Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 3009–3014

Neuraminidase pharmacophore model derived from diverse classes of inhibitors

Jian Zhang,[a] KunQian Yu,[a] Weiliang Zhu[a,*] and Hualiang Jiang[a,b,*]

[a]*Drug Discovery and Design Centre, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Shanghai Institutes for Biological Sciences, and Graduate School, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China*
[b]*School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China*

Received 13 December 2005; revised 3 February 2006; accepted 21 February 2006
Available online 10 March 2006

Abstract—A three-dimensional pharmacophore model was developed based on 22 currently available inhibitors, which were carefully selected with great diversity in both molecular structure and bioactivity, for discovering new potent neuraminidase (NA) inhibitors to fight against avian influenza virus. The best hypothesis (Hypo1), consisting of five features, namely, one positive ionizable group, one negative ionizable group, one hydrophobic point, and two hydrogen-bond donors, has a correlation coefficient of 0.902, a root mean square deviation of 1.392, and a cost difference of 72.88, suggesting that a highly predictive pharmacophore model was successfully obtained. The application of the model shows great success in predicting the activities of 88 known NA inhibitors in our test set with a correlation coefficient of 0.818 with a cross-validation of 98% confidence level. Accordingly, our model should be reliable in identifying structurally diverse compounds with desired biological activity.

An influenza virus, called avian influenza virus (AIV) that should infect only avian species, was found to infect humans, causing acute disease and rapid death.[1] Despite considerable knowledge of viral infectivity, no therapeutic measure is highly and specifically effective in controlling this disease. Studies showed that neuraminidase (NA), a glycoprotein embedded in the viral envelope, plays a key role at the final stage of infection when NA cleaves sialic acid from cell surface and progeny virions facilitating virus release from infected cells.[2,3] When the influenza virus is deficient in NA activity, virus progeny aggregate at the surface of the infected cell, severely impairing further spread of viruses to other cells. Therefore, NA has been recognized as a main target for developing agents against AIV infection.

Over the last two decades, a number of classes of NA inhibitors have been developed and showed to be somewhat effective in controlling AIV infection in humans.[4–11] Zanamivir (from GlaxoSmithKline and Biota)[12,13] and oseltamivir (from Hoffman La Roche and Gilead Sciences)[14] have been approved by FDA for the treatment and prevention of the influenza, but were recently questioned about recovery rate. Therefore, developing drugs against AIV is of crucial importance. Our literature survey revealed that NA inhibitors in different classes possess different scaffolds (Fig. 1). Thus, quantitative structure–activity relationship (QSAR) for different classes of inhibitors could be useful in digging out valuable information for developing new potent NA inhibitors. It is widely accepted that pharmacophore model is a well-behaved approach to quantitatively explore common chemical characteristics among a considerable number of structures with great diversity, and qualified pharmacophore model could also be used as a query for searching chemical databases to find new chemical entities. Theodora et al. developed a pharmacophore model based on 18 NA inhibitors, but without statistical validation.[15] Recently, many new inhibitors were discovered with great diversity in both chemical structure and biological activity.[4–11] As the disease might be widely spread with high possibility to kill millions of people, there is an urgent request to discover potent NA inhibitors as drugs to prevent or to cure the

Keywords: Avian influenza virus; Neuraminidase; Inhibitor; Pharmacophore model; QSAR.
* Corresponding author. Tel.: +86 21 50807188; fax: +86 21 50807088 (W.Z.); e-mail addresses: wlzhu@mail.shcnc.ac.cn; hljiang@mail.shcnc.ac.cn

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.02.054

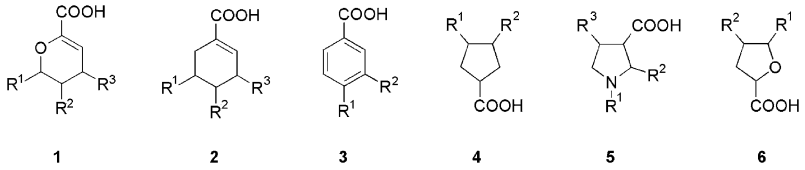

Figure 1. Diverse scaffold for NA inhibitors.[4–11]

virus. Therefore, new information about structure–activity relationship based on more compounds with greater diversity in their structures and activities should be helpful in discovering new NA inhibitors for curing the disease. In this study, we present a hypothetical image of the primary pharmacophore features responsible for the bioactivity of six different classes of NA inhibitors using the Catalyst software (Fig. 1),[16–18] with aims to understand the inhibitory mode of various classes of NA inhibitors and to discover new anti-AIV drug leads.

Materials. A set of 110 different compounds has been collected from difference Refs. 4–11 and 14, which could be classified into six different classes (Fig. 1). We selected 22 compounds as training set with the following two rules: (a) both training and test sets should have structures from each class of compounds to ensure structural diversity. If one class has only one compound, it was assigned to the training set (e.g., compound **19**); (b) both training and test sets should cover the molecular bioactivities (IC_{50}) as wide as possible. If there is only one compound with maximum or minimum order of bioactivity in a class, this compound was assigned to the training set. The training set consists of 22 compounds with significant structural diversity and wide coverage of molecular bioactivities in terms of IC_{50} ranging from 0.75 nM to 250 μM (Fig. 2 and Table 1). Both Zanamivir (compound **18**) and GS4071 (compound **3**) are included in the training set. To validate our pharmacophore hypothesis, 88 compounds with available IC_{50} values were used as a test set (Table S1 and Fig. S1).[4–9,11,14]

The compounds were built using Catalyst 2D–3D sketcher,[16] and a family of representative conformations was generated for each compound using the best conformational analysis method with Poling algorithm[19] and CHARMM forcefield parameters.[20] A maximum number of 250 conformations of each compound were selected using 'best conformer generation' option with a constraint of 20 kcal/mol energy thresholds above the global energy minimum to ensure maximum coverage of the conformational space.

Pharmacophore generation. Based on the conformations for each compound, Catalyst4.10 software package[15] was employed to construct possible pharmacophore models. When generating a hypothesis, catalyst attempts to minimize a cost function consisting of two terms. One penalizes the deviation between the estimated activities of the training set molecules and their experimental values; the other penalizes the complexity of the hypothesis. Uncertainty influences the first step, called the constructive phase, in the hypothesis generating process.[21] The default uncertainty value of 3 was used for the compound activity, representing the ratio of the uncertainty range of measured biological activity against the actual activity for each compound. Analysis of functional groups on each compound in the training set revealed that five chemical features, viz., hydrogen-bond acceptor (HA), hydrogen-bond donor (HD), hydrophobic group (HY), positive ionizable point (PI), and negative ionizable point (NI), could effectively map all of the critical chemical features. Hence, the five features were selected to form the essential information in this hypothesis generation process.

HypoGen mode. The best predictive hypothesis (Hypo1), produced by HypoGen encoded in Catalyst4.10, has five features: one positive ionizable feature, one negative ionizable feature, one hydrophobic point, and two hydrogen-bond donor (Fig. 3), which was characterized by the highest cost difference, the lowest rms divergence, and the best correlation coefficient. Remarkably, the three highest active compounds (compounds **22**, **18**, and **3** in Table 1) can be nicely mapped onto the Hypo1 model by the best fit values, which are shown in Figures 4A–C, indicating that the Hypo1 model provides reasonable pharmacophoric characteristics of the NA inhibitors for the components of their activities.

Cost analysis. In addition to generating a hypothesis, catalyst also provides two numbers to help assessing the validity of the hypothesis. The first is the cost of an ideal hypothesis, which is a lower bound on the cost of the simplest possible hypothesis that still fits the data perfectly. The second is the cost of the Null hypothesis, which presumes that there is no statistically significant structure in the data, and that the experimental activities are normally distributed about their mean. A generated hypothesis with a score that is substantially below that of the Null hypothesis is likely to be statistically significant and bears visual inspection. The greater the difference between the score of the generated hypothesis and the score of the Null hypothesis, the less likely it is that the hypothesis reflects a chance correlation. The total fixed cost of the run is 87.99, the cost of the Null hypothesis is 186.22, and the total cost of the Hypo1 is 113.34. Then, the cost range between Hypo1 and the fixed cost is 25.35, while that between the Null hypothesis and Hypo1 is 72.88. Noticeably, the total cost of Hypo1 was much closer to the fixed cost than to the Null cost. Furthermore, a high correlation coefficient of 0.902 was observed with rms value of 1.392 and the configuration cost of 12.869, demonstrating that we have successfully developed a reliable pharmacophore model with high predictivity.

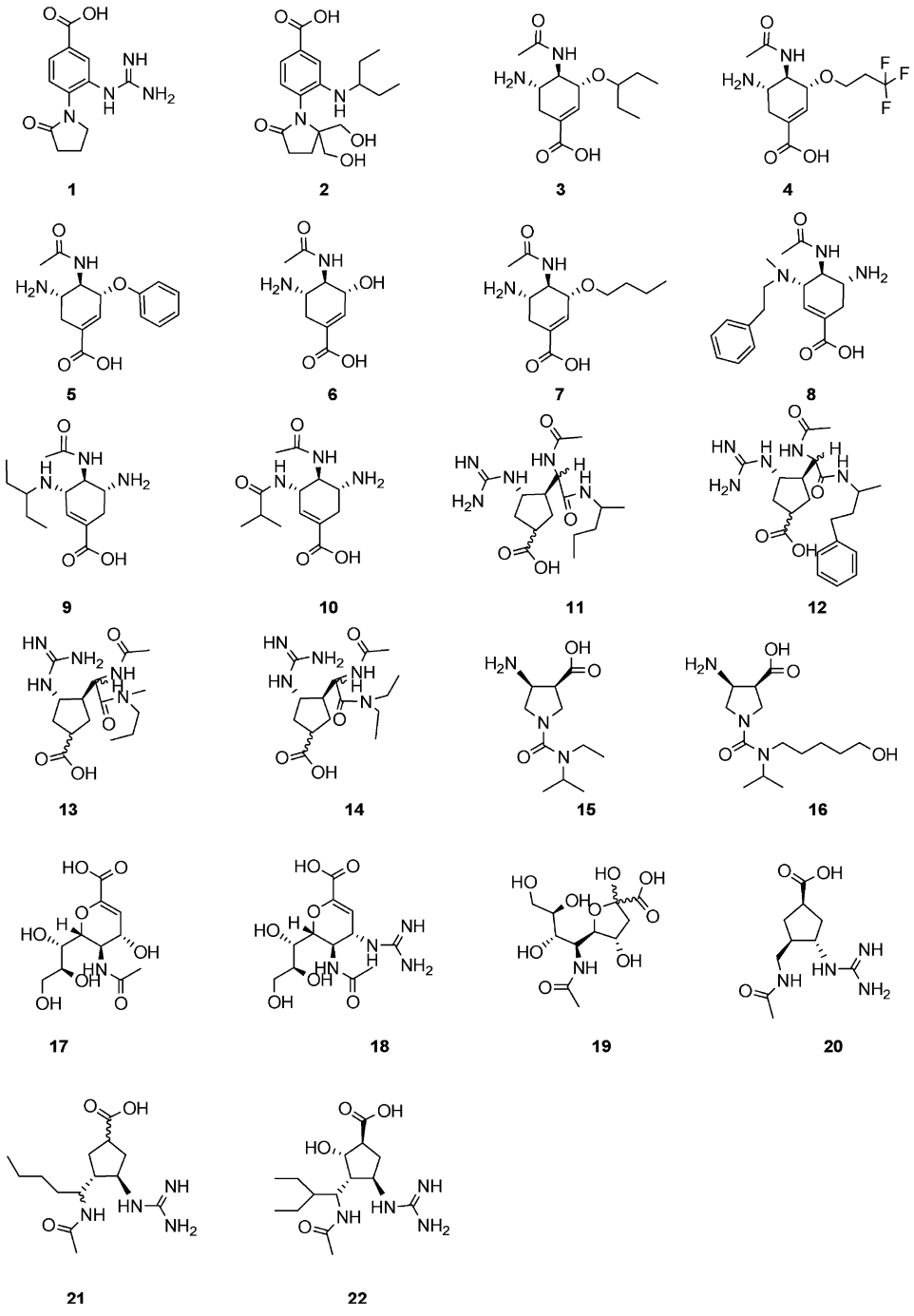

Figure 2. Chemical structures of the 22 training set molecules applied to HypoGen pharmacophore generation.

Score hypothesis. To verify Hypo1's discriminability among NA inhibitors with different order of magnitude activity, all training set compounds were classified by their activity as highly active (<100 nM, +++), moderately active (100–10,000 nM, ++), and inactive (>10,000 nM, +). The actual and estimated NA inhibitor activities of the 22 compounds based on Hypo1 are listed in Table 1. All the compounds, except for compounds **14**, **15**, and **19**, were classified correctly (Table 1). The discrepancy between the actual and estimated activity observed for the three compounds was only about 1 order of magnitude, which might be an artifact of the program that uses different number of degrees of freedom for these compounds to mismatch the pharmacophore model. Interestingly, all highly active compounds map the second hydrogen-bond donor feature (HD2), and, with few exceptions, they also map the positive ionizable feature (PI), revealing that these two features should be mainly responsible for the high molecular bioactivity, thus, should be taken into account in discovering or designing novel NA inhibitors.

Validation of the constructed pharmacophore model. The 88 compounds in the test set were mapped onto the best pharmacophore hypothesis Hypo1, and the actual activity versus estimated activity is shown in Table S1 in the

Table 1. Output of the score hypothesis process on the training set

Compound (Reference)	True IC_{50} (nM)	Estimated IC_{50} (nM)	Error factor[c]	Fit value[d]	Activity scale[e]	Estimated activity scale	Mapped features				
							HD1	HD2	HY	NI	PI
1 (4)	250,000	170,000	−1.5	6.82	+	+	–	+	+	+	–
2 (4)	48	22	−2.2	10.70	+++	+++	+	+	+	+	–
3[a] (5)	1	28	28	10.60	+++	+++	–	+	+	+	+
4 (5)	225	690	3.1	9.21	++	++	+	+	+	+	–
5 (5)	530	200	−2.7	9.75	++	++	+	+	+	+	–
6 (6)	6300	2900	−2.1	8.58	++	++	+	+	–	+	+
7 (6)	300	860	2.9	9.12	++	++	+	+	+	+	–
8 (7)	100	51	−2	10.34	+++	+++	–	+	+	+	+
9 (7)	11	31	2.8	10.55	+++	+++	–	+	+	+	+
10 (7)	6400	880	−7.2	9.10	++	++	+	+	+	+	–
11 (8)	3200	1600	−1.9	8.83	++	++	+	–	+	+	+
12 (8)	230	420	1.8	9.43	++	++	+	+	+	+	–
13 (8)	940	1100	−1.2	8.99	++	++	–	+	+	+	+
14 (8)	15	170	12	9.81	+++	++	–	+	+	+	+
15 (9)	1600	39,000	24	7.46	++	+	–	–	+	+	+
16 (9)	2000	800	−2.5	9.14	+	+	+	–	+	+	+
17 (10)	10,000	6000	−1.7	9.92	++	++	+	+	–	+	–
18[b] (10)	1.3	5.7	9	10.98	+++	+++	+	+	–	+	+
19 (10)	40,000	7100	−5.6	8.19	+	++	+	+	–	+	–
20 (10)	115,000	18,000	−6.4	7.81	+	+	+	–	–	+	+
21 (10)	100	76	−1.3	10.17	+++	+++	–	+	+	+	+
22 (10)	0.75	0.24	3.2	12.67	+++	+++	+	+	+	+	+

[a] GS4071.
[b] Zanamivir.
[c] The error factor is computed as the ratio of the measured activity to the activity estimated by the hypothesis or the inverse if estimated is greater than measured.
[d] Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule.
[e] Activity scale: +++, IC_{50} < 100 nM (highly active); ++, 10,000 nM > IC_{50} > 100 nM (moderately active); +, IC_{50} > 10,000 nM (inactive).

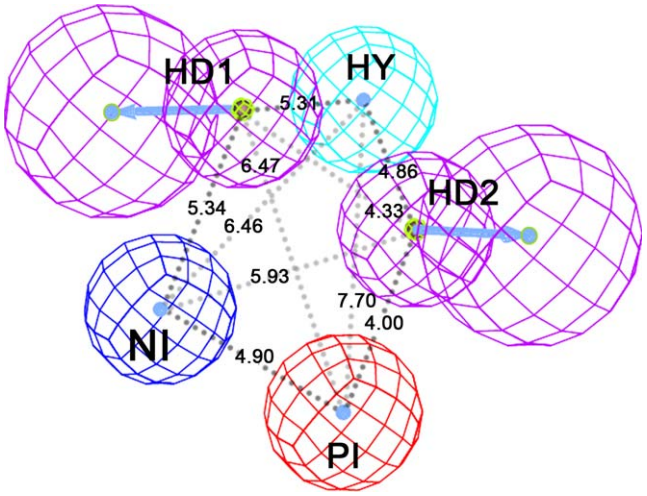

Figure 3. The best hypothesis model Hypo1 produced by the HypoGen module in Catalyst4.10 software package.[15] Pharmacophore features are color-coded with light-blue for hydrophobic groups, red for positive ionizable group, blue for negative ionizable, and magenta for hydrogen-bond donor. Distance between pharmacophore features is reported in angstroms. HY, hydrophobic group; HD1, hydrogen-bond donor 1; HD2, hydrogen-bond donor 2; PI, positive ionizable group; NI, negative ionizable group.

supporting information. A correlation coefficient of 0.818 generated using the test set compounds shows a good correlation between the actual and estimated activities. Detailedly, 18 of 20 highly active, 38 of 47 moderately active, and 15 of 21 inactive compounds were predicted correctly. Highly active compounds **45** and **54** were underestimated as moderately active; five moderately active compounds were underestimated as inactive; four moderately active and six inactive compounds were overestimated as false positive. In summary, all the compounds in the test set were predicted correctly or better than their actual activity, with the exception of seven compounds that were underestimated.

The most potent compound **59** in the test set mapping on the Hypo1 is shown in Figure 4D. The conformation of the compound fits HD2, PI, NI, and HY features, which is in good accordance with the assumption based on the training set that HD2 and PI are essential to high activity in inhibiting NA.

Fisher test. To further evaluate the statistical relevance of the model, the Fischer method[22] was applied. With the aid of the CatScramble program, the experimental activities in the training set were scrambled randomly, and the resulting training set was used for a HypoGen run. All parameters were adopted from the initial HypoGen calculation. This procedure was reiterated 49 times. None of the outcome hypotheses has lower cost score than the initial hypothesis. Table 2 lists the 10 lowest total cost values of the resulting 49 hypothesis. Accordingly, this result indicates that there is a 98% chance for the best hypothesis to represent a true correlation in the training set activity data.[15,23–25]

In conclusion, a ligand-based computational approach was employed to identify molecular structure requirements as effective neuraminidase inhibitors, for discov-

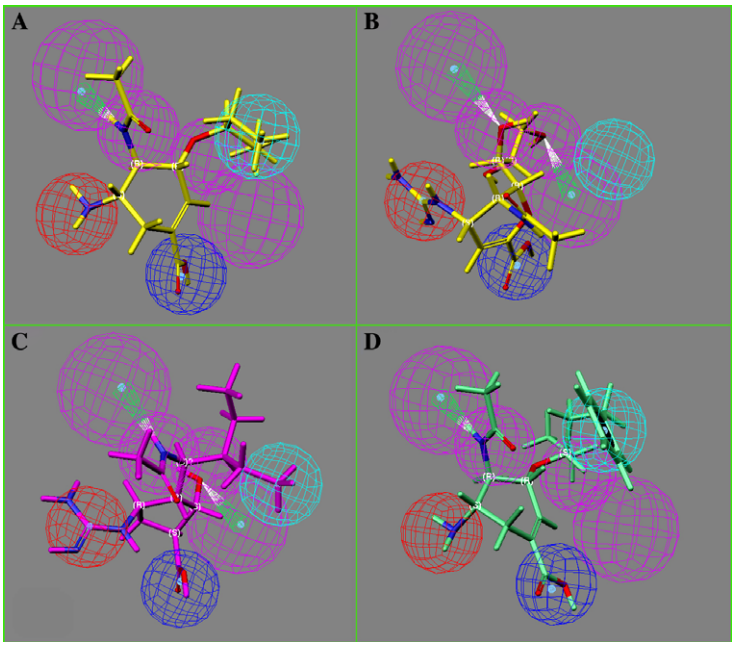

Figure 4. Mapping of the four most highly active compounds on the best hypothesis model Hypo1. (A) Compound **3** (GS4071); (B) compound **18** (Zanamivir); (C) compound **22**; (D) compound **59**. Pharmacophore features are color-coded with the same as Figure 3.

Table 2. Output parameters of the 10 lowest cost hypothesis resulting from the statistical evaluation procedure according to the Fischer method[22]

Hypothesis	Correlation	rms	Total cost
1	0.756	2.091	137.719
2	0.745	2.133	138.920
3	0.745	2.131	139.024
4	0.717	2.226	139.166
5	0.758	2.101	139.561
6	0.688	2.320	144.141
7	0.677	2.349	145.000
8	0.687	2.324	145.804
9	0.710	2.247	145.812
10	0.695	2.297	147.123
Hypo1	0.902	1.392	113.340

ering drugs to prevent and cure avian influenza virus (AIV). A highly predictive pharmacophore model was generated based on 22 training set compounds, which consists of two hydrogen-bond donors, one hydrophobic point, one positive ionizable group, and one negative ionizable group. The utility of our pharmacophore model on 88 test set compounds showed that the model is able to accurately differentiate various classes of NA inhibitors. Thus, our pharmacophore model should be helpful in identifying novel lead compounds with improved inhibitory activity through 3D database searches and useful to designing novel NA inhibitors.

Acknowledgments

This work was supported by grants from the State Key Program of Basic Research of China (2004CB518901) and Shanghai Science and Technology Commission (03DZ19228, 05JC14092, and 05QMX1464).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2006.02.054.

References and notes

1. John, S. T. *Vaccine* **2002**, *20*, 77.
2. Air, G. M.; Laver, W. G. *Proteins* **1989**, *6*, 341.
3. Bucher, D. J.; Palese, P. *The Biologically Active Proteins of Influenza Virus: Neuraminidase*; Academic: New York, NY, 1975, p 83.
4. Atigadda, V. R.; Brouillette, W. J.; Duarte, F.; Ali, S. M.; Babu, Y. S.; Bantia, S.; Chand, P.; Chu, N.; Montgomery, J. A.; Walsh, D. A.; Sudbeck, E. A.; Finley, J.; Luo, M.; Air, G. M.; Laver, G. M. *J. Med. Chem.* **1999**, *42*, 2332.
5. Williams, M. A.; Lew, W.; Mendel, D. B.; Tai, C. Y.; Escarpe, P. A.; Laver, W. G.; Stevens, R. C.; Kim, C. U. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 1837.
6. Kim, C. U.; Lew, W.; Williams, M. S.; Wu, H.; Zhang, L.; Chen, X.; Escarpe, P. A.; Mendel, D. B.; Laver, W. G.; Stevens, R. C. *J. Med. Chem.* **1998**, *41*, 2451.
7. Lew, W.; Wu, H.; Mendel, D. B.; Escarpe, P. A.; Chen, X.; Laver, G.; Graves, B. J.; Kim, C. U. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3321.
8. Chand, P.; Babu, Y. S.; Bantia, S.; Rowland, S.; Dehghani, A.; Kotian, P. L.; Hutchison, T. L.; Ali, S.; Brouillette, W.; El-Kattan, Y.; Lin, T.-H. *J. Med. Chem.* **2004**, *47*, 1919.
9. Wang, G. T.; Chen, Y.; Wang, S.; Gentles, R.; Sowin, T.; Kati, W.; Muchmore, S.; Giranda, V.; Stewart, K.; Sham, H.; Kempf, D.; Laver, W. G. *J. Med. Chem.* **2001**, *44*, 1192.
10. Babu, Y. S.; Chand, P.; Bantia, S.; Kotian, P.; Dehghani, A.; El-Kattan, Y.; Lin, T.-H.; Hutchison, T. L.; Elliott, A. J.; Parker, C. D.; Ananth, S. L.; Horn, L. L.; Laver, G. W.; Montgomery, J. A. *J. Med. Chem.* **2000**, *43*, 3482.

11. Chand, P.; Babu, Y. S.; Bantia, S. *J. Med. Chem.* **1997**, *40*, 4030.
12. Taylor, N. R.; von Itzstein, M. *J. Med. Chem.* **1994**, *37*, 616.
13. von Itzstein, M.; Wu, W.-Y.; Jin, B. *Carbohydr. Res.* **1994**, *259*, 301.
14. Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H. T.; Zhang, L. J.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Steven, R. C. *J. Am. Chem. Soc.* **1997**, *119*, 681.
15. Theodora, S.; Thierry, L. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 184.
16. CATALYST 4.10.; Accelrys, Inc., San Diego, CA, 2005, <http://www.accelrys.com/>.
17. Kurogi, Y.; Güner, O. F. *Curr. Med. Chem.* **2001**, *8*, 1035.
18. Oliver, F. F.; Viktor, K.; Jan, D.; Thierry, L. *J. Med. Chem.* **2004**, *47*, 2750.
19. Smellie, A.; Teig, S.; Towbin, P. *J. Comput. Chem.* **1994**, *16*, 171.
20. Brooks, B. R.; Brucolleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J. *J. Comput. Chem.* **1983**, *4*, 187.
21. Güner, O. F. In *Pharmacophore Perception Development and Use in Drug Design*; International University Line: La Jolla, CA, 2000; p 173.
22. Fischer, R. *The Principle of Experimentation Illustrated by a Psycho-Physical ExpeHafner Publishing Co*, 8th ed.; Hafner Publishing: New York, 1966; Chapter II.
23. Krovat, E. M.; Langer, T. *J. Med. Chem.* **2003**, *46*, 716.
24. Tafi, A.; Costi, R.; Botta, M.; Di Santo, R.; Corelli, F., et al. *J. Med. Chem.* **2002**, *45*, 2720.
25. Funk, O. F.; Kettmann, V.; Drimal, J.; Langer, T. *J. Med. Chem.* **2004**, *47*, 2750.