Pacific Symposium on Biocomputing 4:413-414 (1999)

FIELD-BASED SIMILARITY FORCING IN ENERGY MINIMIZATION AND MOLECULAR MATCHING

James R. Blinn, Douglas C. Rohrer, and Gerald M. Maggiora
Pharmacia & Upjohn
301 Henrietta Street
Kalamazoo, MI 49007 USA

A new field-based similarity forcing procedure for matching conformationally-flexible molecules is presented. The method extends earlier work on similarity matching of molecules based upon the program MIMIC, by directly coupling a similarity function to a molecular mechanics force field. In this way conformational energetics are fully accounted for in the similarity matching process. Simultaneous similarity/conformational searches can then be undertaken within a Monte Carlo or molecular dynamics framework. Here, a Monte Carlo approach is used to provide a simple example of two HIV-1 reverse transcriptase inhibitors, nevirapine and αAPA, that illustrates the basic characteristics of the method and suggests areas for further investigation.

1. Introduction

Molecular similarity is an important tool in drug design.[1,2] This is especially true in areas where large numbers of molecules must be handled such as in diversity assessment and in the selection of compounds for screening and purchase.[3] Even with the recent explosion in high-throughput technologies in the pharmaceutical industry, the need remains for low-throughput methods that can provide *detailed* assistance in the design of potent new drug molecules. Similarity methods can be of assistance here too, particularly when detailed three-dimensional structural information on the macromolecular target is unavailable.

A variety of strategies have evolved to address this important problem domain. Of the recent approaches, the most notable are the 3D QSAR methods, exemplified by CoMFA, that have proliferated over the last several years.[4] A characteristic of all of these methods is their reliance, to varying degrees, on the nature of the environment surrounding the molecules under study. Molecular fields (*e.g.*, electrostatic) and pseudo-fields (*e.g.*, steric and lipophilic) are generally used, but shape features of the ligands and the binding sites of the target macromolecules are also used. In essentially all field-based methods, the molecules under study must be aligned in some way, usually by field matching. Because of the highly non-linear character of most field-based matching functions, many critical points (*i.e.*, maxima, minima, and saddle points) exist and finding the *global maximum* of the matching function can be difficult. The need for computationally efficient algorithms is paramount. Developing such algorithms is made even more difficult in cases where conformationally-flexible molecules must be matched.

The study reported here briefly describes some of our recent efforts to explore new, more computationally-efficient algorithmic approaches to field-based similarity

matching of small, conformationally-flexible molecules. This work extends the field-based matching capabilities of the program MIMIC.[5-8] In the current version of MIMIC conformational flexibility is treated by a multi-step procedure that becomes computationally intractable as the number of molecules and their conformational flexibility increases. In the new approach, a *field-based similarity restraint* is coupled to the molecular force field, which provides a natural way for incorporating conformational energetics with field-based similarity searching. A significant benefit of this approach is that standard minimization, Monte Carlo (*MC*), and molecular dynamics (*MD*) procedures found in most molecular modeling packages can now be used in similarity searching. A somewhat related approach has been described by Klebe *et al*.[9]

In the present work we focus upon the use of *MC* procedures to study pairwise molecular matching, but the corresponding *MD* procedures have also been implemented. An example is presented of the pairwise matching of two HIV-1 reverse transcriptase inhibitors, nevirapine and αAPA, depicted in Figure 1.[9,10]

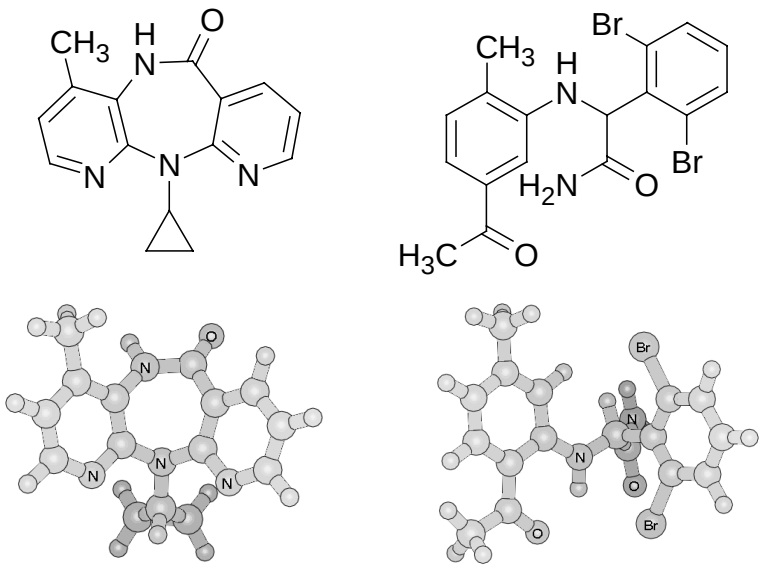
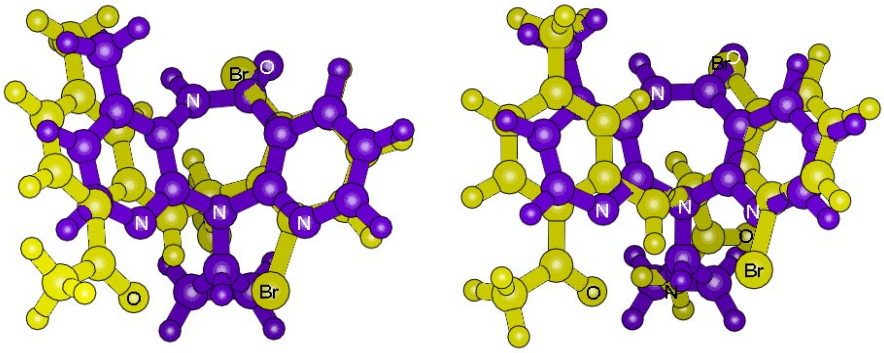


Figure 1. *Top left*: 2D structure of nevirapine. *Top right*: 2D structure of αAPA. *Bottom left*: 3D structure of nevirapine taken from the x-ray structure of its complex with HIV-1 reverse transcriptase—Reference 10. *Bottom right*: 3D structure of αAPA taken from the x-ray structure of its complex with HIV-1 reverse transcriptase—Reference 11.

To test the validity of the method, the computational results achieved for pairwise similarity matching are compared to a *surrogate experimental result*, which is obtained in the following manner. First, the crystallographically-determined main-

chains of the two inhibitor-protein complexes are superimposed using only the steric-field of MIMIC, inducing a relative alignment of the two inhibitors.[8] Removal of the superimposed main-chains leads directly to the aligned inhibitors shown in Figure 2a, a structure tantamount to an "experimentally-derived" alignment. It was shown in our earlier work that using MIMIC to align nevirapine and αAPA, based upon the conformations found crystallographically for these inhibitors in their complexes, produced a relative alignment in qualitative agreement with the experimentally-derived alignment.[5,6,8] As will be shown below, a similar study, but one allowing conformational-flexibility, also leads to a result in qualitative accord with experiment. Although not shown here, simultaneous multi-molecule alignments can also be carried out with this procedure. A number of examples of such alignments obtained with fixed geometries using MIMIC are described in our earlier works.[6,8]



(a) **(b)**

Figure 2. Comparison of (a) the "experimentally-derived" nevirapine/αAPA alignment with (b) the best alignment obtained by the conformationally-flexible field-based similarity matching procedure described in this work (S_{AB} = 0.6265).

2. Methodology

The salient features of the molecular field-based procedures used here are briefly described below. Additional details are given in our earlier works. In the present work, two fields are used, one relating to the *molecular steric volume* (*MSV*) and the other to the *molecular electrostatic potential* (*MEP*), respectively, although other types of fields such as those related to lipophilicity[12,13] or electrotopological states[14] could also be used.

The *MSV* field at a point in space, **r**, for molecule *A* is given by

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

where $f_i^{MSV}(\mathbf{r})$ defines the steric volume of the *i*-th atom, which is represented by a spherically-symmetric Gaussian function located at its nucleus, $\mathbf{R}_i$, *i.e*,

$$f_i^{MSV}(\mathbf{r}) = \alpha_i \cdot \exp\left[-\beta_i |\mathbf{r} - \mathbf{R}_i|^2\right] \quad . \tag{2}$$

The coefficient α_i and the exponent β_i are optimized for each atom as described in earlier works.[5,6] The corresponding *MEP* field is given by the classical expression for the electrostatic-potential field of a set of fixed, partial charges q_i, *i.e.*,

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

where $\mathbf{r}$ and $\mathbf{R}_i$ are the same as given in the *MSV* field term. In the current work, Mulliken charges based on the AM1 Hamiltonian as implemented in the MOPAC program[15] are used. No attempt was made to assess the sensitivity of the results to the atomic charge representation used; work on this problem is on-going. The singularity at the nucleus due to the 1/*r* term is avoided by a three-Gaussian approximation[16] that renders the *MEP* field terms as sums of Gaussians making them analogous in form to the *MSV* field terms. Earlier studies in our laboratory have shown that the relatively crude approximations used to describe the *MSV* and *MEP* fields are, nevertheless, sufficiently accurate to produce meaningful results for a number of different molecular overlays.

Given the two fields, a similarity index based upon Carbo's early work[17] is used to characterize the pairwise similarity,

$$S_{AB}^{MF} = \frac{\int F_A^{MF}(\mathbf{r}) F_B^{MF}(\mathbf{r}) \, d\mathbf{r}}{\sqrt{\int F_A^{MF}(\mathbf{r}) F_A^{MF}(\mathbf{r}) \, d\mathbf{r}} \sqrt{\int F_B^{MF}(\mathbf{r}) F_B^{MF}(\mathbf{r}) \, d\mathbf{r}}} \quad , \tag{4}$$

where the superscript *MF* = *MSV* or *MEP*. The terms in the denominator, which normalize the value of S_{AB}^{MF} to the unit interval [0,1] for the positive-definite *MSV* field and to the interval [-1,1] for the *MEP* field, are related to the self-similarity of molecules *A* and *B*, respectively. The joint effect of the two fields is accounted for by taking a weighted sum of the two field terms

$$S_{AB} = \lambda S_{AB}^{MSV} + (1 - \lambda) S_{AB}^{MEP} \quad , \tag{5}$$

where $0 \le \lambda \le 1$. Here $\lambda = 0.67$ is used, corresponding to at 2:1 ratio of the *MSV* : *MEP* field terms.[6]

Generally, in similarity studies using MIMIC the next step would involve a set of operations designed to determine optimal pairwise alignments. These steps include a detailed conformational analysis of each molecule and selection of a number of important prototypical conformations. Each of the prototype conformations of each molecule is then aligned with the prototypical conformations of all other molecules leading to an explosion in the amount of computation. This is further exacerbated by the multiple solutions obtained for each pair of rigid conformations, which arise from the use of numerous starting geometries to ensure that the global maximum solution is obtained — each calculation being carried out by a gradient-based optimization procedure.[16] In order to investigate alternative strategies for the determination of optimal pairwise alignments, a number of alternative approaches are currently under investigation.

The present work focuses on a new procedure that combines energetics for conformational searching with field-based molecular similarity. The latter is incorporated as a fictitious energy term which provides a similarity-based restraint that influences the relative alignments of each pair of molecules, and indirectly their conformations, subjected to the procedure. This can be formally written

$$E_{AB}^{total} = E_A^{conf} + E_B^{conf} + E_{AB}^{sim} \quad , \tag{6}$$

where the total energy, E_{AB}^{total}, is given as a simple sum of the conformational energies, E_A^{conf} and E_B^{conf}, and the fictitious energy due to the restraining similarity term, E_{AB}^{sim}. The conformational energy includes all relevant energy terms that depend upon molecular structure including, if desired, bond stretching and angle bending in addition to torsional and non-bonding terms. Specifically, the AMBER force field[18] as implemented in MacroModel,[19] *i.e.*, AMBER*, is used in all calculations reported here, but the procedure is completely general and can be implemented for any molecular mechanics force field and any energy/gradient-based method.

The similarity-based energy term is given by

$$E_{AB}^{sim} = K_{sim} \cdot (1 - S_{AB})^2 \tag{7}$$

where K_{sim} is an adjustable proportionality constant, which lies in the range of 5-20 kcal/mol in the present work. Other possible forms for this term are currently under

investigation. Substituting Eq. 5 into Eq. 7 and rearranging terms yields the following expression

$$E_{AB}^{sim} = K_{sim}^{MSV} \cdot \left(- S_{AB}^{MSV} \right) + K_{sim}^{MEP} \cdot \left(- S_{AB}^{MEP} \right) \tag{8}$$

where $K_{sim}^{MSV} : K_{sim}^{MEP} = \lambda : 1 - \lambda$. In the cases studied here, where $\lambda = 0.67$, $K_{sim}^{MSV} : K_{sim}^{MEP} = 2$ and thus, the ratio of the *MSV* to the *MEP* field contributions found to be optimal in our earlier MIMIC calculations (*vide supra*), is maintained.

The *MC* procedure used here is a modification of that developed by Chang *et al.*[20] and implemented in the BatchMin 4.5, which is part of the MacroModel program suite.[19] Briefly, arbitrary conformations are generated for each of the molecules being compared, and the molecules are placed within close enough proximity to at least weakly overlap one another. The relative orientation of the molecules in this initial "overlay" is arbitrary. Random perturbations of the initial conformations are then carried out, and the similarity-restrained energy E_{AB}^{sim} of the resulting conformationally-perturbed "complex" is minimized by a gradient-based procedure. The resulting structure is then used as input to the next iteration. The "structural perturbations" carried out in this and in all subsequent iterations involve both intramolecular (*i.e.*, conformational) and intermolecular (*i.e.*, rotational and translational) degrees-of-freedom. Numerous iterations are carried out to ensure broad coverage of the search space, resulting in the identification of many solutions each corresponding to a particular $min\left(E_{AB}^{total} \right)$. After a significant number of cycles, solutions previously obtained begin to reappear. Eventually a point is reached where new, low-energy solutions are no longer found, at which time the optimization process is terminated. Because of the extensive conformational and similarity searching carried out by this procedure, the particular starting geometry chosen for optimization is unimportant. The paper by Chang *et al.*[20] should be consulted for further details on the sampling scheme and convergence criteria typically employed in this procedure.

3. Results and Discussion

A single search process of 5000 *MC* steps was carried out to identify similarity-restrained energy minima, $min\left(E_{AB}^{total} \right)$, lying within 10 kcal/mol of the lowest-energy pairwise alignment All rotational degrees-of-freedom of essential single bonds in nevirapine and αAPA are included in the optimization. Equivalent atoms, such as arise about an axis of rotation in phenyl rings, are also treated to remove structurally-redundant solutions. The *starting geometry* is depicted in Figure 3: both

nevirapine and αAPA have been altered significantly from their minimum-energy conformations, rotated about their respective center-of-mass and then translated to separate the two molecules. From a comparison of the starting geometry with the structure of the optimal alignment shown in Figure 2b, it is clear that there is a substantial difference between the two.

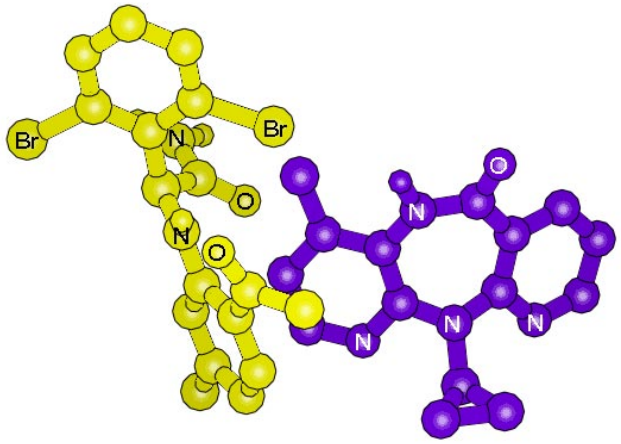


Figure 3. Starting geometry used in the confomrationally-flexible field-based similarity search process.

Of the energy minima obtained, 826 represented unique structures. The solution with the lowest E_{AB}^{total}, which corresponds to a similarity of S_{AB} = 0.6265, is depicted in Figure 2b. The conformations found for nevirapine and αAPA lie quite close to their *nearest* conformational minima, differing by only about 0.1 kcal/mol in each case. Comparison of the structure corresponding to the best conformationally-flexible solution with the experimentally-derived structure depicted in Figure 2a clearly shows that both are in *qualitative* agreement — note that the correct enantiomorph of nevirapine is obtained. Since the *raison d'etre* of the field-based similarity approach described here [*See also*, References 5-8] is to provide assistance in the design of drugs, such qualitative structural information is generally adequate.

Although the best alignment solution corresponds to the experimentally-derived one, the second through sixth solutions involve the other nevirapine enantiomorph. Moreover, the second-best solution lies only 0.07 kcal/mol higher (0.01 kcal/mol higher strain energy and 0.002 lower similarity) than the best solution. If the correct nevaripine enantiomorph is held rigid, allowing only αAPA to flex in the calculations, 97 solutions are obtained within the 10 kcal/mol threshold. As is the case where both molecules are allowed to flex, the best solution again corresponds

to the experimentally-derived one. However, the second-best solution lies more than 1.6 kcal/mol above the best solution and corresponds quite closely with the seventh solution obtained in the case of totally flexible matching. These observations clearly indicate the importance of molecular conformation to the matching process, and the need to ensure that the force field used properly accounts for conformation.

The best solution obtained here compares favorably with the best pairwise solution produced by MIMIC for the nevirapine/αAPA pair using fixed conformations corresponding to those found in the crystal structures of their respective complexes with HIV-1 reverse transcriptase.[10,11] Interestingly, the value of S_{AB} = 0.6265 obtained here for the best flexible, pairwise-alignment solution is better than the value of S_{AB} = 0.6021 obtained by MIMIC for the best pairwise alignment based upon the fixed, crystallographically-determined conformations of nevirapine and αAPA.

4. Summary and Conclusions

A new method is described that combines the field-based similarity procedure in MIMIC with the powerful *MC*-based conformational-searching procedure in BatchMin.[19] This combination significantly enhances the capabilities of field-based similarity methods, such as MIMIC, that use fixed conformations in the matching process. The coupling is obtained simply by appending a similarity-based energy term to the sum of the conformational energies of both molecules. Although the molecules are in close proximity they are coupled only by their pairwise similarity through the energy term, which acts as a restraint on the conformational spaces of the two molecules. All intermolecular interactions due to terms in the molecular force field are set to zero. Because of the thoroughness of the *MC*-based conformational search, the choice of starting geometry is reasonably arbitrary. This is in clear distinction to the case in MIMIC where numerous starting geometries are systematically generated for each pair of fixed conformations of the molecules being match, in order to ensure that the best global similarity solution is obtained.

A test of the new method was carried out on the pair of HIV-1 reverse transcriptase inhibitors nevirapine and αAPA. The best out of the 931 solutions obtained is found to be in qualitative agreement with a surrogate experimental alignment obtained from the crystal structures of the two inhibitor complexes with HIV-1 reverse transcriptase. Preliminary results from a *ternary alignment* of nevirapine, αAPA, and TIBO, another potent reverse transcriptase inhibitor, also appear to be in agreement with the experimentally-based alignment obtained from the crystal structures of the three inhibitor complexes.

While these results are certainly encouraging, considerable analysis needs to be done on this and other classes of inhibitors before the capabilities of the new method are fully characterized. In addition to investigating a variety of molecular systems,

improvements in the computational algorithms are also needed to increase the flexibility and usefulness of this new tool.

Acknowledgements

The authors would like to thank the members of the CADD Unit at Pharmacia & Upjohn, Kalamazoo, Michigan for many helpful discussions; Jeff Howe, the Director of the Unit, for his support, and Jordi Mestres, now at Organon in The Netherlands, for his many contributions to MIMIC, which made this work possible.

References

1. M.A. Johnson and G.M. Maggiora (Eds.), *Concepts and Applications of Molecular Similarity*, John Wiley & Sons, New York, 1990.
2. P. Dean (Ed.), *Molecular Similarity in Drug Design*, Blackie Academic, London, 1995.
3. M.S. Lajiness, "Dissimilarity-Based Compound Selection Techniques," *Perspectives in Drug Discovery and Design* (1997), **7/8**, 65-84.
4. H. Kubinyi, G. Folkers, and Y.C. Martin (Eds.), *3D QSAR in Drug Design: Recent Advances*, Kluwer/ESCOM, The Netherlands, 1998.
5. D.C. Rohrer, "3D Molecular Similarity Modeling in Computational Drug Design," in *Molecular Similarity And Reactivity: From Quantum Chemical to Phenomenological Approaches*, R. Carbó, Ed., Kluwer, The Netherlands, 1995, pp. 141-161.
6. J. Mestres, D.C. Rohrer, and G.M. Maggiora, "MIMIC: A Molecular-Field Matching Program. Exploiting Applicability of Molecular Similarity Approaches, *J. Comput. Chem.* (1997), **18**, 934-954.
7. J. Mestres, D.C. Rohrer, and G.M. Maggiora, "A Molecular Field-Based Similarity Approach to Pharmacophoric Pattern Recognition," *J. Mol. Graphics Model.* (1997), **15**, 114-121.
8. J. Mestres, D.C. Rohrer, and G.M. Maggiora, "A Molecular Field-Based Similarity Study of Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors," *J. Comput.-Aided Mol. Design* (1998), in press.
9. G. Klebe, T. Mietzner, and F. Weber, "Different Approaches Toward an Automatic Structural Alignment of Drug Molecules: Application to Sterol Mimics, Thrombin and Thermolysin Inhibitors," *J. Comput.-Aided Mol. Design* (1994), **8**, 751-778.
10. S.J Smeardon, *et al.*, "Structure of the Binding Site for Nonnucleoside Inhibitors of the Reverse Transcriptase of Human Immunodeficiency Virus Type 1," *Proc. Nat. Acad. Sci. USA* (1994), **91**, 3911-3915.

11. J. Ding, *et al*., "Structure of HIV-1 Reverse Transcriptase in a Complex with the Non-Nucleoside Inhibitor Alpha-APA R95854 at 2.8 Angstroms Resolution," *Structure(London)* (1995), **3**, 365-379.
12. G.E. Kellog, S.F. Semus, and D.J. Abraham, "HINT: A New Method of Empirical Hydrophobic Field Calculation for CoMFA," *J. Comput.-Aided Mol. Design* (1991), **5**, 545-552.
13. Q. Du, G.A. Arteca, and P.G. Mezey, "Heuristic Lipophilicity Potential for Computer-Aided Rational Drug Design," *J. Comput.-Aided Mol. Design* (1997), **11**, 503-515.
14. G.E. Kellog, L.B. Kier, P. Gaillard, and L.H. Hall, "E-State Fields: Applications to 3D QSAR," *J. Comput.-Aided Mol. Design* (1997), **10**, 513-520.
15. J.J.P. Stewart, *MOPAC7 Version2 manual*, QCPE, Bloomington, Indiana, 1993.
16. A.C. Good and W.G. Richards, "Rapid Evaluation of Shape Similarity Using Gaussian Functions," *J. Chem. Inf. Comput. Sci*. (1993), **33**, 112-116.
17. R. Carbó, L. Leyda, and M. Arnau, "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.
18. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, "An All Atom Force Field for simulation of Proteins and Nucleic Acids," *J. Comput. Chem.* (1986), **7**, 230-252.

19. F. Mohamadi, N.G.J. Richards, W.C, Guida, R. Liskamp, C. Caufield, G. Chang, T. Hendrickson, and W.C. Still, "MacroModel: An Integrated Software System for Modeling Organic and Bio-Organic Molecules Using Molecular Mechanics," *J. Comput. Chem.* (1990), **11**, 440-467. [*A single, minor modification of the AMBER* force field was made to the OO-C2-N2-CU torsional parameter, changing its value from 1.5 to 2.5, since the initial value of the trosional parameter yielded an incorrect geometry for the amide nitrogen. The new value corrected this deficiency.*]
20. G. Chang, W.C. Guida, and W.C.Still, "An Internal Coordinate Monte Carlo Method for Searching Conformational Space," *J. Am. Chem. Soc.* (1989), **111**, 4379-4386