Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 6267–6272

Improved 3D-QSAR CoMFA of the dopamine transporter blockers with multiple conformations using the genetic algorithm

Hongbin Yuan and Pavel A. Petukhov*

Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, 833 S. Wood Street, Chicago, IL 60612, USA

Received 3 August 2006; revised 6 September 2006; accepted 7 September 2006
Available online 5 October 2006

Abstract—A 3D-QSAR/CoMFA was performed for a series of 42 piperidine-based dopamine transporter (DAT) blockers. The overall process consisted of three major steps: (1) a pharmacophore model was built using the Genetic Algorithm Similarity Program (GASP); (2) the Flexible Superposition (FlexS) technique was applied to generate multiple conformations for each of the ligands based on the pharmacophore; (3) the Genetic Algorithm was employed to optimize the selection of the ligand conformations for the CoMFA modeling. The CoMFA models were found to be more detailed in the putative binding site by exploring multiple conformations of each ligand. The comparison of the contour maps shows that, in general, these models are comparable and the differences between them result from the ability of the flexible 3α-substituents of the ligands to adopt multiple conformations satisfying the same pharmacophore model. These findings provide guidance for the design and improvement of compounds with DAT activity, which is important for the development of a treatment of cocaine addiction and certain neurological disorders.

Since the development of the Comparative Molecular Field Analysis (CoMFA)[1] as one of the earliest 3D-QSAR techniques, it remains the most popular QSAR method mainly because the graphic results of CoMFA models can provide a direct way to visualize the structure–activity relationship. Meaningful conformations and suitable alignments of lead compounds for building interpretable and predictive models are essential for 3D-QSAR/CoMFA and ligand-based drug design in general. In the absence of the structural data of a ligand–protein complex, the conformational search and the alignment of ligands can be performed using the methods based on the atom RMS fitting, pharmacophore or feature alignments.[2–4] Compared to the scaffold superposition utilizing the atom RMS fitting, the feature alignment approaches (e.g., GASP, DISCO, FlexS, HipHop, and GALAHAD) are more advantageous in aligning flexible and diverse compounds.[5–9] The problem of finding the conformations of ligands and their alignments becomes dramatically more complex with the increase of diversity and flexibility of the chemical structures of ligands. An additional challenge should be conquered before a successful 3D-QSAR model can be constructed—the selection of one out of multiple conformations of flexible molecules that match a desired molecular template or a pharmacophore model.

We have reported a pharmacophore-based CoMFA modeling study of the piperidine-based blockers of the dopamine transporter (DAT).[3] Substantial evidence has suggested that DAT is the key site of action for cocaine in the central nervous system even though the mechanisms that mediate the addictive character of cocaine are more complex.[10–15] DAT and the other monoamine transporters have been studied extensively as the targets for the development of therapies for addiction and neurological diseases,[14,16–18] thus justifying further interest in the monoamine transporter blockers.

Despite the success of our previous CoMFA modeling, it became clear that on one hand the model can benefit from exploring multiple conformations of the ligands; but on the other hand the construction of the models is significantly impeded due to a large number of combinations of different ligand conformations possible for analysis. In an effort to overcome this limitation, we performed a CoMFA modeling of the dopamine transporter blockers utilizing the genetic algorithm (GA) in the

Keywords: 3D-QSAR; CoMFA; Multiple conformations.
* Corresponding author. Tel.: +1 312 996 4174; fax: +1 312 996 7107; e-mail: pap4@uic.edu
URL: http://medchem.pharm.uic.edu/

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.09.037

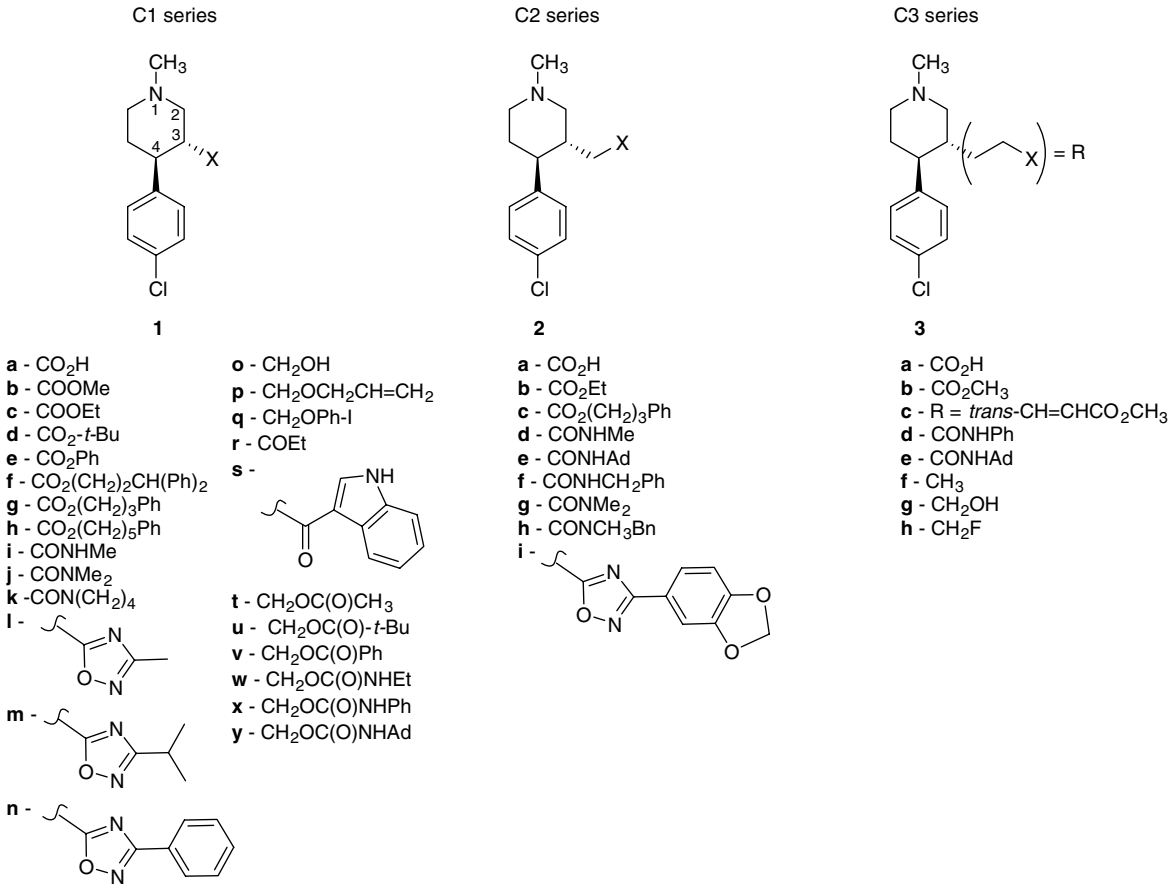

Figure 1. Three series of the piperidine-based cocaine analogues used in the CoMFA studies. The ligands were assigned into the series C1, C2 or C3 according to the number of carbon atoms between the attachment point of the 3α-substituent and its first polar atom.

selection of the ligand conformations. GA is an optimization approach that is based on the biological evolution and natural selection principles.[19,20] In the last decade, GA has been widely used in the modeling and optimization of QSAR for variable selection.[21–25] Herein we present the details of the improved CoMFA by exploring multiple conformations of the ligands.

All molecular modeling studies were performed on a Redhat Linux WS 3.0 computer using the SYBYL 7.0 software package.[26] All compounds were tested for their ability to inhibit high affinity uptake of [^{3}H]DA using rat nerve endings (synaptosomes) obtained from brain regions enriched in DAT. The dataset is identical to that published in our previous report.[3] A total of 42 piperidine-based molecules were divided into three series C1, C2, and C3 according to the number of carbon atoms between the attachment point of the 3α-substituent and the first polar atom in the 3α-substituent (Fig. 1 and Table 1). The MOPAC[27] charges for all the molecules were calculated using the AM1 method.[28]

Three molecules **1p**, **2i**, and **3c** were selected to represent the three series C1, C2, and C3, respectively, in the search of the pharmacophore. The pharmacophore searching step was accomplished using the Genetic Algorithm Similarity Program (GASP).[5] The three ligands (one from each series) were chosen based on the following three criteria. First, a representative molecule should be DAT-active; second, it should have meaningful functional groups, which could be used in the pharmacophore search; and third, the 3α-substituents of these molecules should be relatively rigid with only a few rotatable bonds, thus limiting the number of distinct pharmacophore models obtained for further evaluation by CoMFA modeling. Once the pharmacophore was generated, the three-dimensional structure of **2i**, the largest one among the three molecules selected in the search of the pharmacophore, was used as a template in the structure alignment for all the ligands. This step was performed using an incremental construction algorithm and a scoring function based on intermolecular interactions and overlapping density functions implemented in the Flexible Superposition (FlexS) technique.[6] The number of alignments per ligand was 30. To avoid energetically unfavorable conformations of the compounds, the five lowest energy conformations of each ligand generated by FlexS were kept in its conformational library for further studies.

The steric and electrostatic field energies were calculated using an sp^3 carbon atom with a +1 charge as a probe. Principal Component Analysis (PCA)[29] and Partial Least Squares (PLS) regression[30,31] were performed to correlate the biological data and molecular fields. The maximum number of principal components was set to

Table 1. Observed and predicted DAT activities of the ligands

Compound	DAT, pK_i	Model 1	Model 2	Model 3
1a	3.76	3.70	3.68	3.64
1b	6.63	6.66	6.64	6.72
1c	6.28	6.32	6.29	6.17
1d	6.50	6.49	6.61	6.56
1e*	6.04	5.36	5.48	5.54
1f	5.94	5.88	6.00	5.96
1g	5.85	5.90	5.84	5.79
1h	7.28	7.26	7.30	7.32
1i	5.00	5.09	5.16	5.29
1j	5.67	5.56	5.51	5.56
1k	6.04	5.99	6.03	5.98
1l	6.73	6.79	6.76	6.77
1m	6.41	6.43	6.44	6.32
1n*	6.30	5.53	5.71	5.83
1o	6.30	6.37	6.37	6.40
1p	7.22	7.19	7.23	7.27
1q	6.10	6.10	6.09	6.04
1r*	6.53	6.61	6.59	6.84
1s	6.59	6.60	6.53	6.55
1t	6.22	6.22	6.23	6.29
1u	6.57	6.58	6.55	6.56
1v	6.56	6.58	6.50	6.56
1w	6.74	6.70	6.74	6.76
1x*	7.04	7.04	7.22	7.41
1y	7.25	7.27	7.28	7.30
2a	5.52	5.60	5.57	5.51
2b	7.10	7.13	7.08	7.07
2c	7.44	7.50	7.40	7.44
2d	6.03	6.04	6.08	6.12
2e	6.15	6.15	6.19	6.18
2f	6.57	6.57	6.51	6.54
2g	7.80	7.80	7.75	7.69
2h*	6.78	6.44	6.94	6.73
2i	7.36	7.34	7.41	7.38
3a	5.89	5.88	5.89	5.96
3b*	7.16	7.33	7.31	7.36
3c	7.28	7.25	7.31	7.26
3d	6.70	6.69	6.62	6.65
3e	6.35	6.34	6.37	6.37
3f	7.70	7.55	7.54	7.55
3g	7.58	7.57	7.62	7.63
3h	7.37	7.41	7.39	7.32

The test compounds are labeled with a '*' symbol.

6. Leave-one-out cross-validation was initially utilized to evaluate the predictive capability of the models. PLS procedures without cross-validation were performed to create the CoMFA models. The selection of the ligand conformations by GA was integrated with CoMFA. The code was written in the Sybyl Programming Language (SPL). The quality of the final models was further verified using leave-N-out (10%) cross-validation. A set of external test compounds (**1e**, **1n**, **1r**, **1x**, **2h**, and **3b)** with distinct chemical structures representing the three series of compounds was used to evaluate the CoMFA models in our previous report.[3] For comparison, the same test set was also used in the present study.

The overall strategy of the pharmacophore-driven alignment, generation and selection of the ligand conformations, CoMFA modeling and validation of the final models are shown in Figure 2. One of the two GASP pharmacophores corresponding to the best CoMFA models reported previously was utilized in the present study (Fig. 3).[3] Theoretically any one of the five conformations for each of the 36 training compounds can be selected for the CoMFA modeling. Although the conformations of each ligand were generated by FlexS based on the same pharmacophore and thus very similar, the overall alignments of various conformations for all the ligands may result in significantly different 3D-QSAR/CoMFA models. To explore multiple conformations of the ligands, a genetic algorithm analysis was applied in the selection of the ligand conformations for the CoMFA modeling. The genetic algorithm consisted of the following steps:

Initialization: Generate an initial population P_1 of CoMFA models using one randomly selected conformation from the conformation library of each ligand. The population size was 50.

Repeat:

(1) Crossover. Exchange the conformations of corresponding ligands for any two models in the population P_i. The crossover ratio was set to 50:50.
(2) Mutation. For randomly selected ligands, replace the conformations obtained in step 1 with randomly selected conformations in the library. Store the results as a temporary population P_{tmp}. The mutation rate was set to 0.08.
(3) Selection. Generate new CoMFA models for P_{tmp}. Compare their q^2 values (or other fitness functions) with those generated for population P_i and keep the best models in population P_{i+1}.

Until: The 500 generations limit is reached or the 50 best models remain unchanged for 10 consecutive generations.

To evaluate the performance of the GA analysis, a total of 20 GA runs were performed. The best CoMFA model in each GA run was constructed for comparison. Out of the 20 models the top three models were selected for further analysis (Table 2). The standard errors of estimate for the training compounds by the three models are between 0.053 and 0.088. The standard errors of prediction by the leave-one-out cross-validation for the three models are in the range from 0.255 to 0.265. The values of the leave-one-out q^2 are 0.918, 0.916 and 0.908 for the three models, respectively. Knowing the risk of utilizing the leave-one-out q^2 as a criterion for selecting the best model,[32] the training compounds were further divided into 10 groups and one group of ligands was omitted and predicted each time. The leave-10%-out squared correlation coefficients are between 0.823 and 0.92 for the top three CoMFA models. The contributions from steric and electrostatic fields are similar in all models and range from 0.633 to 0.674 for the steric fields and from 0.326 to 0.367 for the electrostatic fields. Overall, these results indicate that models 1–3 are statistically comparable and consistent to our previously published results.[3]

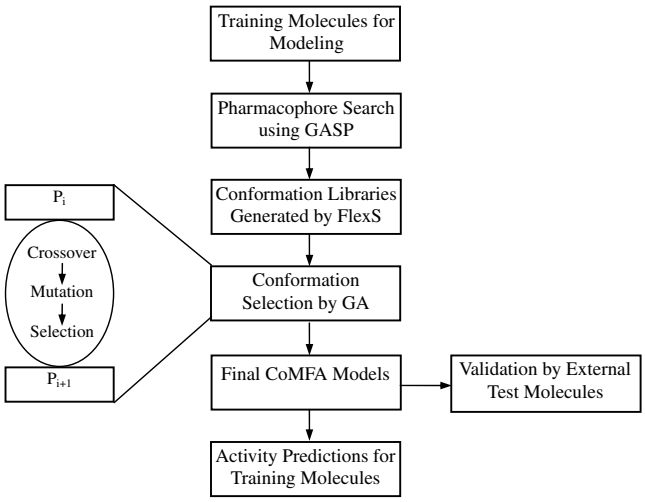

Figure 2. Flowchart of CoMFA integrated with GA.

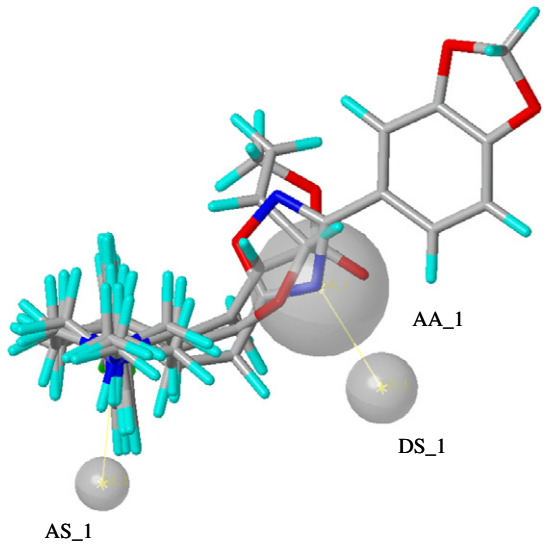

Figure 3. The pharmacophore generated by GASP using **1p**, **2i**, **3c**: AA_1, H-bond acceptor atom in the ligands; AS_1, H-bond acceptor site in the DAT; DS_1, H-bond donor site in the DAT.

Table 2. Summary of statistics and field contributions for the top three CoMFA models

	Model 1	Model 2	Model 3
No. of training compounds	36	36	36
Optimal no. of components	6	6	5
q^2 (Leave-One-Out)	0.918	0.916	0.908
Standard error of prediction	0.255	0.257	0.265
Cross-validation (10%)	0.920	0.879	0.823
r^2	0.996	0.994	0.990
Standard error of estimate	0.053	0.069	0.088
F values	1350.0	803.4	581.2
Probability of $r^2 = 0$	0	0	0
Field contributions			
Steric	0.638	0.674	0.633
Electrostatic	0.362	0.326	0.367
No. of test compounds	6	6	6
r^2 of test set	0.726	0.767	0.753
Standard error of test set	0.265	0.199	0.256

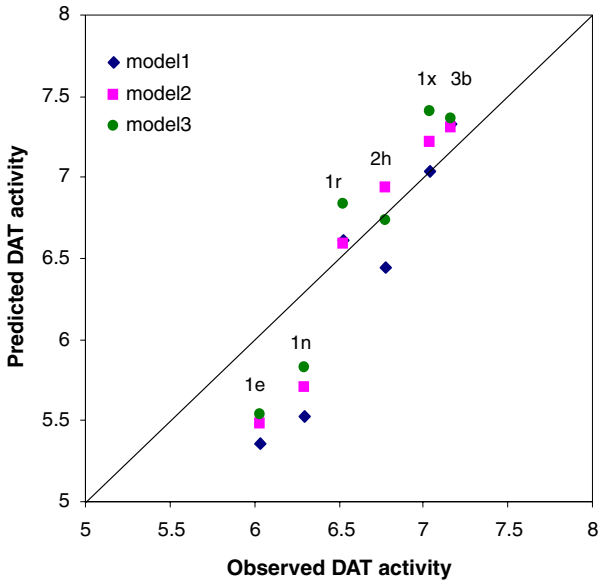

Figure 4. Observed and CoMFA predicted DAT activities (pK_i) of the test compounds.

In a further test the CoMFA models were validated by an external test set of six compounds. Models 1–3 generally performed well in the prediction of the activities of the test ligands, although they were relatively less successful in predicting the activities of ligands **1e** and **1n** (Fig. 4). The comparison of the structures of **1e** and **1n** shows that in both ligands the 3α-substituent is rigid and contains either the phenyl or phenylisoxazole group. Ligands with these structural features are underrepresented in the training set since they were found to be poorly active[33] and the synthetic efforts were quickly stirred toward more active compounds.[34] The r^2 values of the test set for the top three models range from 0.726 to 0.767, which are greater than the acceptable level for the predictive power of QSAR models.[32] The standard errors of prediction for the test compounds are between 0.199 and 0.265, which are consistent with the standard errors of prediction for the leave-one-out cross-validation results.

The contour maps for the top three CoMFA models are illustrated using compound **2i** as a reference (Fig. 5). There is a large steric favorable area (green) behind the piperidine ring and below the 3α-substituent where bulky groups increase the binding affinity. The steric favorable area close to the piperidine ring is consistent with the higher overall DAT blocking activity of the tropane-based compounds as the additional two-carbon bridge in the tropane-based compounds would be positioned close to this steric favorable area. In addition, several steric unfavorable (yellow) areas exist near the 3α-substituent in all three models. The steric unfavorable area on the right side of the 3α-substituent is consistent with the prohibited region proposed earlier.[33] Whereas the majority of the steric features are located near the 3α-substituent, there are also several sterically unfavorable regions and one favorable region (model 1) near the 4-chloro-

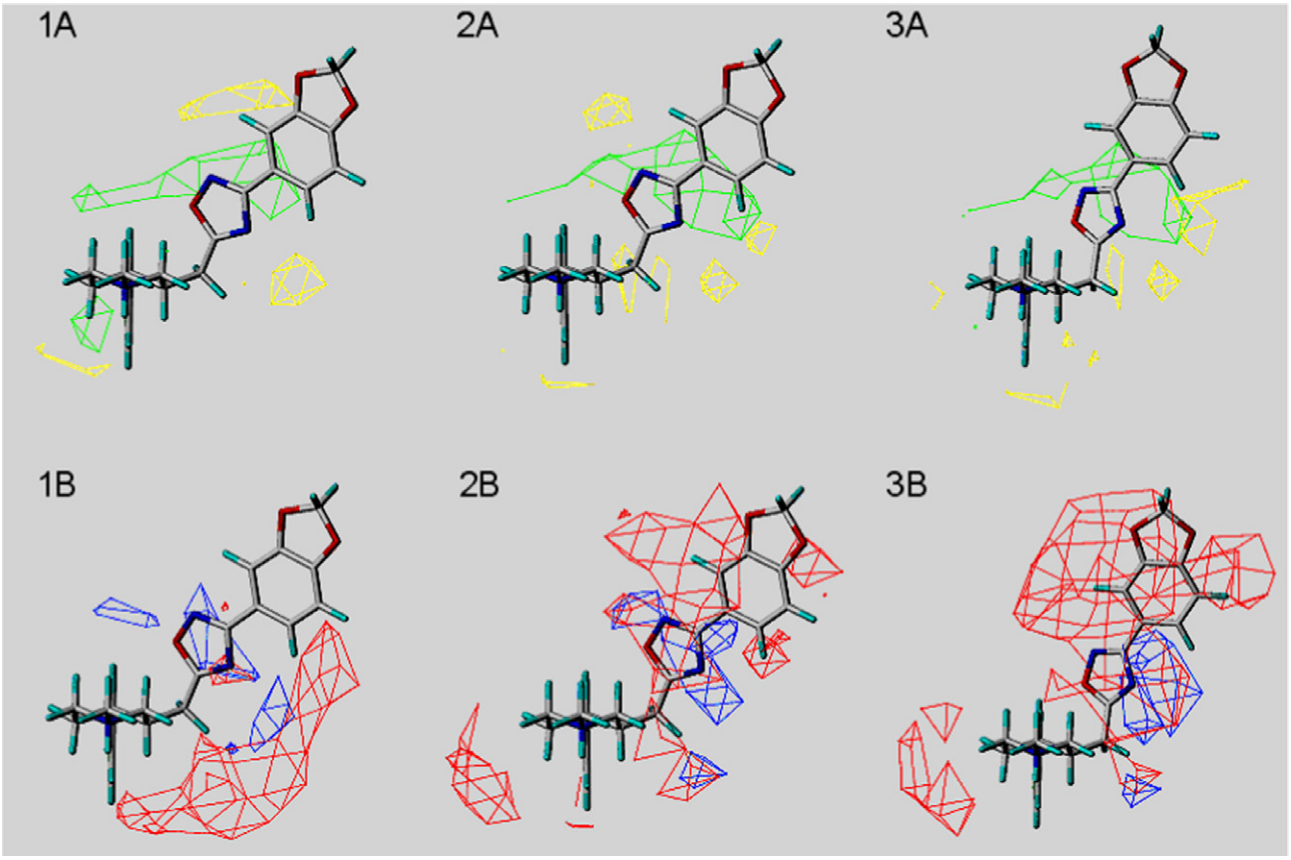

Figure 5. Contour maps of the top three CoMFA models. 1A, 2A and 3A are the steric fields for models 1, 2 and 3, respectively. 1B, 2B and 3B are the electrostatic fields for models 1, 2 and 3, respectively. In the steric contour maps, greater affinity is correlated with more bulky groups near green and less bulky groups near yellow; in the electrostatic contour maps, greater affinity is correlated with more positive charge near blue and more negative charge near red.

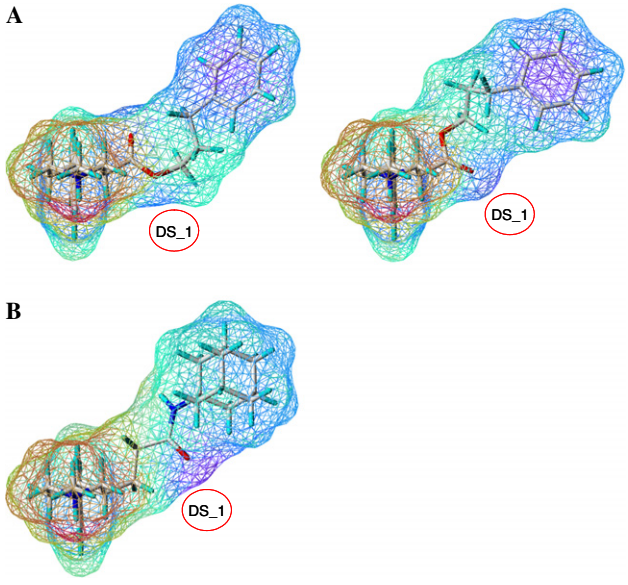

Figure 6. (A) Comparison of the conformations of **1g** in models 1 (left) and 2 (right). (B) The same conformation of **3e** in models 1 and 2. The solvent accessible surfaces of the ligands are colored by electrostatic potential. Note that the opposite conformation of **3e** does not pass the pharmacophore requirements. The DS_1 site corresponds to the position of the H-bond donor site DS_1 in the pharmacophore.

phenyl ring, suggesting that the 4-chlorophenyl ring may be limited in its ability to rotate freely and, in fact, actively participate in the binding. This observation is consistent with the CoMFA models previously reported for the tropane-based ligands.[35–37]

Obvious dissimilarity between model 1 and models 2 and 3 comes from the electrostatic fields. In model 1, several unconnected positive charge favorable areas (blue) are below the 3α-substituent, whereas a large area (red) favorable to negative charge exists on the right side of the 3α-substituent. In models 2 and 3, the large negative charge favorable area is positioned on the top and both sides of the 3α-substituent. Models 2 and 3 have an extra negative charge favorable area (red) on the left side of the piperidine ring compared to model 1.

Taking into account the differences between the models one may ask a question: 'Why the CoMFA fields in the models based on the same pharmacophore are different?' We believe that the answer to this question lies in the phenomenon of the multiple interaction points possible for the majority of the polar functional groups and the fundamental differences between the steric and electrostatic fields. Unlike the electrostatic interactions, the ste-

ric bulk is virtually non-directional. For instance, the shape of the solvent accessible area of the 3α-substituent in ligand **3e** is close to a cylinder with a C_∞ axis of symmetry (Fig. 6B). From the viewpoint of the steric bulk all sides of the substituent are similar. The electrostatic potential of the amide group in ligand **3e**, on the other hand, has a positive/negative alternating electrostatic field pattern that has no symmetry at all. Clearly, the two orientations of the amide group C=O—up, NH—down, and C=O—down, NH—up would have similar solvent accessible area profiles but different electrostatic potential profiles. On the other hand, despite the difference in the orientation of the ester group in ligand **1g** the two conformations are sterically and electronically similar from the viewpoint of the pharmacophore model used to generate the conformations (Fig. 6A). In fact, the ligand polar groups may have multiple favorable ways in which they can interact with receptors.[38] The ability of the 3α-substituent to adopt conformations that are similar in shape but different electronically becomes more evident taking into account that the majority of the ligands have relatively flexible 3α-substituents. Therefore, even for the same ligand the conformation library may contain entities that would have similar steric but overall different electrostatic CoMFA fields. Multiple methods to validate the statistical quality and stability of CoMFA models are available.[39,40] The applicability of these validation methods is limited to the large datasets with densely populated chemical space that are typically not available on the early stages of the medicinal chemistry efforts. Whereas we cannot justify which of the models should receive the preference, the presence of two or more statistically acceptable models may indicate that the additional modifications of the ligands should be planned to rigidify the substituents in those conformations that were suggested by the different models.

In summary, an improved CoMFA of the dopamine transporter blockers was performed by exploring the multiple conformations of each ligand using the genetic algorithm. Three CoMFA models with the comparable statistical results were identified based on the same pharmacophore. The comparison of their contour maps suggests that both the similarity in the steric fields and the differences in the electrostatic fields among these models are due to the fact that the flexible 3α-substituents abiding the pharmacophore requirements can adopt the conformations similar in shape but different electronically.

References and notes

1. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
2. Bostrom, J.; Bohm, M.; Gundertofte, K.; Klebe, G. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1020.
3. Yuan, H.; Kozikowski, A. P.; Petukhov, P. A. *J. Med. Chem.* **2004**, *47*, 6137.
4. Lemmen, C.; Lengauer, T. *J. Comput. Aided Mol. Des.* **2000**, *14*, 215.
5. Jones, G.; Willett, P.; Glen, R. C. *J. Comput. Aided Mol. Des.* **1995**, *9*, 532.
6. Lemmen, C.; Lengauer, T.; Klebe, G. *J. Med. Chem.* **1998**, *41*, 4502.
7. Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. *J. Comput. Aided Mol. Des.* **1993**, *7*, 83.
8. Wellsow, J.; Machulla, H. J.; Kovar, K. A. *Quant. Struct.—Act. Relat.* **2002**, *21*, 577.
9. Krovat, E. M.; Langer, T. *J. Med. Chem.* **2003**, *46*, 716.
10. Kuhar, M. J.; Ritz, M. C.; Boja, J. W. *Trends Neurosci.* **1991**, *14*, 299.
11. Carroll, F. I.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. *J. Med. Chem.* **1992**, *35*, 969.
12. Boja, J. W.; Vaughan, R.; Patel, A.; Shaya, E. K.; Kuhar, M. J. The dopamine transporter. In *Dopamine Receptors and Transporters*; Niznik, H. B., Ed.; Marcel Dekker: New York, 1994; pp 611–644.
13. Schenk, S. *Pharmacol. Biochem. Behav.* **2000**, *67*, 363.
14. Singh, S. *Chem. Rev.* **2000**, *100*, 925.
15. Castanon, N.; Sceare-Levie, K.; Lucas, J. J.; Rocha, B.; Hen, R. *Pharmacol. Biochem. Behav.* **2000**, *67*, 559.
16. Carroll, F. I. *J. Med. Chem.* **2003**, *46*, 1775.
17. Newman, A. H.; Kulkarni, S. *Med. Res. Rev.* **2002**, *22*, 429.
18. Uhl, G. R.; Hall, F. S.; Sora, I. *Mol. Psychiatry* **2002**, *7*, 21.
19. Holland, J. H. *Sci. Am.* **1992**, *267*, 66.
20. Forrest, S. *Science* **1993**, *261*, 872.
21. Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B. Q.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. *J. Am. Chem. Soc.* **1997**, *119*, 10509.
22. Hasegawa, K.; Miyashita, Y.; Funatsu, K. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 306.
23. Hasegawa, K.; Kimura, T.; Funatsu, K. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 112.
24. Turner, D. B.; Willett, P. *J. Comput. Aided Mol. Des.* **2000**, *14*, 1.
25. Yuan, H.; Parrill, A. L. *Bioorg. Med. Chem.* **2002**, *10*, 4169.
26. SYBYL (Version 7.0), Tripos, Inc., 1699 South Hanley Road, St. Louis, MO 63144.
27. MOPAC [QCPE No. 455], Quantum Chemistry Exchange Program, Indiana University, Creative Arts Bldg. 181, Bloomington, IN 47405; 6.0 ed.
28. Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
29. Cramer, R. D. *J. Am. Chem. Soc.* **1980**, *102*, 1837.
30. Dunn, W. J., 3rd; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. *Quant. Struct.—Act. Relat.* **1984**, *3*, 131.
31. Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. *Quant. Struct.—Act. Relat.* **1988**, *7*, 18.
32. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Model.* **2002**, *20*, 269.
33. Petukhov, P. A.; Zhang, M.; Johnson, K. J.; Tella, S. R.; Kozikowski, A. P. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2079.
34. Petukhov, P. A.; Zhang, J.; Wang, C. Z.; Ye, Y. P.; Johnson, K. M.; Kozikowski, A. P. *J. Med. Chem.* **2004**, *47*, 3009.
35. Carroll, F. I.; Mascarella, S. W.; Kuzemko, M. A.; Gao, Y.; Abraham, P.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. *J. Med. Chem.* **1994**, *37*, 2865.
36. Lieske, S. F.; Yang, B.; Eldefrawi, M. E.; MacKerell, A. D., Jr.; Wright, J. *J. Med. Chem.* **1998**, *41*, 864.
37. Newman, A. H.; Izenwasser, S.; Robarge, M. J.; Kline, R. H. *J. Med. Chem.* **1999**, *42*, 3502.
38. Taylor, R. *Acta Crystallogr. Section D* **2002**, *58*, 879.
39. Clark, R. D.; Sprous, D. G.; Leonard, J. M. *Prous Science SA* **2001**, 475.
40. Golbraikh, A.; Shen, M.; Xiao, Z. Y.; Xiao, Y. D.; Lee, K. H.; Tropsha, A. *J. Comput. Aided Mol. Des.* **2003**, *17*, 241.