immediately. The percentage contribution to the total MFS at a particular grid point $\mathbf{r}_i$, $C_{AB}(\mathbf{r}_i)$, is given by

$$C_{AB}(\mathbf{r}_i) = \frac{S_{AB}(\mathbf{r}_i)}{S_{AB}} \times 100 \quad (4)$$

By representing the $C_{AB}(\mathbf{r}_i)$ values at each of the points on a surface surrounding the superimposed molecules, it is possible to identify molecular regions where the contribution to the total MFS is higher or lower than a user-specified value. Alternatively, isosimilarity surface plots can be used to represent molecular regions contributing a given percentage to the total MFS. The use of these molecular field similarity surfaces allows visual recognition of common steric and electrostatic field pharmacophoric patterns among sets of active and inactive molecules.

APPLICATION TO HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 PROTEASE INHIBITORS

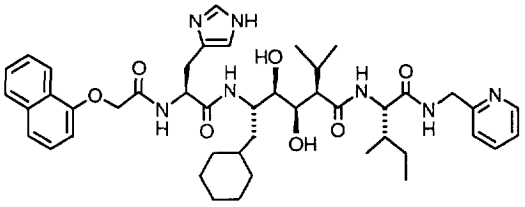
A study of a set of human immunodeficiency virus type 1 (HIV-1) protease inhibitors[48] was conducted to illustrate the potential applicability of the graphical MFS approach. HIV-1 protease inhibitors have been the subject of several similarity studies using feature-based[21], CoMFA[49–51], and field-based[31–33,52] approaches. In the present work, four peptidic inhibitors belonging to structurally diverse classes were selected (Scheme 1), namely, a dihydroxyethylene isostere (PNU-75875)[53], a member of the sulfonamide class (VX-478)[54], a hydroxyethylamine isostere (JG-365)[55], and a pseudo-symmetric member of the hydroxyethylbenzamide class (A-74704)[56]. To focus the discussion purely on similarity issues and not on conformational issues, molecular structures were taken directly from the crystal structures of the inhibitors bound at the active site of the HIV-1 protease[57]. The Protein Data Bank (PDB) codes[58] for PNU-75875, VX-478, JG-365, and A-74704 are 1HIV, 1HPV, 7HVP, and 9HVP, respectively. Throughout this work, PDB codes are used to designate the respective inhibitors. In the next sections, the strategy used to obtain a *consensus molecular alignment* for the set of four inhibitors is presented, followed by a *visual postalignment analysis* of the steric and electrostatic field similarities based on the use of molecular field similarity surfaces.

Molecular field alignments

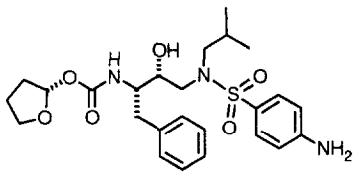
The systematic search algorithm implemented in MIMIC is used to explore different overlay possibilities between pairs of molecules and to locate the set of best possible alignments[34]. Basically, one molecule is kept fixed (the reference molecule) while the other (the adapting molecule) is systematically placed in a number of unique starting orientations about the reference molecule. The MFS between the two molecules is then optimized for each of the starting orientations, using common gradient-seeking techniques. This procedure ensures a wide and uniformly distributed exploration of the similarity space induced by the molecular fields of the two molecules[34]. It has been previously shown that field-weighted similarities help compensate for strong field-based dependencies of the molecular alignments[34]. Accordingly, throughout this work the similarity between molecules will be evaluated using a 2:1 weighting of the steric and the electrostatic field similarities, respectively.

The five best molecular alignments obtained from the explorations of the similarity spaces induced by the six pairs of inhibitors are gathered in Table 1. Two interesting features of Table 1 are worth commenting on. First, in five of six cases, the best alignment solution is in excellent agreement with the "experimental" inhibitor alignment (marked in boldface), obtained by superimposing the C_α carbons of the respective HIV-1 protease structures to which the inhibitors are bound. In the one case in which the best alignment does not correspond to the experimental one, {1HIV, 9HVP}, the second-best solution matches the experimental alignment, although both similarity values are close. The second interesting result is that there is a clear gap between the similarity values obtained for the best and second-best alignment solutions and for the rest of the alignments. This indicates that all pairs of inhibitors seem to have strong preferences for essentially two relative orientations. This is not too surprising given their elongated shape and C_2 pseudosymmetric character, which, in this particular case, helps reduce the number of relevant alignments. The series of best and second-best alignments obtained for three selected pairs of inhibitors, {1HIV, 1HPV}, {1HIV, 9HVP}, and {9HVP, 1HPV}, are shown in Figure 1.

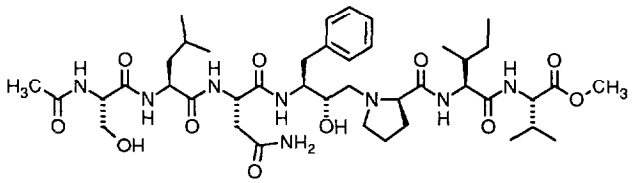
The notion of *alignment consistency* arises when trying to



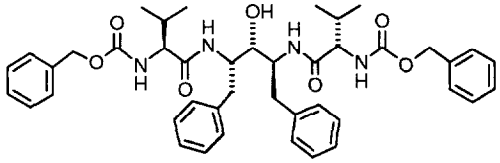
PNU-75875 (1HIV)



VX-478 (1HPV)



JG-365 (7HVP)



A-74704 (9HVP)

Scheme 1

Table 1. Similarity values corresponding to the five best pairwise molecular alignments obtained from the explorations of the similarity spaces induced by the six pairs of molecules.[a]

Molecular pair	1	2	3	4	5
{1HIV, 1HPV}	**0.4315**	0.4152	0.3505	0.3335	0.3294
{1HIV, 7HVP}	**0.6161**	0.3981	0.3836	0.3827	0.3736
{1HIV, 9HVP}	0.5691	**0.5643**	0.3746	0.3695	0.3631
{7HVP, 1HPV}	**0.4468**	0.4048	0.3599	0.3286	0.3258
{9HVP, 1HPV}	**0.4893**	0.4610	0.3645	0.3269	0.3267
{9HVP, 7HVP}	**0.5808**	0.5676	0.3749	0.3740	0.3704

[a] Alignments in agreement with the relative binding orientations obtained experimentally are presented in boldface.

extrapolate the alignment of a set of more than two molecules by directly superimposing the different pairwise alignment solutions. To understand this concept consider, for example, three molecules, *A*, *B*, and *C*, and select *A* as the reference molecule and *B* and *C* as the adapting molecules. The direct superimposition of *B* and *C* at the optimized best alignment solutions of the pairs {A, B} and {A, C} may not correspond to the best alignment solution of the pair {B, C}. An example of this situation is illustrated in Figure 1. Taking the best alignment solutions of the pairs {1HIV, 1HPV} and {1HIV, 9HVP}, the direct superimposition of 1HPV and 9HVP does not agree with the best alignment solution of the pair {9HVP, 1HPV} but with the second-best alignment solution. Instead, taking the best alignment solution of the pair {1HIV, 1HPV} and the second-best alignment solution of {1HIV, 9HVP}, the direct superimposition of 1HPV and 9HVP is in good agreement with the best alignment obtained for the pair {9HVP, 1HPV}. Thus, it is evident that, when dealing with a set of molecules, some quantitative measure beyond the pairwise similarities is needed to derive *consensus molecular alignments*, i.e., the set of best possible multimolecule alignments.

A consensus MFS accounting for multimolecule relationships among a set of *M* molecules, $S_{\{M\}}$, can be evaluated by averaging all possible pairwise similarities between the molecules in the set as

$$S_{\{M\}} = \frac{\sum_{I>J} S_{IJ}}{(M^2 - M)/2} \qquad (5)$$

Initially, molecules are superimposed according to the alignment solutions obtained from the pairwise similarity space explorations (Table 1). Each direct superimposition is taken as the starting orientation from which the alignment of all molecules is then optimized. In principle, all pairwise alignment solutions should be considered. However, in practice, it is reasonable to expect combinations of the best pairwise alignment solutions to lead to the best multimolecule alignments.

Table 2 collects the five best consensus multimolecule alignments obtained for the four possible sets of three inhibitors and the final set of four inhibitors. The pairwise alignments used for the starting orientation of the multimolecule alignments are indicated in parentheses. As is observed, in all three-molecule cases the four best alignments come from the superimposition of combinations of the best and second-best pairwise alignments. For example, for the set {1HIV, 1HPV, 9HVP} the best consensus three-molecule alignment (with an MFS value of 0.4937) was obtained from the best and second-best pairwise alignments of the {1HIV, 1HPV} and {1HIV, 9HVP} pairs, respectively (see Figure 1). In all cases, the best three-molecule alignments are in agreement with the experimentally derived alignment (indicated in boldface in Table 2). Thus, it is possible to identify the "correct" three-molecule alignments even though originally the highest ranked similarity value for the {1HIV, 9HVP} pair was not in agreement with the experimentally observed relative binding orientation. Finally, the best alignment obtained for all four molecules also is in agreement with the experimental result[59].

Obtaining quality alignments has always been one of the major concerns of MFS approaches. For instance, it is well established that the success of a CoMFA study is highly dependent on the quality of the "alignment rule" or choice of superimposition of the molecules under study[49,50]. The experimentally observed relative binding orientation of the inhibitors and the best relative binding model obtained using the present MFS approaches are depicted in Color Plate 1. As is seen, the two multimolecule alignments are in excellent agreement with one another (the RMS difference is 0.368 Å). Thus, the use of consensus MFSs appears to be a robust procedure for assessing the quality of multimolecule alignments. Other MFS approaches can then benefit from these alignments by using them as the basis for their postalignment analyses.

Molecular field similarity surfaces

Once a consensus molecular alignment has been derived, grid-based cumulative MFS indices can be used to construct molecular field similarity surface plots. The grid used in these similarity calculations was defined by extending its limits 5 Å away from the common Cartesian coordinates of the ensemble of the four inhibitors. Strictly for esthetic reasons, the grid spacing of the dot-surface plots was set to 0.25 Å. Two different approaches can be followed for constructing similarity surfaces. For positive-definite fields such as the MSV, it may be more convenient to construct a plot by representing those points on a given surface (for example, the surface obtained by the union of the van der Waals molecular surfaces in the alignment) contributing more (colored in green) or less (colored in yellow) than a user-specified percentage to the total MFS value. For nonpositive-definite fields such as the MEP, isosimilarity surface plots may be preferable. Such plots are obtained by representing those points making a positive (colored in red or blue depending on whether the similarity arises

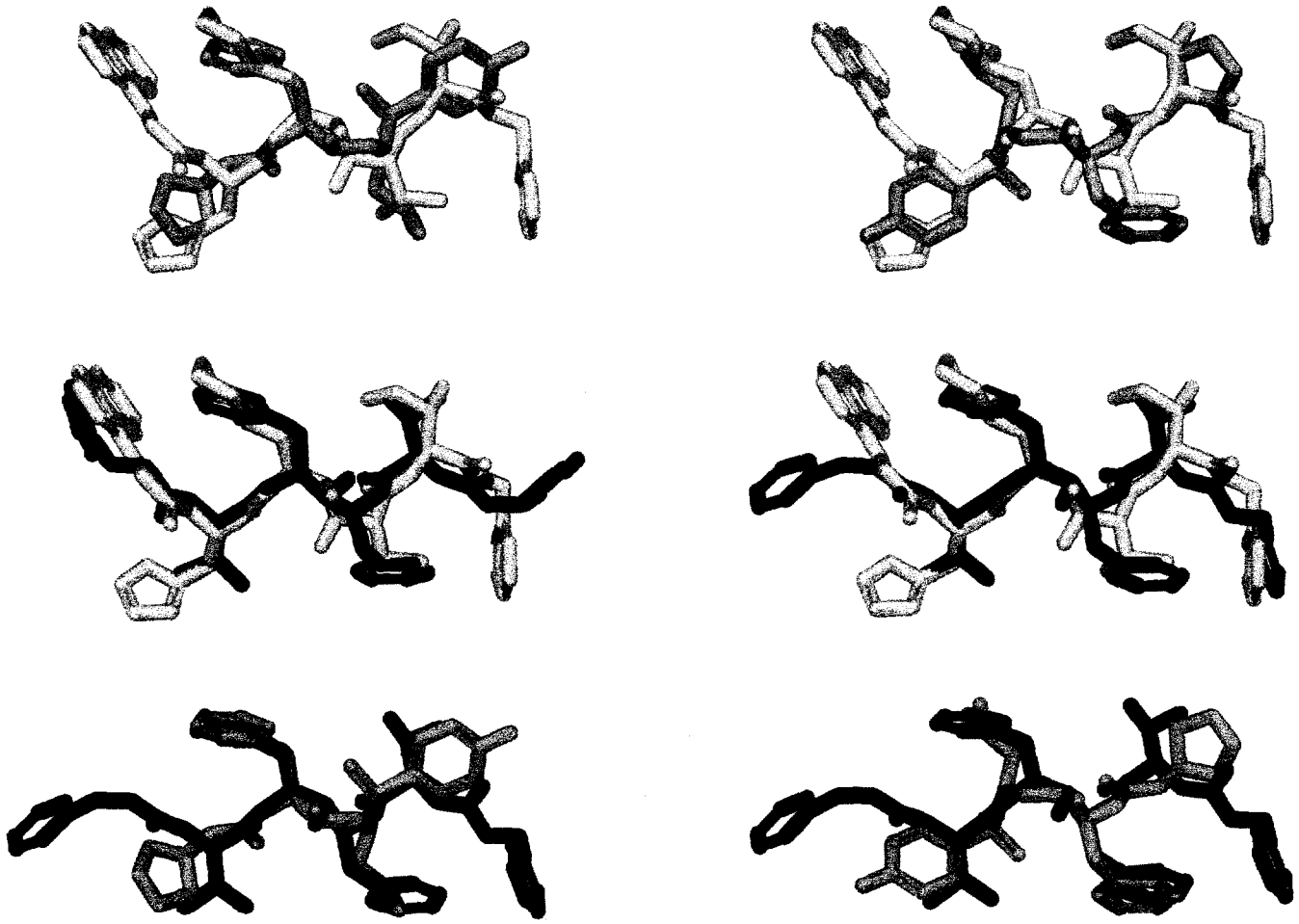


Figure 1. Best (left) and second-best (right) pairwise alignments for three selected pairs of inhibitors (see Table 1). The inhibitor pairs are (from top to bottom): {1HIV, 1HPV}, {1HIV, 9HVP}, {9HVP, 1HPV}. 1HIV is in white, 1HPV in gray, and 9HVP in black.

primarily from similar negative or positive MEP regions, respectively) or negative (colored in yellow) contribution to the total MFS, relative to a given user-specified percentage value.

The steric and electrostatic field similarity surfaces obtained for the best relative binding model derived from the set of four HIV-1 protease inhibitors are presented in Color Plate 2. The steric field similarity plot on the union van der Waals surface embedding the four inhibitors clearly identifies regions of high steric similarity in the P1, P2, P1′, and P2′ subsites, whereas low steric similarity is preferentially located in the P3, P4, P3′, and P4′ subsites[60]. A much more careful examination is required for analyzing the information contained in the electrostatic field isosimilarity surface plot. Obviously, the selection of different grid-point percentage contributions to the total similarity gives rise to different isosurfaces. Thus, some field-based similarity features appear to be emphasized depending

Table 2. Similarity values corresponding to the five best consensus molecular alignments obtained for the four sets of three molecules and the final set of four molecules[a]

Molecular set	1′		2′		3′		4′		5′	
{1HIV, 1HPV, 7HVP}	**0.4949**	**(1, 1)**	0.4725	(2, 1)	0.4111	(2, 2)	0.4070	(1, 2)	0.4062	(3, 1)
{1HIV, 1HPV, 9HVP}	**0.4937**	**(1, 2)**	0.4831	(2, 1)	0.4829	(1, 1)	0.4769	(2, 2)	0.4038	(4, 2)
{1HIV, 7HVP, 9HVP}	**0.5858**	**(1, 2)**	0.5833	(1, 1)	0.5137	(2, 1)	0.5090	(2, 2)	0.4483	(1, 5)
{9HVP, 1HPV, 7HVP}	**0.5029**	**(1, 1)**	0.4893	(2, 2)	0.4850	(1, 2)	0.4815	(2, 1)	0.4138	(3, 1)
{1HIV, 1HPV, 7HVP, 9HVP}	**0.5184**	**(1, 1, 2)**	0.5114	(1, 1, 1)	0.5042	(2, 1, 1)	0.5028	(2, 1, 2)	0.4750	(2, 2, 1)

[a] The numbers in parentheses refer to the pairwise alignments (with respect to the first molecule in the list) used to obtain the initial multiple-molecule superimposition (see Table 1). Alignments in agreement with the relative binding orientation obtained experimentally are presented in boldface.

on the percentage contribution selected. For instance, higher percentage contribution isosimilarity surfaces (|0.01|%) essentially contain regions of negative-MEP similarity. These regions are concentrated in the two common sets of carbonyl groups pointing towards Ile-50 and Ile-50′, the common hydroxyl group making a hydrogen bond with Asp-25 and Asp-25′, the two sets of carbonyl groups interacting with Asp-29 and Asp-29′ (not present in 1HPV), and around the groups overlapping in the P1 and P2 subsites. However, when a lower percentage contribution is selected (|0.001|%) a wider variety of similar electrostatic features is encountered. The resultant isosimilarity surface is presented in Color Plate 2. Those highly similar negative-MEP regions previously identified remain visible in the plot but other electrostatic regions contributing to the similarity at a lower level become apparent. For instance, although some negative-MEP similarity (in red) is present in the P1′ subsite, its electrostatic requirements do not seem as constrained as in the P1 subsite. Also, a positive-MEP similarity region (in blue) is identified in the P2′ subsite. Finally, electrostatically diverse regions (in yellow) are found primarily in the most external P3, P4, P3′, and P4′ subsites.

These molecular field similarity surfaces were then placed within the binding site of the HIV-1 protease to provide a more visual sense of the steric and electrostatic requirements of the set of inhibitors examined in this study. The results are shown in Color Plate 3. As can be observed, the yellow-colored surface regions are located predominantly in those subsites of the inhibitors that have minimal interaction with the steric and electrostatic characteristics of the receptor site. This visual information provides an illustrative example of the potential applicability of MFS approaches in drug design.

In summary, essentially two different types of insights can be obtained from the construction of molecular field similarity surfaces. On the one hand, they permit visual identification of molecular regions where the local field-based similarity contributes significantly to the total field-based similarity. For a group of active molecules, regions where a high contribution to the total steric field similarity is found can be associated with sterically constrained regions, which are likely to be buried deep within the binding site. Accordingly, regions contributing preferentially to the total electrostatic field similarity can be associated with sites where electrostatic interactions are of key importance for binding. These common molecular field regions constitute what can be referred as *field-based pharmacophoric patterns.*

On the other hand, those molecular regions where the local field-based similarity contributes weakly, in the case of positive-definite fields, or negatively, in the case of nonpositive-definite fields, to the total field-based similarity are also revealed by this method. For a group of active molecules, these regions may be associated with less restrictive subsites of the binding site, possibly reflecting potential protein conformational flexibility or simply being adjacent to the protein exterior. As these sites do not give rise to dominant ligand–protein interactions, they do not play a critical role in the biological activity of the molecules.

Both of these aspects provide the ability to use molecular field similarity surfaces for constructing a unique composite similarity picture, capable of visually identifying consensus and nonconsensus regions among the molecular fields described by an appropriate series of active or inactive molecules. These graphical tools represent a basis from which relationships between the structural characteristics of inhibitors and their biological activities, or other relevant properties, can be inferred. These tools are especially useful when structural information on the protein receptor site is not available. In these cases, the drug design process can proceed by maximally exploiting the information contained in the structures of the effector molecules themselves.

CONCLUSIONS

As exemplified by the results presented in this work, recognition of field-based pharmacophoric patterns can be accomplished using similarity approaches based on the molecular field properties of molecules. However, a compulsory first step is the construction of good-quality molecular alignments. For this purpose, the use of consensus molecular field similarities appears to be a robust procedure for assessing quality multimolecule alignments and for deriving satisfactory relative binding models. In a second stage, postalignment visual analyses can be performed using molecular field similarity surfaces as a means for gaining a better understanding of the local steric and electrostatic field similarities among molecules. Drug designers can benefit from the visual insights provided by these molecular similarity approaches by designing novel active molecules that mimic the identified highly similar steric and electrostatic patterns, while being more permissive with those structural and electrostatic motifs located in low-similarity regions.

ACKNOWLEDGMENTS

We thank Joe Moon for his assistance in the graphical representation of the dot-surface plots. J.M. acknowledges the financial support provided by the Spanish Ministerio de Educacion y Ciencia (reference no. PF95-40314447).

REFERENCES

1 Golender, V.E. and Vorpagel, E.R. Computer-assisted pharmacophore identification. In: *3D QSAR in Drug Design: Theory, Methods, and Applications* (H. Kubinyi, ed.). Escom Science, Leiden, 1993, pp. 137–149

2 Liljefors, T. and Pettersson, I. Computer-aided development of three-dimensional pharmacophore models. In: *A Textbook of Drug Design and Development* (P. Krogsgaard-Larsen, T. Liljefors, and U. Madsen, eds.). Harwood Academic, Amsterdam, 1996, pp. 60–93

3 Martin, Y.C. 3D database searching in drug design. *J. Med. Chem.* 1992, **35,** 2145–2154

4 Willett, P. Scarching for pharmacophoric patterns in databases of three-dimensional chemical structures. *J. Mol. Recogn.* 1995, **8,** 290–303

5 Johnson, M.A. and Maggiora, G.M. (eds.). *Concepts and Applications of Molecular Similarity.* Wiley, New York, 1990

6 Kubinyi, H. (ed.). *3D QSAR in Drug Design: Theory, Methods, and Applications.* Escom Science, Leiden, 1993

7 Sen, K. (ed.). *Topics in Current Chemistry: Molecular Similarity I,* Vol. 173. Springer-Verlag, Berlin, 1995

8 Dean, P.M. (ed.). *Molecular Similarity in Drug Design.* Blackie Academic, London, 1995
9 Carbó, R. (ed.). *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches.* Kluwer, Amsterdam, 1995
10 Van de Waterbeemd, H. (ed.). *Structure–Property Correlations in Drug Research.* Academic Press, London, 1996
11 Hopfinger, A.J. A QSAR investigation of dihydrofolate reductase inhibition by Baker triazines based upon molecular shape analysis. *J. Am. Chem. Soc.* 1980, **102,** 7196–7206
12 Cramer, R.D., III, Patterson, D.E., and Bunce, J.D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 1988, **110,** 5959–5967
13 Good, A.C., So, S.-S., and Richards, W.G. Structure–activity relationships from molecular similarity matrices. *J. Med. Chem.* 1993, **36,** 433
14 Klebe, G. Structural alignment of molecules. In: *3D QSAR in Drug Design. Theory, Methods, and Applications* (H. Kubinyi, ed.). Escom Science, Leiden, 1993, pp. 173–199
15 Jakes, S.E. and Willett, P. Pharmacophoric pattern matching in files of three-dimensional chemical structures. Selection of interatomic distance screens. *J. Mol. Graphics* 1986, **4,** 12–20
16 Perkins, T.D.J. and Dean, P.M. An exploration of a novel strategy for superposing several flexible molecules. *J. Comput.-Aided Mol. Design* 1993, **7,** 155–172
17 Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A., and Dunn, D.A. The conformational parameter in drug design: The active analogue approach. In: *Computer-Assisted Drug Design* (E.C. Olson and R.E. Christoffersen, eds.), Vol. 112. American Chemical Society, Washington, D.C., 1979, pp. 205–222
18 Mayer, D., Naylor, C.B., Motoc, I., and Marshall, G.R. A unique geometry of the active site of angiotensin-converting enzyme consistent with structure–activity studies. *J. Comput.-Aided Mol. Design* 1987, **1,** 3–16
19 Van Drie, J.H., Weininger, D., and Martin, Y.C. ALADDIN: An integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructure searching of three-dimensional molecular structures. *J. Comput.-Aided Mol. Design* 1989, **3,** 225–251
20 Jones, G., Willett, P., and Glen, R.C. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput.-Aided Mol. Design* 1995, **9,** 532–549
21 Barnum, D., Greene, J., Smellie, A., and Sprague, P. Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.* 1996, **36,** 563–571
22 Chau, P.L. and Dean, P.M. Molecular recognition: 3D surface structure comparison by gnomonic projection. *J. Mol. Graphics* 1987, **5,** 97–100
23 Tokarski, J.S. and Hopfinger, A.J. Three-dimensional molecular shape analysis—quantitative structure–activity relationship of a series of cholecystokinin-A receptor antagonists. *J. Med. Chem.* 1994, **37,** 3639–3654
24 Jain, A.N., Dietterich, T.G., Lathrop, R.H., Chapman, D., Critchlow, R.E., Jr., Bauer, B.E., Webster, T.A., and Lozano-Perez, T. Compass: A shape-based machine learning tool for drug design. *J. Comput.-Aided Mol. Design* 1994, **8,** 635–652
25 Kearsley, S.K. and Smith, G.M. An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. *Tetrahedron Comput. Methods* 1990, **3,** 615–633
26 Hermann, R.B. and Herron, D.K. OVID and SUPER: Two overlap programs for drug design. *J. Comput.-Aided Mol. Design* 1991, **5,** 511–524
27 Good, A.C., Hodgkin, E.E., and Richards, W.G. Utilization of Gaussian functions for the rapid evaluation of molecular similarity. *J. Chem. Inf. Comput. Sci.* 1992, **32,** 188–191
28 Sanz, F., Manaut, F., Rodríguez, J., Lozoya, E., and López-de-Briñas, E. MEPSIM: A computational package for analysis and comparison of molecular electrostatic potentials. *J. Comput.-Aided Mol. Design* 1993, **7,** 337–347
29 Klebe, G., Mietzner, T., and Weber, F. Different approaches toward an automatic structural alignment of drug molecules: Applications to sterol mimics, thrombin and thermolysin inhibitors. *J. Comput.-Aided Mol. Design* 1994, **8,** 751–778
30 Apaya, R.P., Lucchese, B., Price, S.L., and Vinter, J.G. The matching of electrostatic extrema: A useful method in drug design? A study of phosphodiesterase III inhibitors. *J. Comput.-Aided Mol. Design* 1995, **9,** 33–43
31 McMartin, C. and Bohacek, R.S. Flexible matching of test ligands to a 3D pharmacophore using a molecular superposition force field: Comparison of predicted and experimental conformations of inhibitors in three enzymes. *J. Comput.-Aided Mol. Design* 1995, **9,** 237–250
32 Perkins, T.D.J., Mills, J.E.J., and Dean, P.M. Molecular surface–volume and property matching to superpose flexible dissimilar molecules. *J. Comput.-Aided Mol. Design* 1995, **9,** 479–490
33 Grant, J.A., Gallardo, M.A., and Pickup, B.T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* 1996, **17,** 1653–1666
34 Mestres, J., Rohrer, D.C., and Maggiora, G.M. MIMIC: A molecular-field matching program. Exploiting the applicability of molecular similarity approaches. *J. Comput. Chem.* 1997, **18,** 934–954
35 Mestres, J., Rohrer, D.C., and Maggiora, G.M. Submitted (1997)
36 Blaney, F., Finn, P., Phippen, R., and Wyatt, M. Molecular surface comparison: Application to drug design. *J. Mol. Graphics* 1993, **11,** 98–105
37 Blaney, F.E., Edge, C., and Phippen, R.W. Molecular surface comparison. 2. Similarity of electrostatic vector fields in drug design. *J. Mol. Graphics* 1995, **13,** 165–174
38 Gasteiger, J. and Li, X. Mapping the electrostatic potential of muscarinic and nicotinic agonists with artificial neural networks. *Angew. Chem. Int. Ed. Engl.* 1994, **33,** 643–646
39 Barlow, T.W. Self-organizing maps and molecular similarity. *J. Mol. Graphics* 1995, **13,** 24–27
40 Beck, B., Glen, R.C., and Clark, T. The inhibition of α-chymotrypsin predicted using theoretically derived molecular properties. *J. Mol. Graphics* 1996, **14,** 130–135

41 Rohrer, D.C. 3D molecular similarity modelling in computational drug design. In: *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches* (R. Carbó, ed.). Kluwer, Amsterdam, 1995, pp. 141–161
42 Mulliken atomic charges were obtained using MOPAC at the AM1 semiempirical level: Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 1985, **107,** 3902–3909
43 Carbó, R., Leyda, L., and Arnau, M. How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *Int. J. Quantum Chem.* 1980, **17,** 1185–1189
44 Hodgkin, E.E. and Richards, W.G. Molecular similarity based on electrostatic potential and electric field. *Int. J. Quantum Chem. Quantum Biol. Symp.* 1987, **14,** 105–110
45 Petke, J. Cumulative and discrete similarity analysis of electrostatic potentials and fields. *J. Comput. Chem.* 1993, **14,** 928–933
46 Good, A.C. The calculation of molecular similarity: Alternative formulas, data manipulation and graphical display. *J. Mol. Graphics* 1992, **10,** 144–151
47 Reynolds, C.A., Burt, C., and Richards, W.G. A linear molecular similarity index. *Quant. Struct.-Act. Relat.* 1992, **11,** 34–35
48 Kempf, D.J. and Sham, H.L. HIV protease inhibitors. *Curr. Pharm. Design* 1996, **2,** 225–246
49 Waller, C.L., Oprea, T.I., Giolitti, A., and Marshall, G.R. Three-dimensional QSAR of human immunodeficiency virus (I) protease inhibitors. 1. A CoMFA study employing experimentally-determined alignment rules. *J. Med. Chem.* 1993, **36,** 4152–4160
50 Oprea, T.I., Waller, C.L., and Marshall, G.R. Three-dimensional quantitative structure–activity relationship of human immunodeficiency virus (I) protease inhibitors. 2. Predictive power using limited exploration of alternate binding modes. *J. Med. Chem.* 1994, **37,** 2206–2215
51 Oprea, T.I., Waller, C.L., and Marshall, G.R. 3D-QSAR of human immunodeficiency virus (I) protease inhibitors. 3. Interpretation of CoMFA results. *Drug Design Discov.* 1994, **12,** 29–51
52 Rohrer, D.C. 3D molecular similarity methods: In search of a pharmacophore. In: *Experimental and Computational Approaches to Structure-Based Drug Design* (P.W. Codding, ed.). Kluwer, Amsterdam, 1997 (in press)
53 Thanki, N., Rao, J.K.M., Foundling, S.I., Howe, W.J., Moon, J.B., Hui, J.O., Tomasselli, A.G., Heinrikson, R.L., Thaisrivongs, S., and Wlodawer, A. Crystal structure of a complex of HIV-1 protease with a dihydroxyethylene-containing inhibitor: Comparisons with molecular modeling. *Protein Sci.* 1992, **1,** 1061–1072
54 Kim, E.E., Baker, C.T., Dwyer, M.D., Murcko, M.A., Rao, B.G., Tung, R.D., and Navia, M.A. Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* 1995, **117,** 1181–1182
55 Swain, A.L., Miller, M.M., Green, J., Rich, D.H., Schneider, J., Kent, S.B.H., and Wlodawer, A. X-Ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 1990, **87,** 8805–8809
56 Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A., and Knigge, M. Design, activity, and 2.8 Å crystal structure of a C_2 symmetric inhibitor complexed to HIV-1 protease. *Science* 1990, **249,** 527–533
57 For a discussion on the comparative ability of molecular field similarity approaches to assess the correct relative binding model when using crystal structures directly or when dealing with a set of representative structures derived from conformational analysis, see Ref. 35
58 Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanoushi, T., and Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112,** 535–542
59 Note that, as the value of $S_{\{M\}}$ is optimized and, consequently, the relative orientation of all molecules is rearranged, the final best multimolecule alignment should not be dependent on the molecule selected as the reference molecule
60 Binding pocket nomenclature according to: Schecter, I. and Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 1967, **27,** 157–162