



A molecular-field-based similarity study of non-nucleoside HIV-1 reverse transcriptase inhibitors

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

Computer-Aided Drug Discovery, Pharmacia & Upjohn Inc., Kalamazoo, MI 49001, U.S.A.

Received 18 May 1998; Accepted 17 August 1998

Key words: molecular alignments, molecular-field similarity, non-nucleoside HIV-1 reverse transcriptase inhibitors, pharmacophore patterns, protein structure alignments

Summary

This article describes a molecular-field-based similarity method for aligning molecules by matching their steric and electrostatic fields and an application of the method to the alignment of three structurally diverse non-nucleoside HIV-1 reverse transcriptase inhibitors. A brief description of the method, as implemented in the program MIMIC, is presented, including a discussion of pairwise and multi-molecule similarity-based matching. The application provides an example that illustrates how relative binding orientations of molecules can be determined in the absence of detailed structural information on their target protein. In the particular system studied here, availability of the X-ray crystal structures of the respective ligand–protein complexes provides a means for constructing an ‘experimental model’ of the relative binding orientations of the three inhibitors. The experimental model is derived by using MIMIC to align the steric fields of the three protein P66 subunit main chains, producing an overlay with a 1.41 Å average rms distance between the corresponding C_α ’s in the three chains. The inter-chain residue similarities for the backbone structures show that the main-chain conformations are conserved in the region of the inhibitor-binding site, with the major deviations located primarily in the ‘finger’ and RNase H regions. The resulting inhibitor structure overlay provides an experimental-based model that can be used to evaluate the quality of the direct a priori inhibitor alignment obtained using MIMIC. It is found that the ‘best’ pairwise alignments do not always correspond to the experimental model alignments. Therefore, simply combining the best pairwise alignments will not necessarily produce the optimal multi-molecule alignment. However, the best simultaneous three-molecule alignment was found to reproduce the experimental inhibitor alignment model. A pairwise consistency index has been derived which gauges the quality of combining the pairwise alignments and aids in efficiently forming the optimal multi-molecule alignment analysis. Two post-alignment procedures are described that provide information on feature-based and field-based pharmacophoric patterns. The former corresponds to traditional pharmacophore models and is derived from the contribution of individual atoms to the total similarity. The latter is based on molecular regions rather than atoms and is constructed by computing the percent contribution to the similarity of individual points in a regular lattice surrounding the molecules, which when contoured and colored visually depict regions of highly conserved similarity. A discussion of how the information provided by each of the procedures is useful in drug design is also presented.

Introduction

The relationship of a molecule’s structure to its biological activity has played a critical role in many aspects of the design of new drug molecules [1].

In recent years the availability of detailed, three-dimensional (3D) structural information of proteins and their complexes with various ligands has revolutionized the drug-design process. Nevertheless, in most instances such information is still unavailable, and the design process must rely upon a more indirect

*Present address: Department of Molecular Design & Informatics, N.V. Organon, 5340 BH Oss, The Netherlands.

**To whom correspondence should be addressed.

approach. In such cases, identification of the ‘molecular characteristics’ that give rise to biological activity is of paramount importance. These ‘characteristics’ can be represented at different levels by *molecular properties* (such as Log *P*, molar refractivity, volume, and surface charge to name a few), *molecular features* (hydrogen bond donors and acceptors, and charged, polar, aromatic and hydrophobic groups), or *molecular fields* (steric, electrostatic, and hydrophobic). The presence and relative positioning of mutual structural features within sets of active molecules, in fact, constitute a *pharmacophore*, a conceptual structure that has a distinguished history in drug design [2,3]. In the absence of detailed 3D structural information of the target protein, a pharmacophore provides useful, albeit greatly simplified, structural information that can be of significant assistance in drug design. Inferring a pharmacophoric pattern from a set of bioactive molecules is, however, non-trivial, and many different approaches aiming at identifying similarities among molecules have been developed [4,5].

To address this problem, the present work uses a molecular-field-based similarity procedure, as embodied in the program MIMIC. A number of other molecular-field matching methods have been developed over the last few years [6–12]. The main differences among them are the type of molecular fields they use and the ways they approach molecular-field-based similarities. The MIMIC approach aligns two or more molecules by matching their steric and electrostatic fields [13,14]. The degree of field matching is taken as a measure of the similarity of the molecules being aligned. Field-based methods are generally less sensitive to the specific structural features of molecules than are methods that rely upon the correspondence of specific atomic and molecular features. Thus, in the search for optimal alignments when the molecules under study are structurally quite diverse, field-based methods tend to be much better than a traditional atom-based approach [15].

Besides the assessment of quality relative binding orientations, identification of specific characteristics common to the set of aligned molecules (pharmacophoric patterns) is straightforward within the MIMIC approach. Both feature-based and field-based pharmacophoric patterns can be derived. The former, closer to the classical pharmacophore concept, is assessed through the calculation of *maximum similarity loci* [6], whereas the latter is based on the definition of *molecular-field similarity surfaces* [7] which graphically portray those regions surrounding the molecules

showing low, medium, and high similarity between the molecular fields. The ensemble of alignment and post-alignment analyses currently available enhance the capabilities of similarity-based methods for expediting the molecular design process, as will be illustrated in the remainder of the paper.

A molecular-field-based similarity study of three non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs) is presented. The NNRTI family is comprised of a set of compounds exhibiting substantial structural diversity [16]. In addition, experimental crystal structures [17–19] of these inhibitors complexed to the protein receptor site are available from the Brookhaven Protein Databank (PDB) [20]. These two facts make the NNRTI family especially attractive as a means for testing the suitability of the MIMIC approach for, on the one hand, constructing relative binding models from the structures of the inhibitors alone and, on the other hand, revealing those feature-based and field-based pharmacophoric patterns that may be responsible for the activity of these compounds. The structures of the inhibitors selected in the present study, namely, Nevirapine (NEVI), the TIBO analog R86183 (TIBO), and the α -APA analog R95845 (α APA), are depicted in Table 1 along with some related information.

Methodological aspects

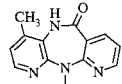
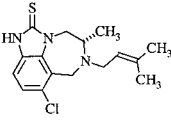
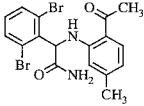
Molecular fields

A brief description of the salient features of the molecular-field-based similarity method used in the current work is provided below. Additional details can be found in our earlier works [7,24]. All the calculations described here are carried out with the program MIMIC. Two molecular fields are employed in MIMIC. One describes the approximate size and shape of a molecule by its *molecular steric volume* (MSV) field. The other describes the approximate electrostatic potential of the molecule by its *molecular electrostatic potential* (MEP) field. Other fields, such as those describing lipophilicity [25], can also be used.

At a point \mathbf{r} in space, the MSV field of molecule *A* is given by

$$F_A^{\text{MSV}}(\mathbf{r}) = \sum_{i \in A} f_i^{\text{MSV}}(\mathbf{r}) \quad (1)$$

Table 1. Non-nucleoside HIV-1 reverse transcriptase inhibitors investigated in this work

Structure	Common name (abbreviation)	Name	PDB [20] reference
	Nevirapine (NEVI)	11-Cyclopropyl-5-11-dihydro-4-methyl-6H-dipyrido-[3,2-b,2',3'-e][1,4]-diazepin-6-one [21]	3HVT [17]
	Tetrahydro-imidazo-benzodiazepinone (TIBO)	(+)-(S)-4,5,6,7-Tetrahydro-5-methyl-6-(3-methyl-2-butenyl)imidazo-8-chloro-[4,5,1-jk][1,4]-benzodiazepin-2(1H)-thione [22]	1HNV [18]
	α -Anilino-phenyl-acetamide (α APA)	(-)- α -((2-Acetyl-5-methylphenyl)amino)-2,6-dibromobenzeneacetamide [23]	1HNI [19]

where f_i^{MSV} defines the steric volume of atom i and is represented by a Gaussian function located at \mathbf{R}_i , the center of atom i :

$$f_i^{\text{MSV}}(\mathbf{r}) = \alpha_i \exp(-\beta_i |\mathbf{r} - \mathbf{R}_i|^2) \quad (2)$$

The coefficients α_i and the exponent β_i are optimized for each atom type [24]. Correspondingly, the MEP field of molecule A at \mathbf{r} is given by classical electrostatics as

$$F_A^{\text{MEP}}(\mathbf{r}) = \sum_{i \in A} \frac{q_i}{|\mathbf{r} - \mathbf{R}_i|} \quad (3)$$

where q_i is the partial charge on atom i and \mathbf{R}_i is its position in space. A three-Gaussian expansion of the $1/r$ term is used to avoid discontinuity of the function at the nuclei [26]. This expansion also simplifies evaluation of the integral described in Equation 4 below. Although all the approximations employed here are rather crude, they nonetheless are adequate to qualitatively reproduce the steric and electrostatic features of most molecules.

Molecular-field-based similarity

Once the appropriate molecular field or fields are chosen, a *similarity measure*, Z_{AB}^{MF} , can be computed between any pair of molecules A and B [27] with respect to the field MF:

$$Z_{AB}^{\text{MF}} = \int F_A^{\text{MF}}(\mathbf{r}) F_B^{\text{MF}}(\mathbf{r}) d\mathbf{r} \quad (4)$$

If B is identical to A , the value obtained from Equation 4 is a measure of the *self-similarity* of A .

As the similarity measure defined in Equation 4 depends upon the size of molecules A and B , to compare molecules of widely differing sizes it is useful to normalize the similarity measure. In the current work, the normalized similarity measure initially described by Carbó et al. [27], termed a pairwise *molecular similarity index* (MSI), S_{AB}^{MF} , is used, viz.

$$S_{AB}^{\text{MF}} = \frac{Z_{AB}^{\text{MF}}}{[Z_{AA}^{\text{MF}} \cdot Z_{BB}^{\text{MF}}]^{1/2}} \quad (5)$$

Depending upon the nature of the molecular field employed, the range of possible values for the pairwise MSI is $[0,1]$ for positive-definite fields such as the MSV and $[-1,1]$ for non-positive-definite fields such as the MEP. A value of unity is achieved only in the case of identical fields; any value less than unity reflects a degree of dissimilarity between the fields and thus between the molecules represented by the fields.

In the current work, S_{AB} (the pairwise MSI for molecules A and B) is obtained as a weighted sum of the pairwise MSIs for each field:

$$S_{AB} = \lambda S_{AB}^{\text{MSV}} + (1 - \lambda) S_{AB}^{\text{MEP}} \quad (6)$$

where $\lambda = 0.66$, corresponding to an approximate 2:1 weighting of the MSV field compared to the MEP field. As is seen in our earlier work [6], the larger contribution of the MSV-field-based component of the similarity partially compensates for the very strong MEP-field-based dependencies of the molecular alignments.

Regardless of the type of function used to characterize molecular-field-based similarity (see e.g. Equations 4, 5, or 6), they all depend upon the *relative position* and *orientation* of the two molecules, *A* and *B*, being compared. Thus, determination of the best molecular-field-based alignment requires an exploration of the highly non-linear *similarity space* of *A* and *B*. The molecular similarity space of *A* and *B*, also called the pairwise similarity space or simply the similarity space of *A* and *B*, is defined as the set of all possible relative positions and orientations of *A* and *B* along with a function that assigns a similarity value to each relative position and orientation. Because of the non-linear nature of the similarity space, numerous maxima exist, each maximum corresponding to a given *pairwise-alignment solution*. The largest maximum corresponds nominally to the ‘best’ pairwise-alignment solution. However, as is seen below for the three NNRTIs examined in the present work, the best solution for each pair of molecules may lead to inconsistencies when three or more molecules are considered simultaneously or in rare cases when the results of field-based similarity searches can be compared with relative binding geometries derived from X-ray crystallographic studies (vide infra). Because of the non-linearity of the similarity functions, it is difficult to identify the global similarity maximum with complete certainty. Convergence of gradient-based optimization methods, such as the one used here, is almost always to the critical point, a maximum in the present case, nearest the starting point of the search. To ensure that all relevant maxima have been identified, a systematic spherical search algorithm has been implemented in MIMIC. In the algorithm, one molecule (the *reference molecule*) is held fixed while the other molecule (the *adapting molecule*) is systematically placed in a number of unique starting orientations about the reference molecule. The field-based similarity of the molecular pair is then optimized for each of the starting geometries using the gradient-based procedure alluded to above. This procedure affords a wide and uniformly distributed exploration of the similarity space, which ensures that all relevant pairwise-alignment solutions are likely to be obtained. Further details and discussion on algorithm and related matters can be found in the original paper [24].

The simultaneous matching of multi-molecule fields is seen below to be crucial for the development of *consistent relative binding orientations* of sets of bioactive molecules, NNRTIs in the present case. The

molecular-field-based similarity of the multi-molecule set $\mathbf{M} = \{A, B, C, \dots\}$ is given by the *average* of the pairwise similarities, i.e.,

$$S_M = \left[\frac{(|\mathbf{M}|^2 - |\mathbf{M}|)}{2} \right]^{-1} \sum_{I < J} S_{IJ} \quad (7)$$

where $|\mathbf{M}|$ is the number of molecules in set \mathbf{M} . Equation 7 provides a computationally efficient means for determining multi-molecule similarity. Multi-molecule similarity spaces can also be defined in an entirely analogous fashion to pairwise similarity spaces, except that the set of relative positions and orientations now are related to three-molecule, four-molecule, or *M*-molecule ensembles, and Equation 7 replaces the appropriate pairwise similarity measure or index.

S_M can also be used to analyze the consistency of pairwise field matchings and molecular alignments compared to the corresponding fully optimized multi-molecule results. A *pairwise consistency index* (PCI) is defined as

$$\text{PCI} = \frac{S_M^\circ}{S_M} \quad (8)$$

where S_M° is the multi-molecule similarity value obtained from Equation 7 by directly superposing the molecules in set \mathbf{M} according to their optimized pairwise alignments, and S_M is the similarity value obtained from full multi-molecule optimization according to Equation 7.

Experimental molecular alignment

In the current study, structures of the three NNRTIs investigated are taken from the X-ray crystal structures of the respective inhibitor–protein complexes (see Table 1). Hydrogen atoms are added to the inhibitor structures in a manner that minimizes unfavorable eclipsing interactions with neighboring atoms. Partial atomic charges are determined using the AM1 Hamiltonian, as implemented in the MOPAC program [28]. The inhibitor structures are held fixed throughout the entire similarity-matching process.

An ‘experimentally derived’ model of the relative binding orientations of the three inhibitors can be obtained from the X-ray crystal structure data as follows. The backbone atoms of the P66 subunit of each of the ligand–protein complexes are employed as a basis for aligning the three complexes simultaneously with MIMIC using only the MSV field. Removal of the aligned protein moieties then leaves the NNRTIs

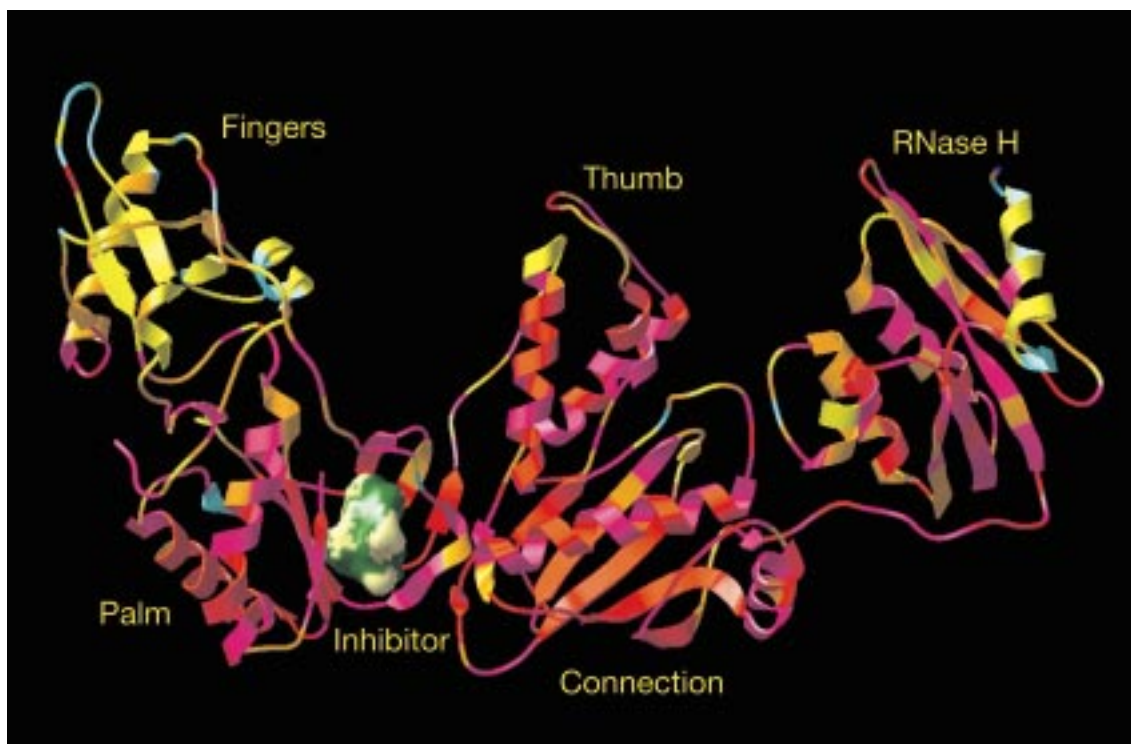


Figure 1. A 'ribbon and tube' representation of the P66 subunit from the 1HNV crystal structure and the MSV-field colored molecular van der Waals surface representation for the union of the three NNRTI molecules located in the binding site. The P66 subunit residues are colored to represent the best inter-chain inter-residue similarity to the other P66 main-chain residues; red, orange, yellow, cyan to blue corresponding to similarities ranging from 1.0, 0.8, 0.6, 0.4, 0.2 to 0.0 respectively.

in an alignment that is tantamount to an experimental model of the relative binding orientations of the three inhibitors.

The validity of the relative binding orientations obtained by MIMIC can then be evaluated by comparing them to the experimentally derived alignments. Such comparisons, however, must be made with caution because of a number of uncertainties that are inherent in the procedure. For example, the optimal field-based alignment of the three P66 backbones (P66-RT) yields a similarity value of 0.6649, indicating that there are some differences among the structures of the three P66 backbones. Using a more conventional measure, the pairwise inter-chain rms distances between the corresponding reverse transcriptase P66 subunit C α 's are 1.66, 1.61 and 0.96 Å for P66-RT(NEVI)/P66-RT(TIBO), P66-RT(NEVI)/P66-RT(α APA) and P66-RT(α APA)/P66-RT(TIBO), respectively, or an average rms distance of 1.41 Å. These results indicate that there are conformational differences between the P66 main chains in these structures. The differences can arise from a number of factors including differences

in the backbone conformation induced by the ligands, in the resolution of the X-ray data, and in the crystallization conditions, which may influence the crystal space group, to name a few. Thus, caution must be exercised in utilizing experimentally derived relative binding orientations as a basis for evaluating the validity of alignments obtained from MIMIC calculations. However, Figure 1 shows the results of evaluating inter-residue MSV similarities between the three P66 backbones. It reveals that the regions of largest variation are located in the 'finger', 'thumb' and RNase H regions and that the 'palm' region containing the inhibitor-binding site is a conformationally conserved region of the P66 subunit.

Once the similarity-based alignment of a set of molecules is completed and validated, several types of *post-alignment analysis* can be performed. These analyses are designed to extract information on key structural features of the molecules under study. Two types of post-alignment analysis are presented below. The first deals with the identification of more traditional pharmacophoric patterns and the second deals

with the graphical portrayal of molecular *similarity fields*.

Feature-based pharmacophoric patterns

One form of post-alignment analysis is aimed at the discovery of pharmacophoric patterns or *pharmacophores*. This can be accomplished once a validated multi-molecule alignment is determined, through an analysis of *atomic similarities*. Atomic similarities are easily computed from the following formula, which is obtained by substituting Equations 1 and 4 into Equation 5 and rearranging terms:

$$S_{A_i B}^{\text{MF}} = \frac{\int F_{A_i}^{\text{MF}}(\mathbf{r}) \cdot F_B^{\text{MF}}(\mathbf{r}) d\mathbf{r}}{\left[\int F_A^{\text{MF}}(\mathbf{r}) \cdot F_A^{\text{MF}}(\mathbf{r}) d\mathbf{r} \right]^{1/2} \left[\int F_B^{\text{MF}}(\mathbf{r}) \cdot F_B^{\text{MF}}(\mathbf{r}) d\mathbf{r} \right]^{1/2}} \quad (9)$$

where A_i represents the contribution for the i th atom of molecule A. The total similarity can be represented as the sum of atomic similarities, and is given by

$$S_{AB}^{\text{MF}} = \sum_i S_{A_i B}^{\text{MF}} \quad (10)$$

Regions of concentrated atomic similarity in pairwise or multi-molecule alignments, called *maximum similarity loci*, correspond to pharmacophoric features. These features are the basis for defining a putative pharmacophore and may include those typically used to define classical pharmacophores such as hydrogen bond donors and acceptors, charged groups, and aromatic rings, but can be more general. Once defined, the pharmacophores can then be used to evaluate and/or search for new molecular patterns that may be of potential value in discovering novel leads that will aid the drug-design process.

Field-based pharmacophoric patterns

Although molecular-field-based similarity calculations are carried out using continuous representations of the appropriate molecular fields (see e.g. Equations 1–3), it is possible to represent the *similarity field* surrounding a set of aligned molecules by mapping the continuous similarity-field values onto a 3D set of grid points into which the aligned molecules are embedded. Alternatively, the similarity values can be mapped onto, for example, solvent-accessible or van der Waals surfaces that surround the aligned molecules. Each grid or surface point \mathbf{r}_i is assigned a

field-based similarity value $S_{AB}^{\text{MF}}(\mathbf{r}_i)$ according to the following discretized variant of Equation 5:

$$S_{AB}^{\text{MF}}(\mathbf{r}_i) = \frac{F_A^{\text{MF}}(\mathbf{r}_i) \cdot F_B^{\text{MF}}(\mathbf{r}_i)}{\left[\sum_j F_A^{\text{MF}}(\mathbf{r}_j) \cdot F_A^{\text{MF}}(\mathbf{r}_j) \right]^{1/2} \left[\sum_j F_B^{\text{MF}}(\mathbf{r}_j) \cdot F_B^{\text{MF}}(\mathbf{r}_j) \right]^{1/2}} \quad (11)$$

where $F_A^{\text{MF}}(\mathbf{r}_i)$ and $F_B^{\text{MF}}(\mathbf{r}_i)$ are the field values at \mathbf{r}_i for molecules A and B, respectively, and the summations are carried out over all the grid or surface points in the set. If the summation is carried out over the relevant set of points, Equation 11 yields

$$S_{AB}^{\text{MF}} = \sum_i S_{AB}^{\text{MF}}(\mathbf{r}_i) \quad (12)$$

Note that combining Equations 11 and 12 yields the discrete form of Equation 5. Computations carried out using either the analytical or numerical approach yield comparable similarity values [6].

The percent of the total field-based similarity at each field point is given by

$$C_{AB}^{\text{MF}}(\mathbf{r}_i) = \frac{S_{AB}^{\text{MF}}(\mathbf{r}_i)}{S_{AB}^{\text{MF}}} \times 100 \quad (13)$$

Representation of $C_{AB}^{\text{MF}}(\mathbf{r}_i)$ values at each of the points on a surface surrounding the superposed molecules or, alternatively, using iso-similarity surfaces enhances visual recognition of common steric and electrostatic field-based patterns of similarity. The recognition of such patterns may be useful in deducing important structural features and pharmacophoric-field patterns that contribute to the bioactivity of the molecules [7]. Application of these graphical representations in the analysis of the three HIV-1 RT inhibitors is presented below.

Application to non-nucleoside HIV-1 RT inhibitors

Initially, an extensive exploration of all possible pairwise alignments is carried out on the three NNRTIs, without regard for information on their actual binding modes. This is followed by a detailed examination of the different pairwise-alignment solutions, which are then combined to generate a set of *preliminary* multi-molecule alignments that are subsequently optimized. Evaluation of the optimized multi-molecule alignments provides a basis for identifying the best relative binding models. These models can then be compared to an experimentally derived relative binding model obtained from X-ray crystal structure data,

as described earlier. In general, however, experimentally derived relative binding models will not exist. In such cases, it is necessary to examine the alignments obtained from MIMIC for internal consistency; this important issue is dealt with below. Finally, two different visual post-alignment analyses are presented to elucidate potential pharmacophoric patterns.

Analysis of pairwise-alignment solutions

Once a family of pairwise-alignment solutions has been obtained for the set of molecules of interest, {NEVI, TIBO, α APA} in the present case, the solutions can be further analyzed for their consistency with respect to optimal multi-molecule alignments. The results of explorations of the similarity spaces corresponding to the three molecular pairs {NEVI, TIBO}, {NEVI, α APA}, and {TIBO, α APA} are collected in Table 2. The five best pairwise-alignment solutions are given for each pair. Similarity values corresponding to these solutions represent the degree of similarity associated with specific 3D molecular alignments for each of the three pairs of molecules. It is important to note that, as discussed earlier, the inhibitor conformations used in this study correspond to those found in the X-ray crystal structures of the NNRTI/HIV-1 RT complexes (see Table 1). The issue of conformational flexibility of the NNRTI ligands will be dealt with in a subsequent paper. Agreement of the pairwise molecular alignments obtained from MIMIC with those derived from experimental crystal structure data, as described earlier, represents the most basic test of the validity of the MIMIC approach.

The molecular alignments in closest agreement with the experimentally derived relative binding orientations are marked in boldface. The similarity values for the pairs of experimentally-derived model overlays given in Table 2 are consistently smaller than the corresponding optimized inhibitor overlays. The lower experimental model inhibitor similarities result from complementary differences with the binding site that cannot be accounted for when comparing only the inhibitor structures. It is observed that for the pairs {NEVI, α APA} and {TIBO, α APA}, the experimentally derived alignment agrees most closely with the best alignment solution found by MIMIC. However, for the pair {NEVI, TIBO}, the second-best alignment solution most closely matches that obtained from ‘experiment’. Table 2 also shows that the values of the different similarity maxima for a given molecular pair are all relatively close in magnitude. Only the {TIBO, α APA} pair shows a clear gap between the best align-

ment solution and the remaining ones. Nevertheless, even when the gap is small, significant structural differences in the pairwise-alignment solutions are observed. This is clearly evident in Figure 2, which depicts the structural superpositions corresponding to the best and second-best pairwise-alignment solutions for the three molecular pairs. The lack of consistency among the pairwise alignments for the three best solutions is clearly evident in Figure 2. For example, it can be seen, taking NEVI (in yellow) as the reference molecule and TIBO (in red) and α APA (in green) as the adapting molecules, that direct superposition of the best alignment solutions associated with the pairs {NEVI, TIBO} and {NEVI, α APA} is inconsistent with the best alignment solution for the {TIBO, α APA} pair. In fact, in this particular case, it corresponds to the second-best alignment solution for the {TIBO, α APA} pair.

From these intermediate pairwise results, it is clear that deriving structural models based upon the relative binding alignments obtained by similarity based methods using pairwise similarity can lead to inconsistent results. Moreover, the results indicate that the best pairwise-alignment solutions are not always in agreement with the corresponding experimentally derived alignments. Thus, similarity methods that are based upon finding a single ‘global’ pairwise similarity maximum may produce incorrect alignments due to inconsistencies that can arise when the best pairwise solution does not correspond to the correct, i.e., the experimentally derived, alignment. This implies that, in principle, several pairwise-alignment solutions should be considered when constructing multi-molecule relative binding models, and it clearly shows the need to go beyond pairwise similarity to ensure that reliable multi-molecule alignments are obtained.

Multi-molecule alignments

Relative structural binding models based on simultaneous multi-molecule alignments rather than on combinations of pairwise alignments should provide more general and reliable results and help avoid some of the inadequacies inherent in using only pairwise alignments. A systematic procedure is used in the present work to obtain optimal multi-molecule alignments: initial superpositions of the molecules in the set are obtained from combinations of pairwise-alignment solutions of a given reference molecule with each other molecule in the set. This procedure establishes a connection among the different pairwise-alignment so-

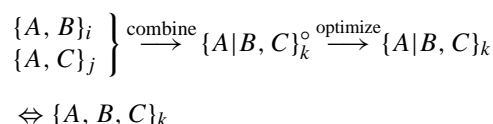
Table 2. Similarity values^a corresponding to the five best molecular overlays obtained from explorations of the similarity spaces induced by the three possible pairs of inhibitors and by the pairs of inhibitor overlays obtained from the experimentally derived model

Molecular pair	1	2	3	4	5	Exp.
{NEVI, TIBO}	0.6309	0.6216	0.5954	0.5759	0.5691	0.4335
{NEVI, α APA}	0.6021	0.5955	0.5815	0.5631	0.5418	1.5604
{TIBO, α APA}	0.6266	0.5700	0.5317	0.5214	0.5020	0.3668

^aOverlays in agreement with the experimentally found relative binding alignments are indicated in boldface.

lutions and the multi-molecule alignment generated from them [7,29].

For example, consider the pairwise solutions $\{A, B\}_i$ and $\{A, C\}_j$, which correspond to the i th and j th pairwise-alignment solutions for A and B , and A and C , respectively. As seen in Scheme 1 below,



Scheme 1

the two pairwise-alignment solutions are combined, taking A as the reference, to yield the unoptimized ‘ternary complex’ $\{A|B, C\}^{\circ}$, where the superscript ‘ \circ ’ indicates that the complex is not fully optimized. The vertical bar within the bracket separates the reference molecule, A , located to the left of the bar, from the adapting molecules, B and C , located to the right of the bar. Note that once the alignment of B and C is given relative to A , the alignment of B with C , $\{B, C\}^{\circ}$, is completely specified and, thus, need not be considered explicitly. In the final step, $\{A|B, C\}^{\circ}$ is optimized and yields, say, the k th ternary-alignment solution, $\{A|B, C\}_k$. Since the relative orientations among all molecules have been optimized simultaneously, the final multi-molecule alignment solutions are independent of the molecule selected as reference. Hence, the distinction accorded to the reference molecule need not be retained, and $\{A|B, C\}_k$ is written as $\{A, B, C\}_k$ (see Scheme 1). In the current notation, the pairwise consistency index, PCI_k , given in Equation 8, can be written as

$$\text{PCI}_k \Leftrightarrow \frac{S\{A|B, C\}_k^{\circ}}{S\{A, B, C\}_k} \quad (14)$$

and is a measure of the degree of consistency or agreement of the ternary alignment derived from purely

pairwise-alignment solutions, $\{A|B, C\}_k^{\circ}$, and that obtained by optimization of the ternary ensemble of molecules, $\{A, B, C\}_k$.

The above process is repeated with different pairwise-alignment solutions and different reference molecules until all the ‘high-quality’ pairwise-alignment solutions have been considered; usually the first five solutions (i.e., $i, j = 1, 2, \dots, 5$) are sufficient for practical purposes. The result is a set of ternary-alignment solutions $\{A|B, C\}_k$, $k = 1, 2, 3, \dots, n \dots$, which, as noted above, are equivalent to $\{A, B, C\}_k$, $k = 1, 2, 3, \dots, n$. As is seen below for the case of the three HIV-1 NNRTIs examined in this paper, the best ternary-alignment solution corresponds most closely to the experimentally derived alignment.

In most cases of practical importance, however, it is not possible to determine whether the experimentally correct solution has been obtained. Therefore, higher level (i.e., quaternary) multi-molecule alignments may be required and, thus, data from an additional molecule must be available. This process can be continued until a self-consistent alignment solution is obtained.

Table 3 contains data for the five best ternary-alignment solutions of the three NNRTIs, i.e., $\{\text{NEVI}, \text{TIBO}, \alpha\text{APA}\}_k$, $k = 1, 2, \dots, 5$. The best ternary-alignment solution, $\{\text{NEVI}, \text{TIBO}, \alpha\text{APA}\}_1$, is 0.6137 and is indicated in boldface in the first row of the table. This solution corresponds closely to the alignment derived from experimental crystallographic data as described earlier.

A comparison of the experimentally derived ternary alignment with that obtained from MIMIC is presented in Figure 3. It is clear from the figure that the general structural features of both ternary alignments are qualitatively, if not quantitatively, quite similar. The pairwise similarity values also given in the first line of the table for $\{\text{NEVI}, \text{TIBO}\}$, $\{\text{NEVI}, \alpha\text{APA}\}$,

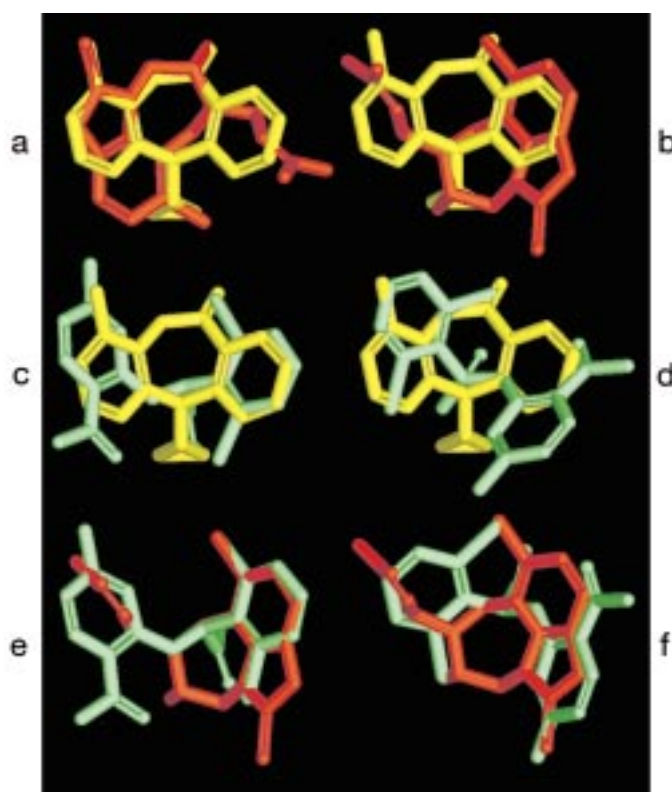


Figure 2. Molecular structure overlays showing the best (left column) and second-best (right column) pairwise similarity alignments for NEVI (yellow), TIBO (red) and α APA (green).

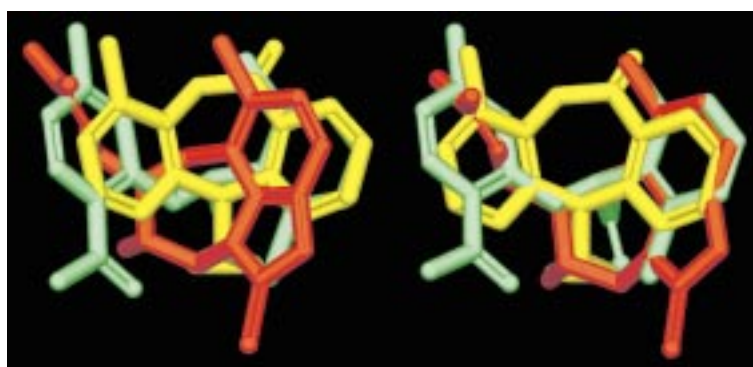


Figure 3. Molecular structure overlays showing the experimental superposition (left) obtained from the similarity alignment of the corresponding P66-backbone crystal structures and the three-molecule similarity alignment (right) of the three NNRTIs; NEVI (yellow), TIBO (red) and α APA (green).

and {TIBO, α APA} are obtained from the corresponding ternary-alignment solution and, thus, are not necessarily better than the values obtained for explicitly pairwise-optimized molecules. This is clearly seen in the column of the table corresponding to the pair {NEVI, TIBO}, where the similarity value obtained

from the ternary alignment, 0.6191, is less than the value obtained for the best and second-best pairwise-alignment solutions, 0.6309 and 0.6216, respectively, as given in Table 2. Interestingly, the similarity value for the pair {NEVI, TIBO}, 0.6220, which is derived from the second-best ternary-alignment solution, is

Table 3. Pairwise consistency analysis of the five best three-molecule overlays collecting the pairwise similarity values^{a,b}, the combined average pairwise similarity value for the set of three molecules^c (S_M), and the pairwise consistency index (PCI) at the initial pairwise superposition (see Table 1)

Rank	Molecular set	{NEVI, TIBO}	{NEVI, α APA}	{TIBO, α APA}	S_M	PCI
I	{NEVI, TIBO, α APA} ₁	0.6191	0.5974	0.6246	0.6137	
	{NEVI TIBO, α APA} ^o	0.6216 (2)	0.6021 (1)	<i>0.5887</i>	0.6041	0.9844
	{TIBO NEVI, α APA} ^o	0.6216 (2)	<i>0.5835</i>	0.6266 (1)	0.6106	0.9948
	{ α APA NEVI, TIBO} ^o	<i>0.5910</i>	0.6021 (1)	0.6266 (1)	0.6066	0.9884
II	{NEVI, TIBO, α APA} ₂	0.6220	0.5655	0.5656	0.5844	
	{NEVI TIBO, α APA} ^o	0.6309 (1)	0.6021 (1)	<i>0.4336</i>	0.5555	0.9505
	{TIBO NEVI, α APA} ^o	0.6309 (1)	<i>0.5449</i>	0.5700 (2)	0.5819	0.9957
	{ α APA NEVI, TIBO} ^o	<i>0.5451</i>	0.6021 (1)	0.5700 (2)	0.5724	0.9795
III	{NEVI, TIBO, α APA} ₃	0.6265	0.4870	0.6236	0.5790	
	{NEVI TIBO, α APA} ^o	0.6309 (1)	0.5069 (12)	<i>0.5063</i>	0.5480	0.9465
	{TIBO NEVI, α APA} ^o	0.6309 (1)	<i>0.4672</i>	0.6266 (1)	0.5749	0.9929
	{ α APA NEVI, TIBO} ^o	<i>0.5539</i>	0.5069 (12)	0.6266 (1)	0.5625	0.9715
IV	{NEVI, TIBO, α APA} ₄	0.5734	0.5371	0.6245	0.5783	
	{NEVI TIBO, α APA} ^o	0.5954 (3)	0.5418 (5)	<i>0.4489</i>	0.5287	0.9142
	{TIBO NEVI, α APA} ^o	0.5954 (3)	<i>0.4528</i>	0.6266 (1)	0.5583	0.9654
	{ α APA NEVI, TIBO} ^o	<i>0.5473</i>	0.5418 (5)	0.6266 (1)	0.5719	0.9889
V	{NEVI, TIBO, α APA} ₅	0.5105	0.5929	0.6252	0.5762	
	{NEVI TIBO, α APA} ^o	0.5344 (11)	0.5955 (2)	<i>0.4144</i>	0.5148	0.8934
	{TIBO NEVI, α APA} ^o	<i>0.5344</i> (11)	<i>0.4388</i>	0.6266 (1)	0.5333	0.9255
	{ α APA NEVI, TIBO} ^o	<i>0.4981</i>	0.5955 (2)	0.6266 (1)	0.5734	0.9951

^aThe italicized values correspond to the two-molecule similarity value obtained as a result of the two optimized pairwise matches to the same reference.

^bThe values in parentheses correspond to the ranking of the pairwise alignment.

^cThe three-molecule alignment that agrees most closely with the experimentally derived relative binding alignment is given in boldface.

greater than the corresponding value of 0.6191, which is derived from the best ternary-alignment solution. There is no inherent reason why these ‘derived’ pairwise similarity values should fall in any particular order, as the optimization procedure seeks only to find solutions that maximize the similarity of the ternary ensemble of molecules.

The three rank I alignments lying below {NEVI, TIBO, α APA}₁, namely {NEVI|TIBO, α APA}^o, {TIBO|NEVI, α APA}^o, and { α APA|NEVI, TIBO}^o, represent ternary alignments constructed from appropriate pairwise-alignment solutions that all converge to the same solution. Consider, for example, {NEVI|TIBO, α APA}^o, which is constructed from {NEVI, TIBO}₂ and {NEVI, α APA}₁, corresponding to the second-best and best pairwise-alignment solutions (indicated in Table 3 by the numbers in parentheses located to the right of the similarity value), 0.6216 and 0.6021 respectively. As noted above, given

two pairwise-alignment solutions, the third is determined; in this case the unoptimized pairwise similarity value for {TIBO, α APA}^o is 0.5887 (indicated in italics in the table). The similarity value, S_M , for {NEVI|TIBO, α APA}^o is obtained by taking the average of the pairwise similarity values: $(1/3)[0.6216 + 0.6021 + 0.5887] = 0.6041$. The PCI is then obtained simply by taking the *ratio* of the similarity values of {NEVI|TIBO, α APA}^o and {NEVI, TIBO, α APA}₁ yielding a value of 0.9844, which indicates reasonably good consistency between the fully optimized and directly superposed pairwise-optimized solutions.

Most importantly, Table 3 clearly shows that direct superposition of the best pairwise-optimized similarity solutions does *not* necessarily lead to the best ternary-alignment solution. For example, construction of the ternary ensemble {NEVI|TIBO, α APA}^o from the best pairwise-alignment solutions {NEVI|TIBO}₁ and {NEVI| α APA}₁ leads upon op-

timization to the second-best ternary-alignment solution, $\{\text{NEVI}, \text{TIBO}, \alpha\text{APA}\}_2$, rather than to the best solution. In fact, two of the three derived ternary ensembles that converge upon optimization to the best ternary-alignment solution are constructed from secondary pairwise-alignment solutions: $\{\text{NEVI}|\text{TIBO}, \alpha\text{APA}\}^\circ$ from $\{\text{NEVI}, \text{TIBO}\}_2$ and $\{\text{NEVI}, \alpha\text{APA}\}_1$ and $\{\text{TIBO}|\text{NEVI}, \alpha\text{APA}\}^\circ$ from $\{\text{NEVI}, \text{TIBO}\}_2$ and $\{\text{TIBO}, \alpha\text{APA}\}_1$. Only $\{\alpha\text{APA}|\text{NEVI}, \text{TIBO}\}^\circ$ is constructed from the best pairwise-alignment solutions, $\{\text{NEVI}, \alpha\text{APA}\}_1$ and $\{\text{TIBO}, \alpha\text{APA}\}_1$.

The above discussion is illustrated graphically in Figure 2. The molecular pairs in the left-hand column (Figures 2a, c and e) correspond to the best pairwise-alignment solutions, while those in the right-hand column (Figures 2b, d and f) correspond to the second-best solutions. If NEVI is taken as the reference molecule, it is clear from the figure that directly superimposing the $\{\text{NEVI}, \text{TIBO}\}$ and $\{\text{NEVI}, \alpha\text{APA}\}$ pairs (Figures 2b and c) yields an alignment for the $\{\text{TIBO}, \alpha\text{APA}\}$ pair that corresponds closely to the best pairwise-alignment solution obtained for that pair (Figure 2e). On the other hand, superposing the best pairwise alignments for these two pairs (Figures 2a and 2c) leads to an alignment that is inconsistent with Figure 2e. Moreover, the ternary ensemble obtained by superposing the pairs in Figures 2b and c, as described above, clearly resembles the best ternary-alignment solution.

This clearly indicates one of the major difficulties encountered in practice, namely how does one choose an appropriate set of pairwise-alignment solutions that will ensure that ternary ensembles constructed from them converge to the best ternary-alignment solution? A pragmatic answer to this question is to include a sufficient set of each of the possible pairwise-alignment solutions. As is seen in Table 3, five pairwise-alignment solutions are generally sufficient to obtain the best ternary-alignment solutions. Interestingly, for the case of the third-best ternary-alignment solution, two of the unoptimized ternary ensembles that converge to the proper ternary solution are constructed from the twelfth-best pairwise-alignment solution, $\{\text{NEVI}, \alpha\text{APA}\}_{12}$. Considerably more studies on many sets of molecules will have to be carried out before a more definitive characterization of this issue can be presented.

Recognition of pharmacophoric-field patterns

In field-based approaches to molecular similarity, multi-molecule alignments constitute the basis from

which post-alignment visual analyses can be performed to identify local similarities in 3D space. One of the most widely used graphical tools for locating regions in space that may contribute to biological activity is the CoMFA coefficient contour map [30]. Other strategies that have been used to assist in the visual recognition of common molecular-field regions include root mean square difference maps [10] and largest deviation electrostatic potential surfaces [31].

The construction of molecular-field-based similarity surfaces has recently been investigated as a means for identifying molecular regions that contribute significantly to the total steric and/or electrostatic similarity of a set of aligned molecules (see Equations 11–13). Regions of high similarity are characterized as sterically and/or electrostatically *restricted*, while regions of low similarity are characterized as sterically and/or electrostatically *permissive*.

Molecular-field-based similarity surfaces generated from the best multi-molecule alignment solution for the three inhibitors under study are presented in Figure 4. Figure 4a depicts an alternate view (cf., Figure 3) of the multi-molecule alignment used to generate the surfaces. Figures 4b and c depict the respective steric- and electrostatic-field contributions to the molecular similarity value at points lying on the surface derived from the union of the van der Waals surfaces of the individual molecules in their optimal multi-molecule alignment.

Since the MSV fields are positive-definite, their corresponding similarity surfaces (Figure 4b) can be represented using only two colors: green for those regions above and yellow for those regions below a selected cutoff percentage of the total MSV similarity. In the current work a value of 0.0001% was selected. Note that use of a higher or lower cutoff value allows identification of the most and the least similar regions, respectively. In contrast, MEP fields are non-positive-definite, i.e., they contain both positive and negative field values. As a consequence, MEP-field similarity surfaces (Figure 4c) differentiate those regions of the surface where the field values of all the molecules have the same sign, either positive (colored in blue) or negative (colored in red), from those regions of the surface where this is not the case (colored in yellow). Electrostatic fields for some of these molecules have been previously reported [32,33]. Figure 4c compiles in a single surface representation the combined information from the electrostatic fields of all the molecules, thus enhancing both the simplicity and the validity of this kind of visual analysis.

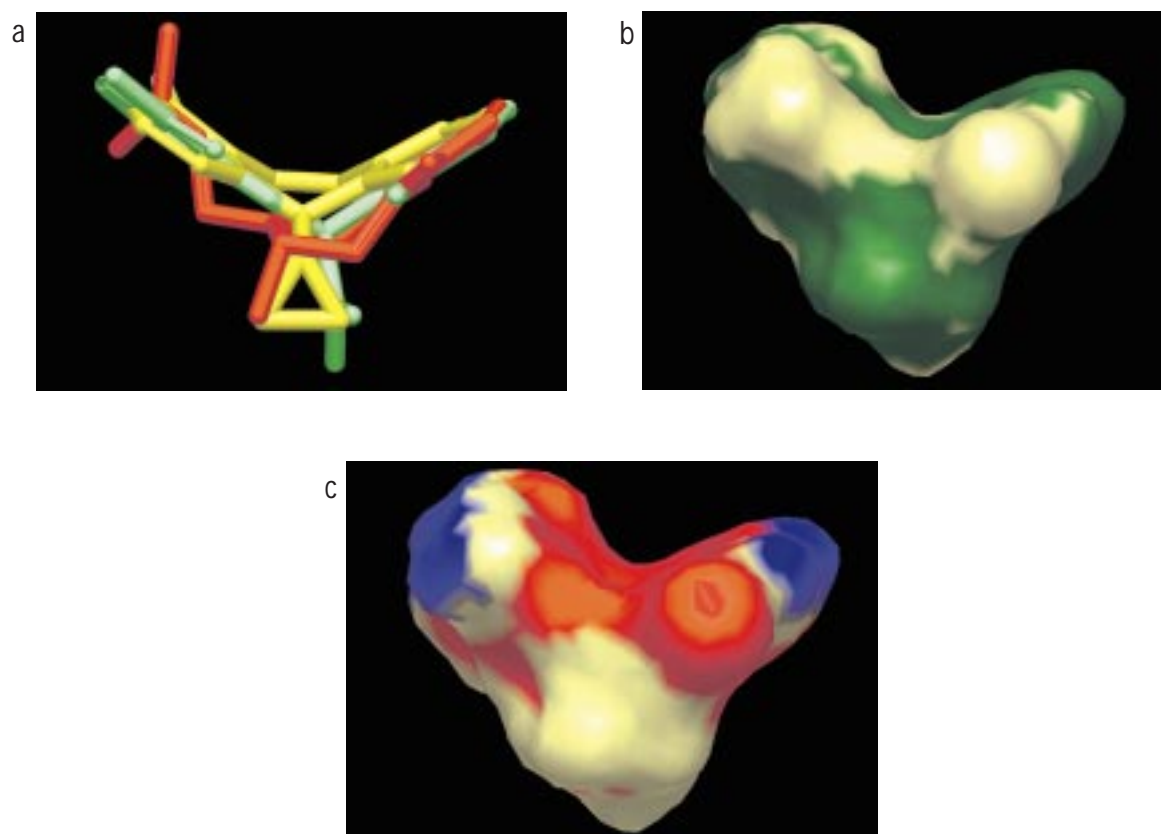


Figure 4. Molecular structure overlay and molecular van der Waals surface representations for the union of the three molecules showing (a) the edge view of the three-molecule similarity alignment of the three NNRTIs; NEVI (yellow), TIBO (red) and α APA (green), (b) the molecular surface showing the regions contributing significantly to the overall MEP-field similarity in green, and (c) the molecular surface showing the regions contributing to the overall MEP-field similarity; all positive MEP in blue, all negative MEP in red and inconsistent MEP in yellow.

The main interest of molecular-field similarity surfaces resides in the fact that the steric and electrostatic characteristics of the receptor site can be inferred from the regions of steric and electrostatic similarity of the aligned molecules themselves. For a selected group of active molecules, regions contributing significantly to the total MSV-field similarity can be associated with sterically constrained regions where the shape of the aligned molecules is sterically constrained by the shape of the receptor cavity. On the other hand, regions contributing significantly to the total MEP-field similarity can be associated with sites where electrostatic interactions may be of key importance for binding. These common molecular-field regions are referred to as *pharmacophoric-field patterns* and define the molecular-field characteristics a molecule should possess in order to be potentially active.

In contrast, those regions where the local field-based similarity contributes weakly, in the case of

MSV fields or, ambiguously, in the case of MEP fields, to the total field-based similarity are associated with less restrictive regions of the binding site. Protein conformational flexibility or greater accessibility of that part of the binding site to the surrounding solvent environment may explain this lack of specificity. In either case, ligand–protein interactions involving these regions are not likely to play a critical role in binding. Therefore, such regions provide potential sites for modification to improve other important biological characteristics of the molecule, such as increased bioavailability or reduced toxicity. Thus, for different reasons, drug designers can certainly benefit from the information obtained from the identification of both high- and low-similarity regions.

To confirm the conclusions drawn from the molecular-field-based similarity surfaces, the three-molecule MSV-field similarity surface is depicted within the binding pocket of the HIV-1 reverse tran-

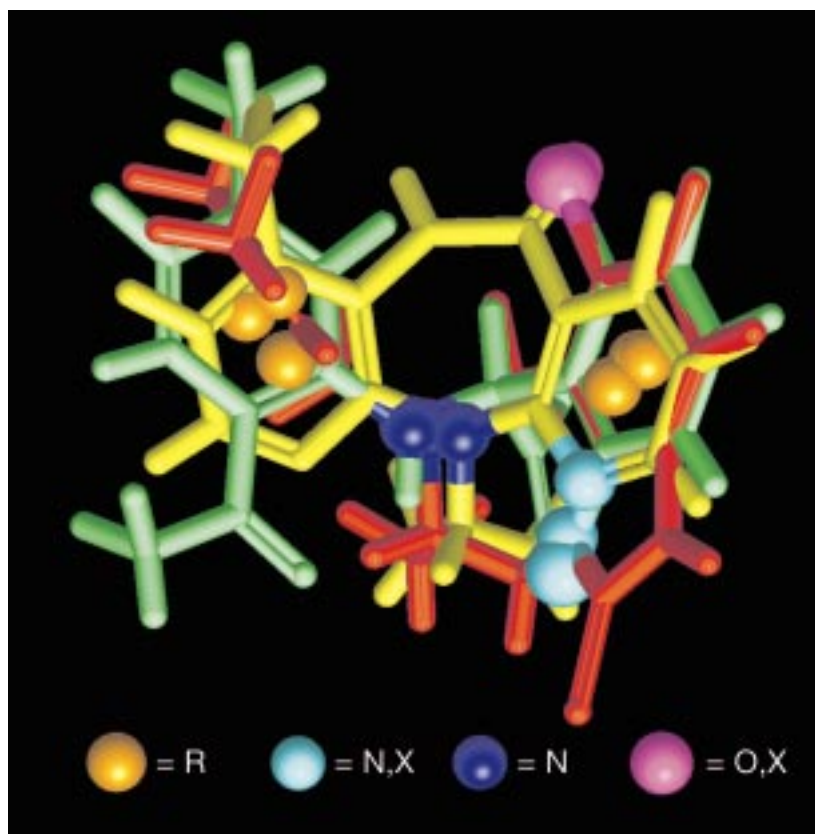


Figure 5. Molecular structure overlay and pharmacophore pattern showing the best NNRTI three-molecule similarity alignment; NEVI (yellow), TIBO (red) and α APA (green). The pharmacophore locations are shown as spheres.

scriptase P66-RT(TIBO) crystal structure (Figure 1). Nevirapine was used to orient three-molecule superposition within the binding site of the protein. As is observed from the figure, the green surfaces, which correspond to sterically restricted regions, clearly identify regions of close complementarity with the protein-binding site which conforms to the shape of active molecules. On the other hand, the yellow surfaces, which correspond to sterically permissive regions, clearly identify those regions exhibiting minimum steric conflict with the steric characteristics of the binding cavity, in this case pointing towards the exterior of the binding pocket. Such information can be of paramount importance in the initial stages of the drug-design process.

Identification of 'classical' pharmacophoric patterns

In addition to the pharmacophoric-field analysis just presented, it is possible within MIMIC to carry out a more classical pharmacophore identification using atomic similarities. From Equation 9 each atom's con-

tribution to a molecule's total similarity can be computed. This is carried out for the three NNRTIs considered in this work, using the optimal three-molecule alignment discussed earlier. Atoms with large atomic-similarity values located within the same region of space but belonging to different molecules constitute a *maximum similarity locus*. The set of maximum similarity loci constitutes a potential pharmacophoric pattern or pharmacophore. Figure 5 illustrates this for five maximum similarity loci represented as clusters of colored spheres. The orange clusters represent the approximate centroids of aromatic rings and the blue cluster represents a grouping of nitrogen atoms, one from each of the three NNRTIs. Both these clusters correspond to the types of molecular features typically used in classical pharmacophore analyses. This, however, is not the case for the cyan and lavender clusters. In the former case, two of the blue spheres correspond to nitrogen atoms on NEVI and TIBO and one corresponds to the bromine atom on α APA. Thus, this pharmacophoric feature does not correspond to a sin-

gle type of molecular feature and as such would not be employed in most classic pharmacophoric pattern analyses. The same is true for the lavender cluster, where one of the spheres corresponds to the carbonyl oxygen of NEVI while the other two correspond to halogen atoms on TIBO and α APA.

Thus, the use of atomic similarity provides a means for deducing objects that are very similar to classical pharmacophores. This adds a new capability to molecular-field-based similarity analysis that allows comparison with other more classical approaches and, more importantly, provides a means for generating pharmacophoric patterns that can be used to search large compound databases for molecules with similar pharmacophoric patterns but belonging to different structural classes.

Conclusions

The above discussion shows quite clearly that the relative binding alignment obtained by MIMIC for the three NNRTIs fixed in their crystallographically determined binding conformations is in qualitative agreement with the 'experimentally' determined binding alignment. The results obtained when conformational flexibility is introduced, which are comparable to those presented above, will be described in a future publication. The discussion also shows the importance of considering more than a single 'best' pairwise-alignment solution and the need to go beyond pairwise to multi-molecule alignments to ensure that an internally consistent relative binding alignment is obtained. The present work has analyzed in detail the connection between the pairwise and three-molecule alignment solutions and delineated the existence of possible alignment inconsistencies at the pairwise level. Finally, it is shown that a number of additional structural insights can be derived from post-alignment visual analyses: one based upon atomic similarities which generate maximum similarity loci that are fundamental to the construction of classical pharmacophores; and another based upon molecular similarity surfaces which depict regions of highly conserved electrostatic or steric similarity surrounding the molecules that can be used as a guide in designing new compounds.

Taken together, molecular-field-based similarity and related analyses provide a powerful set of tools that are useful in the design of bioactive molecules. This is especially true in the absence of detailed structural information on the binding site of the target

protein, with the important caveat that the molecules in question must all bind in a similar manner to the same region of the binding site.

Acknowledgements

We thank Joe Moon and Jim Blinn for integrating MIMIC in the graphical user interface environment of the MOSAIC molecular modeling program. J.M. acknowledges the financial support provided by the Spanish Ministerio de Educacion y Ciencia (reference no. PF95-40314447).

References

1. Bugg, C.E., Carson, W.M. and Montgomery, J.A., *Sci. Am.*, (1993) 92.
2. Gund, P., *Prog. Mol. Subcell Biol.*, 5 (1977) 117.
3. Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A. and Dunn, D.A., In Olson, E.C. and Christofferson, R.E. (Eds.) *Computer-Assisted Drug Design*, ACS Symposium Series, No. 112, American Chemical Society, Washington, DC, 1979, pp. 205–226.
4. Johnson, M.A. and Maggiora, G.M. (Eds.) *Concepts and Applications of Molecular Similarity*, Wiley, New York, NY, 1990.
5. Dean, P.M. (Ed.) *Molecular Similarity in Drug Design*, Blackie Academic, London, 1995.
6. Kearsley, S.K. and Smith, G.M., *Tetrahedron Comput. Methodol.*, 3 (1990) 615.
7. Good, A.C., So, S.-S. and Richards, W.G., *J. Med. Chem.*, 36 (1993) 433.
8. Perkins, T.D.J., Mills, J.E.J. and Dean, P.M., *J. Comput.-Aided Mol. Design*, 9 (1995) 479.
9. Grant, J.A., Gallardo, M.A. and Pickup, B.T., *J. Comput. Chem.*, 17 (1996) 1653.
10. Klebe, G., Mietzner, T. and Weber, F., *J. Comput.-Aided Mol. Design*, 8 (1994) 751.
11. McMartin, C. and Bohacek, R.S., *J. Comput.-Aided Mol. Design*, 9 (1995) 237.
12. Lemmen, C. and Lengauer, T., *J. Comput.-Aided Mol. Design*, 11 (1997) 357.
13. Mestres, J., Rohrer, D.C. and Maggiora, G.M., *J. Comput. Chem.*, 18 (1997) 934.
14. Mestres, J., Rohrer, D.C. and Maggiora, G.M., *J. Mol. Graph. Modelling*, 15 (1997) 114.
15. Klebe, G., In Kubinyi, H. (Ed.) *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM, Leiden, 1993, pp. 173–199.
16. De Clercq, E., *Biochem. Pharmacol.*, 47 (1994) 155.
17. Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. and Steitz, T.A., *Science*, 256 (1992) 1783.
18. Ding, J., Kalysn, D., Moereels, H., Koymans, L., Andries, K., Janssen, P.A.J., Hughes, S.H. and Arnold, E., *Nat. Struct. Biol.*, 2 (1995) 407.
19. Ding, J., Das, K., Tantillo, C., Zhang, W., Clark, Jr., A.D., Jessen, S., Lu, X., Hsiou, Y., Jacobo-Molina, A., Andries, K., Pauwels, R., Moereels, H., Koymans, L., Janssen, P.A.J.,

- Smith, Jr., R.H., Koepke, M.K., Michejda, C.J., Hughes, S.H. and Arnold, E., *Structure*, 3 (1995) 365.
20. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., *J. Mol. Biol.*, 112 (1977) 535.
 21. Merluzzi, V.J., Hargrave, K.D., Labadia, M., Grozinger, K., Skoog, M., Wu, J.C., Shih, C.-K., Eckner, K., Hattox, S., Adams, J., Rosenthal, A.S., Faanes, R., Eckner, R.J., Koup, R.A. and Sullivan, J.L., *Science*, 250 (1990) 1411.
 22. Kukla, M.J., Breslin, H.J., Pauwels, R., Fedde, C.L., Miranda, M., Scott, M.K., Sherrill, R.G., Raeymaekers, A., Van Gelder, J., Andries, K., Janssen, M.A.C., De Clercq, E. and Janssen, P.A.J., *J. Med. Chem.*, 34 (1991) 746.
 23. Pauwels, R., Andries, K., Debyser, Z., Van Daele, P., Schols, D., Stoffels, P., De Vreese, K., Woestenborghs, R., Vandamme, A.-M., Janssen, C.G.M., Anne, J., Cauwenbergh, G., Desmyter, J., Heykants, J., Janssen, M.A.C., De Clercq, E. and Janssen, P.A.J., *Proc. Natl. Acad. Sci. USA*, 90 (1993) 1711.
 24. Rohrer, D.C., In Carbó, R. (Ed.) *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, Kluwer Academic Publishers, Dordrecht, 1995, pp. 141–161.
 25. Audry, E., Dubost, J.P., Colleter, J.C. and Dallet, P., *J. Med. Chem.-Chim. Ther.*, 21 (1986) 71.
 26. Good, A.C., Hodgkin, E.E. and Richards, W.G., *J. Chem. Inf. Comput. Sci.*, 32 (1992) 188.
 27. Carbó, R., Leyda, L. and Arnau, M., *Int. J. Quantum Chem.*, 17 (1980) 1185.
 28. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., *J. Am. Chem. Soc.*, 107 (1985) 3902.
 29. Rohrer, D.C. and Mestres, J., In Coddling, P.W. (Ed.), *Experimental and Conformational Approaches to Structure-Based Drug Design*, Kluwer Academic Publishers, Dordrecht, 1998, pp. 211–222.
 30. Cramer, R.D., Patterson, D.E. and Bunce, J.D., *J. Am. Chem. Soc.*, 110 (1988) 5959.
 31. Beck, B., Glen, R.C. and Clark, T., *J. Mol. Graph.*, 14 (1996) 130.
 32. Mui, P.W., Jacober, S.P., Hargrave, K.D. and Adams, J., *J. Med. Chem.*, 35 (1992) 201.
 33. Schäfer, W., Friebe, W.-G., Leinert, H., Mertens, A., Pool, T., von der Saal, W., Zilch, H., Nuber, B. and Ziegler, M.L., *J. Med. Chem.*, 36 (1993) 726.