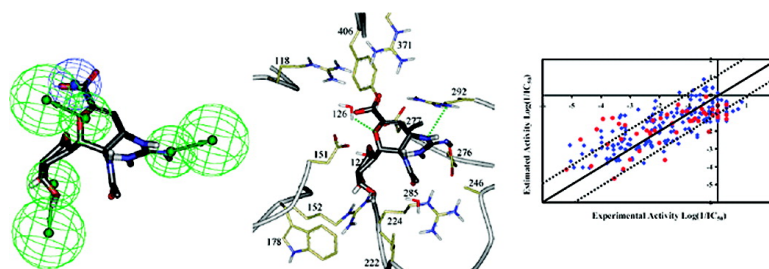


Pharmacophore Modeling, Quantitative Structure#Activity Relationship Analysis, and Shape-Complemented *in Silico* Screening Allow Access to Novel Influenza Neuraminidase Inhibitors

Areej M. Abu Hammad, and Mutasem O. Taha

J. Chem. Inf. Model., **2009**, 49 (4), 978-996 • DOI: 10.1021/ci8003682 • Publication Date (Web): 02 April 2009

Downloaded from <http://pubs.acs.org> on May 2, 2009



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)

Pharmacophore Modeling, Quantitative Structure–Activity Relationship Analysis, and Shape-Complemented *in Silico* Screening Allow Access to Novel Influenza Neuraminidase Inhibitors

Areej M. Abu Hammad and Mutasem O. Taha*

Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman 11942, Jordan

Received October 6, 2008

Neuraminidase (NA) enzyme is one of the valid targets against influenza viruses. With this in mind, the pharmacophoric space of influenza NA was explored using three sets of diverse inhibitors. Subsequently, genetic algorithm and multiple linear regression analysis were employed to select optimal combinations of pharmacophoric models and 2D descriptors capable of yielding self-consistent and predictive quantitative structure–activity relationships (QSARs) against 181 training compounds. The optimal QSAR equations were validated against 43 external test compounds with r^2_{PRESS} values ranging from 0.488 to 0.591. Interestingly, five orthogonal pharmacophores emerged in the optimal QSAR equations suggesting the existence of several distinct ligand/NA binding modes within the NA binding pocket. The resulting pharmacophores were complemented with tight shape constraints and employed as three-dimensional (3D) search queries against the National Cancer Institute (NCI) list of compounds. Several hits exhibited potent inhibitory activities against NA. The highest ranking hit demonstrated an *in vitro* IC_{50} value of 1.8 μM . Docking studies supported the binding modes suggested by our pharmacophore/QSAR analysis.

1. INTRODUCTION

Flu epidemics in humans continue to appear each year due to ineffective control by either vaccines or current antiviral drugs.¹ Recently, many world countries have begun stockpiling the influenza drug Tamiflu to protect their populations against a possible bird flu pandemic.^{2,3}

To develop new antiviral therapies against influenza virus, many viral functions have been selected as targets.⁴ The viral enzyme neuraminidase (NA) has been an active research area for anti-influenza therapy. NA is one of two glycoproteins expressed at the surface of the flu virus and is responsible for viral release from infected cells.^{5,6} Therefore, compounds that inhibit NA should protect the host from viral infection and retard its propagation.⁷ Fortunately, the amino acid residues of the active site of NA are conserved across wild-type influenza viruses,^{8,9} which promote NA as an excellent target for broad-spectrum inhibitors.⁹

The main focus of current efforts toward the development of new NA inhibitors concentrate on structure-based ligand design.^{10,11} The three-dimensional structure of this protein was determined in the early 1980s.⁸ To date, around 100 neuraminidase X-ray complexes are documented in the Protein Data Bank. These structures played an important role in the discovery of several potent and specific inhibitors of the influenza virus NA.¹² Successful drug examples include zanamivir¹¹ and oseltamivir.¹³ However, the continuous threat of imminent pandemics^{14,15} and the emergence of resistant strains to oseltamivir (Tamiflu) is prompting continuous research toward developing new and more effective NA inhibitors.¹⁶

Furthermore, although considered the most reliable structural information that can be used for drug design, crystallographic structures are limited by inadequate resolution¹⁷ and crystallization-related artifacts of the ligand-protein complex.^{18–20} Moreover, crystallographic structures generally ignore structural heterogeneity related to protein anisotropic motion and discrete conformational substates.²¹

The continuous interest in designing new anti-influenza agents, in particular against strain A flu viruses, combined with the potential drawbacks of structure-based design prompted us to explore the possibility of developing anti-NA ligand-based 3D pharmacophore(s) integrated within self-consistent QSAR model(s). This approach avoids the pitfalls of structure-based techniques; furthermore, the pharmacophore model(s) can be used as efficient 3D search query(ies) to mine 3D libraries for new NA inhibitory leads.

Although QSAR analysis is normally conducted to identify appropriate combination(s) of physicochemical or structural descriptors capable of explaining bioactivity variation across a group of bioactive compounds,⁷¹ we implemented QSAR modeling herein as a competition platform to allow the selection of optimal pharmacophore(s) capable, upon combination with other 2D descriptors, of explaining bioactivity variation across a long list of NA inhibitors (224 compounds). We previously reported the use of this innovative approach toward the discovery of new inhibitory leads against glycogen synthase kinase 3β (GSK- 3β),⁶⁵ hormone sensitive lipase (HSL),⁷² bacterial MurF,⁷³ and protein tyrosine phosphatase 1B (PTP 1B).⁷⁴

The fact that most of the reported NA inhibitors are slow binding/transition state analogues (TSAs)^{10,13,22,23} should complicate pharmacophore modeling and the subsequent *in silico* search. TSAs resemble the substrate at its postulated

* Corresponding author e-mail: mutasem@ju.edu.jo.

transition to products implying that such inhibitors require stringent steric and 3D provisions in order to dock into the enzymatic binding site during its sterically demanding high-energy transition state. TSAs are known to be much more tightly bound to the targeted enzyme than their ground-state counterparts (i.e., substrate analogues), which further supports the notion about the pronounced sensitivity of TSA-enzyme complexes to slight misalignments among their complementary attractive groups.^{24–26} This conduct is expected to result in steep or rugged structure–activity surface, which limits the ability of the pharmacophore theory to explain activity/inactivity variations among training compounds. In fact, pharmacophore modeling requires continuous bioactivity variation attributable to the presence or absence of certain chemical features, i.e., it requires smooth SAR surface.

The pronounced sensitivity of TSA to slight structural modifications should also complicate the subsequent use of pharmacophore models as 3D search queries to mine for new hits. Pharmacophore models would be too lax and therefore promiscuous in capturing TSAs as *in silico* hits, i.e., they may identify many inactive hits (false positives).²⁷ In fact, all previous pharmacophore modeling efforts of NA inhibitors were reported without subsequent experimental validation, probably as a consequence to their high false positive rates.^{28,29} Accordingly, we were prompted to hybridize our QSAR-based pharmacophore models with tight ligand shapes and to use the combination as 3D search queries.

In the current study we employed CATALYST-HYPOGEN³⁰ to explore the pharmacophoric space of NA inhibitors utilizing 224 reported influenza A NA inhibitors. Three carefully selected training subsets (chosen based on our previously published detailed probing of ligand-NA interactions)³¹ were used to construct 178 possible binding hypotheses. Subsequently, genetic function algorithm (GFA) and multiple linear regression (MLR) analyses were employed to search for optimal QSAR that combine high-quality binding pharmacophores with other molecular descriptors and capable of explaining bioactivity variation across the whole collection NA inhibitors. The optimal pharmacophores were subsequently merged with tight ligand shapes (corresponding to known potent NA inhibitors) and used as 3D search queries to screen the NCI compounds database for new potential NA inhibitors.^{32,33} The captured hits were subsequently experimentally validated against influenza NA.

CATALYST models drug-receptor interaction using information derived only from the drug structure.^{30,34,35} HYPOGEN identifies a 3D array of a maximum of five chemical features common to active training molecules, which provides a relative alignment for each input molecule consistent with their binding to a proposed common receptor site. The chemical features considered can be hydrogen bond donors and acceptors (HBDs and HBAs), aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionizable groups, and aromatic planes. The conformational flexibility of training ligands is modeled by creating multiple conformers, judiciously prepared to emphasize representative coverage over a specified energy range.^{35,36}

The SHAPE module in CATALYST is a shape-based similarity searching method. The van der Waals surface of a molecule (in certain conformation) is calculated and represented as a set of points of uniform average density on

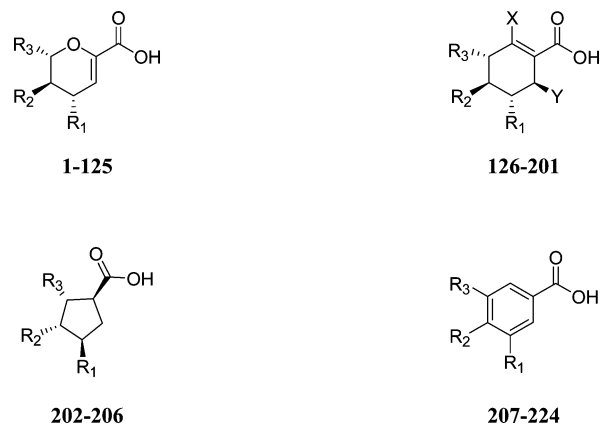


Figure 1. The chemical scaffolds of different NA inhibitors used in modeling. See Table A in the Supporting Information for the chemical structures.

a grid. The surface points enclose a volume on the grid. The geometric center of the set of points is computed along with the three principal component vectors passing through the center. The maximum extents along each principal axis and the total volume are calculated. These provide shape indices that can be compared with the query and used in an initial screening step to eliminate poor matches from further consideration. In CATALYST-SHAPE conformational flexibility is handled by precomputing an ensemble of conformers for each library compound and comparing each conformer with the query shape indices.³³ CATALYST pharmacophores, with or without shape constraints, have been used as 3D queries for database searching and in 3D-QSAR studies.^{33,35–37}

2. RESULTS AND DISCUSSION

2.1. Data Mining and Conformational Coverage. The literature was extensively surveyed to collect many structurally diverse inhibitors of influenza A neuraminidase. From about 400 known inhibitors available in the literature a group of 224 inhibitors from diverse scaffolds were collected: 125 dihydropyrans,^{38–47} 75 cyclohexenes,^{48–53} 4 cyclopentanes,^{54,55} and 17 benzene-based^{56–59} (Figure 1, Table A in the Supporting Information). Other reported scaffolds were excluded because they were either assayed by discrete assay procedures or lacked well-defined stereochemistry. The employed inhibitors were selected in such a way that they were assayed by the same procedure (see ref 70). Furthermore, in order to allow satisfactory pharmacophore and QSAR analysis, the bioactivities of the collected inhibitors were normalized appropriately to alleviate any discrepancies related to minor variations in assay conditions (see section 4.1.2 Data Set in the Experimental Section). The 2D structures of the training inhibitors were imported into CATALYST and converted automatically into plausible 3D single conformer representations *via* the rule-based methods implemented within the package. The resulting single conformer 3D structures were later used as starting point for conformational analysis and in the determination of various molecular descriptors for QSAR modeling.

The conformational space of each inhibitor was extensively sampled utilizing the poling algorithm employed within CATALYST. Poling promotes conformational variation via employing molecular mechanical force field algorithm that

Table 1. Training Subsets Used in Pharmacophore Exploration of NA Inhibitors and Their Corresponding Bioactivities Ranges^a

training subset	number of structures	most active ^b		intermediate ^b		inactive ^b	
		compounds	activity range (nM)	compounds	activity range (nM)	compounds	activity range (nM)
A	23	3, 6, 74, 99, 100, 102, 116, 117	0.5–4.0	7, 24, 87, 148, 159, 207, 209	38.5–1400	76, 78, 145, 146, 195, 203, 208, 210	1720–115000
B	18	14, 31, 32, 45, 132, 135, 136	0.3–1.5	41, 46, 66, 137, 155, 157, 178	3.5–600	64, 65, 92, 97	2100–6000
C	18	34, 37, 56, 127, 130, 141	0.2–1	60, 93, 126, 190, 207, 213	5–500	87, 145, 146, 147, 171, 211	1000–3700

^a Numbers correspond to compounds in Figure 1 and Table A in the Supporting Information. ^b Potency categories as defined by eqs 1 and 2 in section 4.1.4, **Exploration of the Pharmacophoric Space of NA**, under the Experimental Section.

penalizes similar conformers.^{34–36} Conformational coverage was performed employing the “Best” module to ensure extensive sampling of conformational space. Efficient conformational coverage guarantees minimum conformation-related noise during pharmacophore generation and validation stages. Pharmacophore generation and pharmacophore-based search procedures are known for their sensitivity to inadequate conformational sampling within the training compounds.⁶⁰

2.2. Pharmacophore Modeling. CATALYST-HYPOGEN enables automatic pharmacophore construction by using a collection of at least 16 molecules with bioactivities spanning over 4 orders of magnitude.^{34–36} CATALYST-HYPOGEN implements an optimization algorithm that evaluates a large number of potential models within the pharmacophoric “space” of a particular target through fine perturbations to pharmacophore hypotheses generated during the constructive phase and survived the subsequent subtractive phase of the CATALYST modeling algorithm (see section 4.1.4 **Generation of Pharmacophoric Hypotheses** in the Experimental Section).³⁶ The extent of the evaluated pharmacophoric space is reflected by the configuration (config.) cost calculated for each modeling run (see section 4.1.5 **Assessment of the Generated Hypotheses** in the Experimental Section). It is generally recommended that the config. cost of any HYPOGEN run not to exceed 17 corresponding to 2¹⁷ binding hypotheses to be assessed by CATALYST, otherwise thorough analysis of all models cannot be guaranteed.³⁶

The size of the investigated pharmacophoric space is a function of several factors, including the number of training compounds, selected input chemical features, and other CATALYST control parameters.³⁶ Restricting the extent of explored pharmacophoric space might improve the efficiency of optimization by evaluating a limited number of pharmacophoric models. On the other hand, tighter restrictions imposed on the pharmacophoric space might reduce the possibility of discovering optimal pharmacophoric hypotheses as they may occur outside the “boundaries” of the explored space.

Therefore, it was decided to employ a three-stage strategy to effectively explore the pharmacophoric space of neuraminidase inhibitors and to identify pharmacophoric hypotheses representing all possible binding modes assumed by NA inhibitors. Three training subsets were visually selected (i.e., without implementing mathematical or statistical similarity analytical methods, Table 1) for pharmacophore exploration in such a way as to guarantee maximal three-dimensional diversities and NA affinity profiles attributable solely to the

presence or absence of chemical features, i.e., not steric or auxiliary electronic effects (i.e., electron-withdrawing and/or electron-donating effects). Based on the fact that our collected list of NA inhibitors shared 4 chemical fragments, namely, carbocyclic core; carboxylic acid (Figure 1); basic moiety (amine or guanidine, R1 in Table A in the Supporting Information); and amidic side chain (R2, in Table A in the Supporting Information), while they differed significantly in a fifth group (corresponding to R3 in Table A in the Supporting Information), we were prompted to rely on R3 for selecting the training subsets. Our previous docking/CoMFA analyses³¹ suggested that NA inhibitors assume significantly different bound poses depending on the size and hydrophilic/lipophilic character of R3 prompting us to explore their pharmacophoric space employing training subsets based on the size and hydrophilic/lipophilic character of R3: group A characterized with hydrophilic R3 moieties (23 compounds of bioactivity range from 0.5 nM to 115 μ M, Table 1), group B characterized with bulky lipophilic R3 fragments (18 compounds of bioactivity range from 0.3 nM to 6 μ M, Table 1), and group C characterized with lean R3 moieties of mixed hydrophilic/lipophilic properties (18 compounds of bioactivity range from 0.2 nM to 3.7 μ M, Table 1).

In the first stage of our pharmacophore exploration strategy, the pharmacophoric spaces of the selected training subsets were explored under reasonably imposed “boundaries” through four HYPOGEN runs for each training subset (trials 1–4 for each subset, Table 2). The software was restricted to explore pharmacophoric models incorporating from zero to three features of any particular selected feature type instead of the default range of zero to five. The input features were selected in agreement with published SAR studies and crystallographic data which suggest the involvement of four primary regions in binding, namely, a positively charged triad (Arg118, Arg292, and Arg371); a hydrogen-bonding area (Arg152, Trp178, and Asp151); a negatively charged area (Glu119, Asp151, Glu227, Glu276, and Glu277); and a hydrophobic pocket (Ala246, Ile222, Trp178, and the propylene linker of Arg224).³¹ Accordingly, we were prompted to select HBA, HBD, Hydrophobic, Neglon, and Poslon functionalities as possible pharmacophoric features.

To further restrict the explored pharmacophoric space, only 4- and 5-featured pharmacophores (the maximal number of features in HYPOGEN-based models³⁶) were explored, i.e., models of lesser number of features were ignored. This restriction has the advantages of narrowing the extent of investigated pharmacophoric space and allowing a better representation of the feature-rich nature of NA binding

Table 2. Selected Input Features and CATALYST Run Parameters Employed in Exploring Neuraminidase Inhibitors Pharmacophoric Space

pharmacophore modeling runs		selected input features: types and ranges ^b						other run parameters ^b	
training subset ^a	trial	HBA	HBD	Hydrophobic	NegIon	PosIon	Max-Min ^c	spacing ^e	variable weights ^f
A	1	0-3	0-3	0-3	0-3	0-3	4-5	1	0
	2	0-3	0-3	0-3	0-3	0-3	4-5	3	0
	3 ^g	0-3	0-3	0-3	0-3	0-3	5-5	1	0
A	4	0-3	0-3	0-3	0-3	0-3	5-5	3	0
	5	0-3	0-3	---	0-1	0-1	4-4	1	0
	6	1-3	0-2	---	0-1	0-1	4-4	3	0
	7	3-3	---	---	---	1-1	4-4	1	1
	8	1-1	1-1	1-1	1-1	1-1	5-5	3	1
B	1	0-3	0-3	0-3	0-3	0-3	4-5	1	0
	2	0-3	0-3	0-3	0-3	0-3	4-5	3	0
	3	0-3	0-3	0-3	0-3	0-3	5-5	1	0
	4	0-3	0-3	0-3	0-3	0-3	5-5	3	0
	5	1-1	1-1	2-2	---	1-1	5-5	1	1
	6	1-1	1-1	2-2	---	---	4-4	3	1
C	1	0-3	0-3	0-3	0-3	0-3	4-5	1	0
	2	0-3	0-3	0-3	0-3	0-3	4-5	3	0
	3	0-3	0-3	0-3	0-3	0-3	5-5	1	0
	4	0-3	0-3	0-3	0-3	0-3	5-5	3	0
	5	0-2	1-2	2-2	0-1	0-1	4-5	1	0
	6	1-1	1-1	2-2	---	1-1	5-5	1	1
	7	2-2	---	2-2	---	1-1	5-5	3	1

^a The detailed success criteria for each pharmacophore model are shown in Table B in the Supporting Information. ^b The training subsets are as in Table 1. ^c Selected input features and run parameters are in Table 2. ^d See the Experimental Section (4.1.5 Assessment of the Generated Hypotheses) for a detailed description and a discussion regarding the pharmacophoric correlation coefficients and different cost criteria. ^e Calculated employing the Cat.Scramble methods. ^f This run failed to yield any pharmacophore model. ^g The pharmacophore correlation coefficients reported in the log book of each automatic HYPOGEN run.

Table 3. General Performance of Pharmacophoric Models Generated for Neuraminidase Inhibitors^a

training subset ^b	trial ^c	r ^{d,g} range	cost criteria ^a				Fischer confidence range (%) ^e
			configuration	total cost range	fixed	null	
A	1	0.86-0.94	19.0	120.0 - 129.9	97.5	206.1	95.0
	2	0.88-0.93	18.5	123.0 - 129.6	97.0	206.1	95.0
	3 ^f	---	---	---	---	---	---
	4	0.77-0.82	8.8	131.0 - 140.4	87.3	206.1	95.0
	5	0.89-0.89	19.0	127.1 - 130.3	97.5	206.1	95.0
	6	0.86-0.89	18.3	124.1 - 130.1	96.8	206.1	95.0
	7	0.89-0.94	28.3	127.8 - 142.1	106.8	206.1	95.0
	8	0.77-0.91	19.5	122.0 - 151.9	98.0	206.1	95.0
B	1	0.83-0.93	17.0	101.9 - 114.4	78.7	151.6	65.0-95.0
	2	0.83-0.94	16.3	93.8 - 114.0	77.9	151.6	85.0-95.0
	3	0.77-0.88	14.2	97.4 - 111.4	75.9	151.6	90.0-95.0
	4	0.75-0.88	13.5	100.5 - 114.7	75.2	151.6	65.0-90.0
	5	0.74-0.92	22.3	101.3 - 125.1	84.0	151.6	45.0-95.0
	6	0.91-0.94	24.7	102.3 - 114.0	86.4	151.6	95.0
C	1	0.87-0.96	18.0	87.4 - 102.9	79.7	153.5	85.0-95.0
	2	0.91-0.94	16.9	89.9 - 99.6	78.6	153.5	95.0
	3	0.89->0.96	14.6	84.8 - 95.5	76.3	153.5	95.0
	4	0.86-0.95	13.6	84.9 - 99.5	75.3	153.5	90.0-95.0
	5	0.88-0.95	14.8	86.5 - 97.9	76.4	153.5	95.0
	6	0.88-0.95	22.6	93.4 - 105.3	84.3	153.5	95.0
	7	0.88-0.97	23.1	90.3 - 106.5	84.8	153.5	95.0

^a The detailed success criteria for each pharmacophore model are shown in Table B in the Supporting Information. ^b The training subsets are as in Table 1. ^c Selected input features and run parameters are as in Table 2. ^d See the Experimental Section (4.1.5 Assessment of the Generated Hypotheses) for a detailed description and a discussion regarding the pharmacophoric correlation coefficients and different cost criteria. ^e Calculated employing the Cat. Scramble methods. ^f This run failed to yield any pharmacophore model. ^g The pharmacophoric correlation coefficients reported in the log book of each automatic HYPOGEN run.

pocket. Moreover, to further limit the investigated pharmacophoric space, we separately explored optimal interfeature spacing distances at 1 and 3 Å for each training subset.

Modeling trials 1-4 of each training subset in Table 2 show the detailed input pharmacophoric features and run parameters implemented in the first stage of pharmacophore exploration. Eventually, 10 optimal pharmacophoric hypo-

theses were generated for each run, which were automatically ranked according to their corresponding "total cost" values, defined as the sum of error cost, weight cost, and configuration cost (see section 4.1.5 Assessment of the Generated Hypotheses in the Experimental Section).^{30,34-36} Table 3 shows the corresponding cost and statistical criteria of the resulting models from each trial.

In an attempt to restrict the number of explored pharmacophoric models in trials of extended config. costs (>17), i.e., the first and second trials of subset A (A-1 and A-2, Table 3) and the first trial of subset C (C-1, Table 3), a second pharmacophoric exploration stage was implemented. In this phase, we restricted the numbers and types of input pharmacophoric features as well as the number of output features in the resulting models in agreement with optimal pharmacophoric models that emerged from the first stage, as in trials A-5, A-6, and C-5 (in Table 2). For example, in the first trial of training subset A (A-1, config. cost = 19.0, Table 3), the negative and positive ionizable features in the resulting hypotheses ranged from zero to 1 (see Table B in the Supporting Information) prompting us to use this range as input for negative and positive ionizable features in the second stage (trial A-5, Table 2). Similarly, the fact that all successful models in trial A-1 exhibited 4 pharmacophoric features prompted us to restrict our exploration in the second stage to 4-featured models only (trial A-5, Table 2).

Although, this restricted phase succeeded in improving the config. cost of trial C-1 from 18.0 to 14.8 in C-5 (Table 3), it fell short of achieving the same enhancements in trials A1 and A2 (i.e., as compared with trials A5 and A6, respectively, Table 3). Nevertheless, the restriction in this phase allowed sampling and exploration of different regions within the pharmacophoric space of training subset A as evident from the differences in feature types, numbers, and 3D coordinates among pharmacophore models generated in trials A-1 and A-2 compared to trials A-5 and A-6, respectively.

Although high-ranking models from the first two stages probably represent "actual" binding features, however, the fact that they were extracted from somehow wide pharmacophoric space (by exploring 0–1 or 0–3 features of any particular type) suggests they have gross 3D properties, i.e. inaccurate interfeature distances and angles. This assumption is supported by the observed high config. costs even after restrictions in the second stage (e.g., A-5 and A-6 trials >17 , Table 3). Therefore, it was reasoned that further exploration is required to search for models of superior 3D characters within the vicinity of optimal models generated from the two stages, i.e., to search for pharmacophores of identical feature types and numbers and enhanced interfeature distances and angles. Accordingly, the pharmacophoric features of the best models from the first and second stages were employed in a "focused" third stage modeling runs (i.e., runs A-7, A-8, B-5, B-6, C-6, and C-7, Table 3). Accordingly, the modeling trials in this stage were restricted to the pharmacophoric features of models **A-1-1**, **A-4-1**, **B-2-1**, **B-3-1**, **C-3-1**, and **C-4-1** (codes indicate subset-trial-hypothesis, e.g., **A-4-1** is the first pharmacophore of the fourth trial performed on subset A) as they represent the best performing hypotheses in terms of cost values among their peers from a particular training set. Moreover, in this stage, we configured CATALYST-HYPOGEN to explore models of variable feature weights (Table 2). Allowing variable feature weights permits the pharmacophoric features to have different relative impacts on ligand-target affinity.^{36,61} In contrast, the default constant features' weights settings assume that each pharmacophoric feature influences ligand-receptor affinity in a uniform behavior. Allowing feature weights to vary in CATALYST should create pharmacophores more suited to encode the

subtle nature of the receptor–ligand interactions.³⁶ However, variable weight analysis is usually accompanied with explosive expansion in the config. cost, as evident in Table 3; therefore, variable weight configuration was restricted to the best performing hypotheses only. This allowed CATALYST-HYPOGEN to explore variable feature weights without indulging in additional assessment of variable features.

Table 3 shows the success criteria ranges for pharmacophores from the three rounds of modeling trials. Clearly from the table, the trials illustrated variable success criteria, including their Fischer confidence levels (see section **4.1.5 Assessment of the Generated Hypotheses** in the Experimental Section).^{62,63} Moreover, despite our restrictions on the extent of evaluated pharmacophore space, the config. costs of some runs were generally high and exceeded the maximum limit of 17. This conduct is probably related to the flexibilities of some training compounds, and small size combined with extensive fictionalization of other training compounds, which seem to increase the entropy of the hypotheses space.^{30,34–36} Nevertheless, the reasonable cost and confidence criteria of most pharmacophore models should overshadow any drawbacks related to the less-than-optimal config. costs.

Emergence of numerous high-quality pharmacophore models suggest that NA ligands assume multiple pharmacophoric binding modes (poses, i.e., positions and orientations) within the binding pocket. Therefore, it will be rather erroneous to select any particular pharmacophore hypothesis as a sole representative of the ligand-protein binding process.

2.3. QSAR Modeling. Despite the undisputed significance of pharmacophoric hypotheses in understanding ligand-macromolecule affinity and as 3D search queries, their predictive value as 3D-QSAR models is generally limited by steric shielding and bioactivity-enhancing or -reducing auxiliary groups.^{61,64} This point combined with the fact that our NA training subsets furnished numerous reasonable binding hypotheses (as evident from their excellent correlation r coefficients and cost criteria, Table 3) prompted us to employ classical QSAR analysis to search for the best combination of orthogonal pharmacophores and other structural descriptors (connectivity, topological, 2D, etc.,...) capable of explaining bioactivity variation across the training compounds. This task was performed via genetic function approximation-multiple linear regression QSAR (GFA-MLR-QSAR) analysis (see section **4.1.7 QSAR Modeling** in the Experimental Section).

In order to fully represent the chemical diversity of reported influenza NA ligands, we conducted QSAR modeling using large set of inhibitors, i.e., 224 compounds collected from 21 published papers.^{38–59} However, since it is necessary to validate QSAR models against external compounds, we randomly removed 43 NA inhibitors from the collected list (*ca.* 20%) and reserved them to validate the generated QSAR models.^{61,65,72–74} The remaining compounds (181) were employed as training compounds for QSAR modeling. The training list included all three subsets implemented in pharmacophore exploration (subsets **A**, **B**, and **C**, Table 1) in addition to other extra compounds (123 compounds) added to assess the effects of electronic, steric, and topological factors on NA-ligand affinity. Addition of extra compounds should promote the selection of optimal combinations of pharmacophore models complemented with

Table 4. Statistical Performance of the Successful QSAR Models

QSAR model	no. of pharmacophore hypotheses used in QSAR ^a	terms ^b	$r_{(181)}^c$	F-value	$r^2_{(LOO)}^d$	$r^2_{(BS)}^e$	$r^2_{(PRESS)}^f$	PRESS ^g	pharmacophoric terms in QSAR model ^h
I	60	13	0.771	20.518	0.533	0.595	0.591	33.949	A-5-5, B-3-2
II	12	15	0.762	16.243	0.502	0.580	0.586	34.333	A-5-5, B-2-6, C-1-2
III	60	19	0.777	13.672	0.516	0.603	0.575	35.276	A-5-5, B-3-2, C-1-2
IV	12	17	0.78	15.881	0.507	0.609	0.509	40.7403	A-5-5, C-1-2
V	12	17	0.776	15.517	0.503	0.602	0.488	42.491	A-5-5, A-8-1, B-2-6, C-1-2

^a Number of pharmacophore-based best fit columns enrolled in genetic algorithm-multiple linear analysis. The different values are related to the number of clustered pharmacophores, i.e., in one instance they were clustered in 12 models, and in another instance they were clustered into 60. ^b Number of explanatory terms including the intercept. ^c Noncross-validated correlation coefficient for 181 training compounds. ^d Cross-validation correlation coefficients determined by the leave-one-out technique. ^e Bootstrapping correlation coefficient. ^f Predictive r^2 determined for the 43 test compounds. ^g The sum of squared deviations between predicted and actual activity values for every molecule in the test set of 43 compounds. ^h The pharmacophore hypotheses are coded as follows: training subset-trial number-hypothesis serial rank, for example, A-5-5 encodes for the 5th pharmacophore hypothesis from trial 5 performed on subset A (as in Tables 1 and 2), while B-3-2 encodes for the 2nd hypothesis from the 3rd trial performed on subset B, and so on.

electronic, steric, and topological descriptors in QSAR contexts capable of explaining bioactivity variation across diverse NA inhibitors.

Our approach of pharmacophoric exploration via small subsets of diverse 3D structures complemented with subsequent QSAR modeling employing ample training lists should achieve two important goals: (i) provide competition platform for the selection of optimal pharmacophores from a large pool of models and (ii) produce QSAR models capable of explaining bioactivity variations resulting from 3D ligand-receptor recognition aspects (pharmacophore recognition) as well as electronic, steric, and topological factors.^{61,65,72-74}

Table 4 shows the statistical criteria of the resulting top-ranking QSAR models and their associated pharmacophore models. Figure 2 shows the corresponding scatter plots of experimental versus estimated bioactivities for the training and testing inhibitors. The QSAR equations together with definitions of their respective terms are shown in Tables C and D in the Supporting Information. It is clearly evident from Figure 2 that all successful QSAR models overestimated the bioactivities of poorly active NA inhibitors, which probably reflect the fact that most of our collected NA inhibitors are transition state analogues that cannot be adequately described by 2D descriptors.

Interestingly, successful QSARs exhibited several pharmacophore combinations represented by five binding hypotheses, which suggest that NA ligands assume different binding poses (i.e., positions and orientations) within the same binding pocket.^{61,65} In fact, each QSAR-selected pharmacophore model encodes for a different set of ligand-binding interactions pointing to distinct sets of binding amino acid residues within the binding site. Apparently, subtle differences in the 3D distribution of pharmacophoric features (hydrophobic, hydrogen bonding or electrostatic groups) bring about significant differences in the involved binding amino acid residues within the binding pocket, which reflect in a collection of distinct binding pharmacophores and bound ligand poses. This point can be further established by comparing our QSAR-selected pharmacophore models with different docked poses and crystallographically resolved bound structures of some selected NA ligands in section 2.5 (Comparison of QSAR-Selected Pharmacophores with NA Binding Site). Figures 3-7 show the successful pharmacophore models and how they fit some potent inhibitors,

while Tables 5 and 6 show their binding features, success criteria, and X, Y, Z coordinates of their pharmacophoric features.

Emergence of topological, electrotopological, shape, and shadow descriptors in the successful QSARs (see Tables C and D in the Supporting Information) illustrate a certain role played by ligands' geometry and topology in the binding process. This is not unexpected, as most of the collected training compounds, particularly the highest-ranking, are transition state analogues that have their binding behavior greatly influenced by molecular shape (see the Introduction). However, despite the predictive significance of these descriptors, their information content is quite obscure. On the other hand, emergence of highest and lowest frontier orbital energies (HOMO and LUMO) indicates a certain role played by ligands' electronic characteristics in NA binding. Similarly, emergence of several JURS surface charge-related descriptors in association with negative regression coefficients suggests significant hydration-related effects that undermine ligand-NA binding, i.e., pronounced surface charges seem to prefer interaction with water molecules instead of binding within NA binding pocket.

2.4. In Silico Screening of NCI Database and Subsequent Experimental Evaluation. To increase the probability of active captured hits, it was necessary to guarantee the tightest 3D analogy between hit compounds and potent transition state training inhibitors. Therefore, it was decided to decorate our QSAR-selected pharmacophore models with shape-constraints derived from potent training inhibitors. Shape constraints encode for the extent of 3D spatial similarity between screened compounds and template ligands used to build the shape limitations.³²

To generate merged shape-pharmacophore queries, a selected potent training compound was first fitted against the corresponding pharmacophore model, and thereafter, the best-fitted conformer of the inhibitor was used to generate shape constraints that were subsequently merged with the pharmacophore (see section 4.1.8 In Silico Screening of the NCI Database for New Neuraminidase Inhibitors). However, to create the tightest possible shape queries, several shape tolerance values were evaluated.³⁰ Tolerance values that allowed the capture of a reasonable number of hits were used for virtual screening. A range of 20 to 100 hits was considered reasonable as it is manageable for *in vitro*

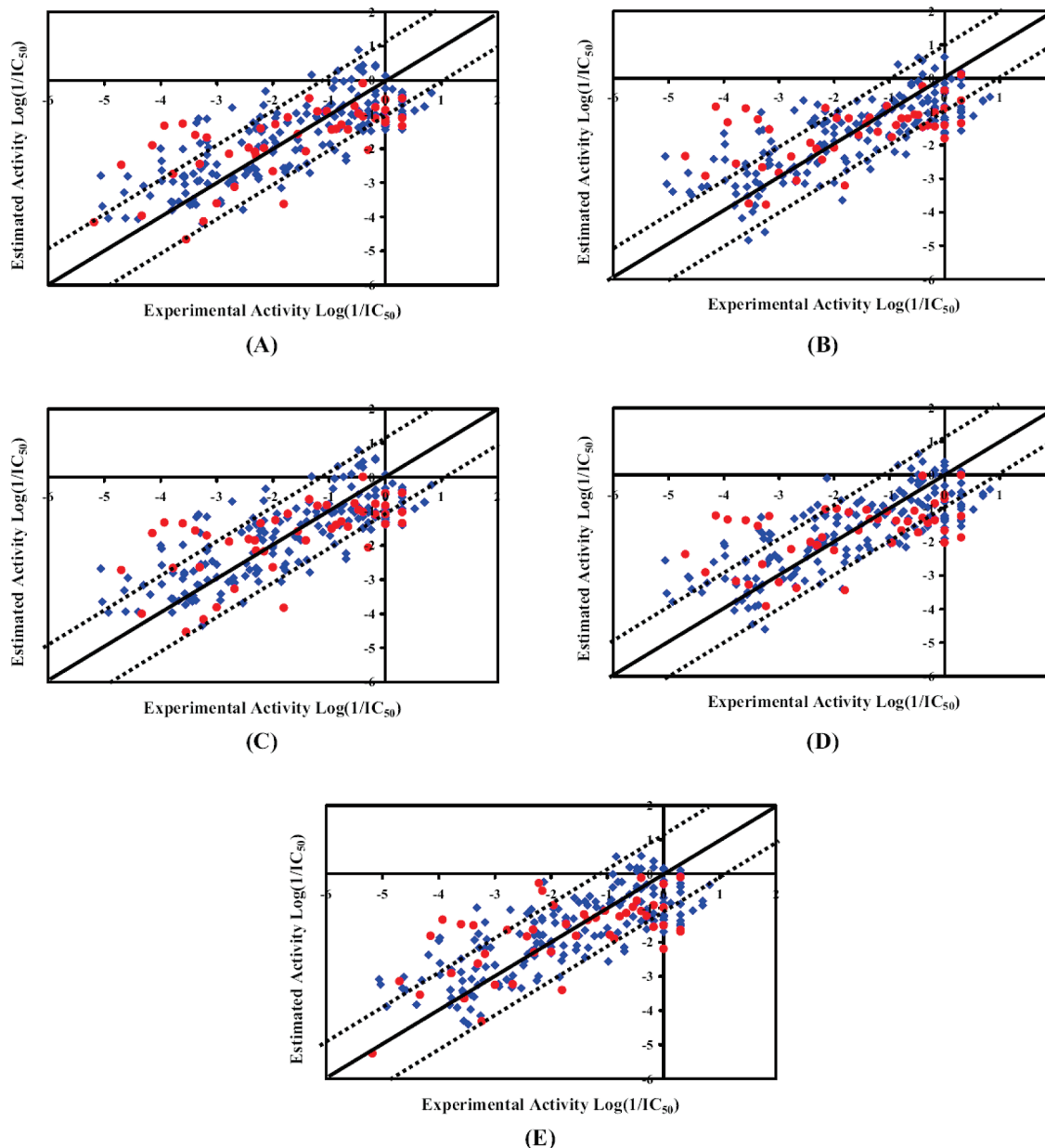


Figure 2. Experimental versus fitted (◆ (blue), 181 compounds) and predicted (● (red), 43 compounds) bioactivities calculated by QSAR models: (A) model I, $r^2_{\text{LOO}} = 0.533$, $r^2_{\text{PRESS}} = 0.591$, (B) model II, $r^2_{\text{LOO}} = 0.502$, $r^2_{\text{PRESS}} = 0.586$, (C) model III, $r^2_{\text{LOO}} = 0.516$, $r^2_{\text{PRESS}} = 0.575$, (D) model IV, $r^2_{\text{LOO}} = 0.507$, $r^2_{\text{PRESS}} = 0.509$, and (E) model V, $r^2_{\text{LOO}} = 0.503$, $r^2_{\text{PRESS}} = 0.488$ (see Table 4). The solid lines are the regression line for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1.0 log point errors margins.

screening and large enough to allow satisfactory diversity within the hit lists. A total of 6 merged shape-pharmacophore queries were generated, as shown in Figures 3–7 and Tables 5 and 6. The generated merged pharmacophore-shape queries were employed as 3D search queries against the NCI database.

Table 7 shows the performances of our shape-complemented pharmacophore queries. Clearly from the table, a total of 290 hit compounds were captured (Table 7). Hits exhibiting reactive covalent bond-forming moieties (alkylating moieties or Michael acceptors) or toxic elements (e.g., arsenic) were excluded. Unfortunately, only 105 molecules were available from the NCI for *in vitro* evaluation against NA. Figure A in the Supporting Information shows the chemical structures of the tested hits.

The fact that our best performing QSAR models were of mediocre predictive qualities (r^2_{press} ranges from 0.591 to 0.488) prompted us to avoid using their predictions to prioritize hits for *in vitro* evaluation. Therefore, we evaluated all available 105 NCI hits against NA (from N9 Noddy Tern virus) with oseltamivir as standard inhibitor. At 10 μM levels, 83 hit molecules illustrated measurable anti-NA inhibition ranging from 0.3%–86.4% (equivalent to 80% enrichment, see Table E in the Supporting Information). Figure 8 shows the chemical structures of hit compounds of inhibitory percentages $\geq 30\%$ (25 compounds), while Table 8 shows their corresponding pharmacophoric fit values, QSAR predictions, and experimental inhibitions.

Hits with promising inhibitory potential at 10 μM , namely, **289**, **296**, and **300**, were further studied to determine their

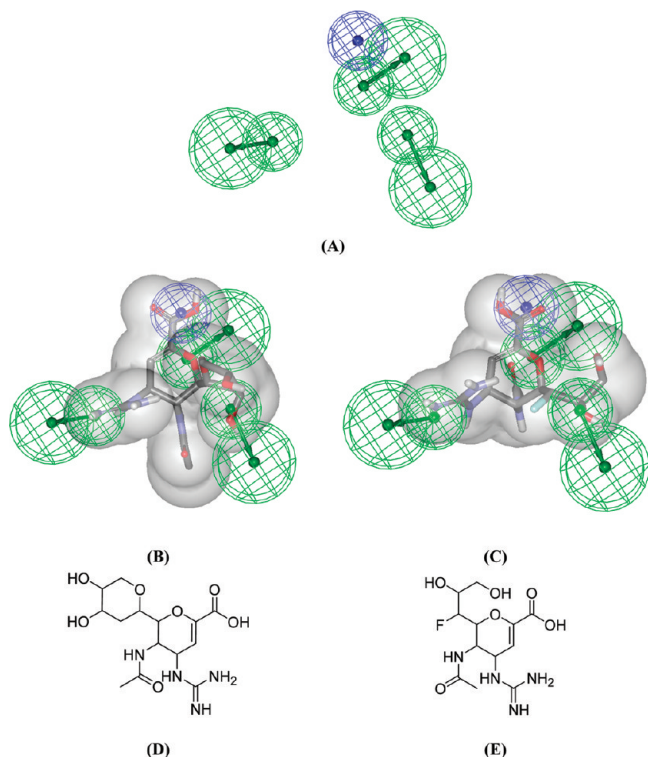


Figure 3. (A) The pharmacophoric features of A-5-5. Blue spheres represent negative ionizable features, while green-vec-tored spheres represent hydrogen-bond acceptors. (B) and (C) potent inhibitors 6 and 99 (structures D and E, respectively, $IC_{50} = 0.5$ nM) fitted against A-5-5 and combined with their corresponding molecular shapes.

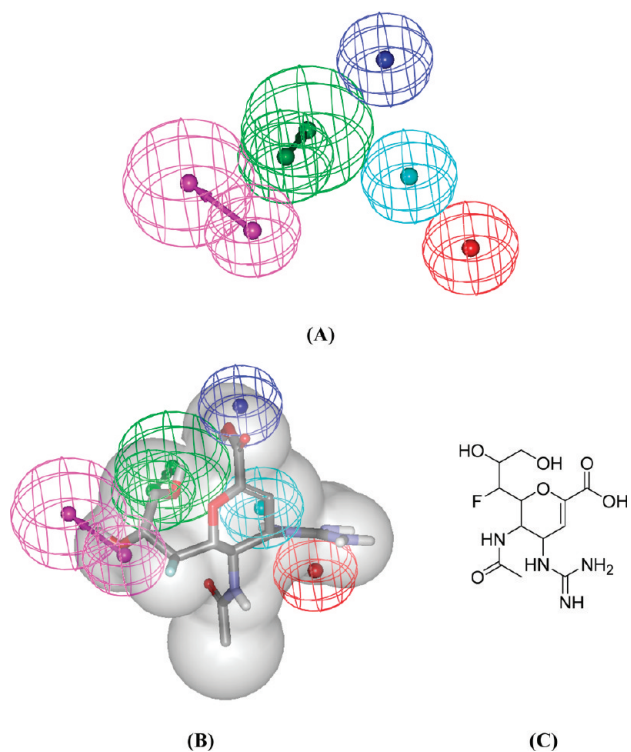


Figure 4. (A) The pharmacophoric features of A-8-1. Light-blue spheres represent hydrophobic features, red spheres encode for positive ionizable features, blue spheres encode for negative ionizable features, while green- and magenta-vec-tored spheres encode for hydrogen-bond acceptor and donor features, respectively. (B) Inhibitor 99 (structure C, $IC_{50} = 0.5$ nM) fitted against A-8-1 and combined with the corresponding molecular shape.

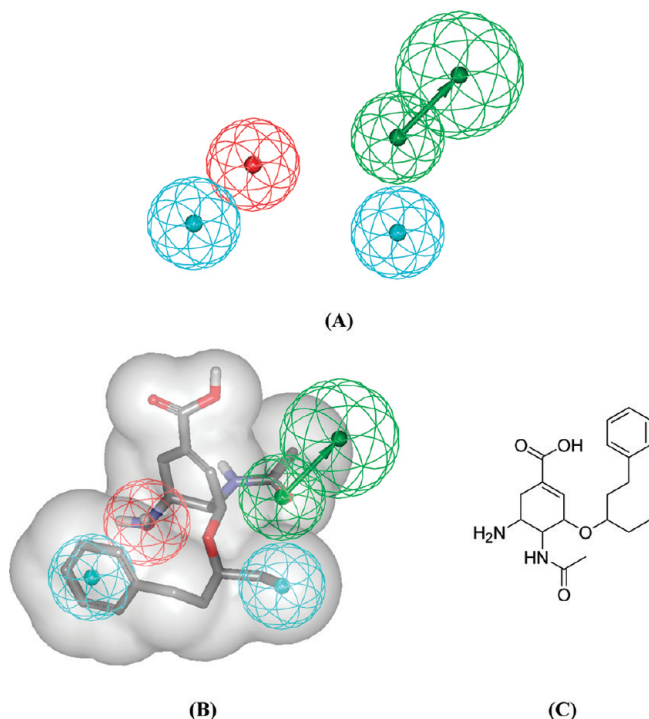


Figure 5. (A) The pharmacophoric features of B-2-6. Light blue spheres represent hydrophobic features, red spheres represent positive ionizable features, while green-vec-tored spheres encode for hydrogen-bond acceptor. (B) Inhibitor 136 (structure C, $IC_{50} = 0.3$ nM) fitted against B-2-6 and combined with the corresponding molecular shape.

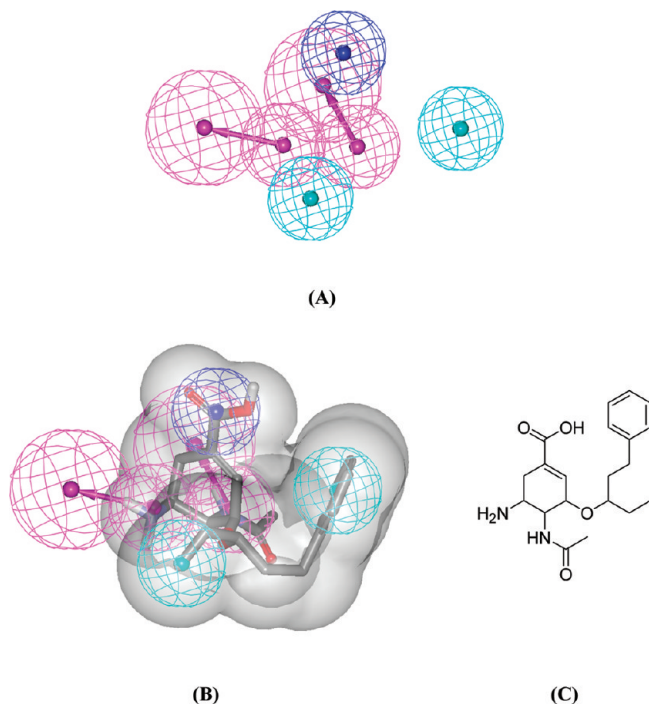


Figure 6. (A) The pharmacophoric features of B-3-2. Light blue spheres represent hydrophobic features; blue spheres encode for negative ionizable features, and magenta-vec-tored spheres encode for hydrogen-bond donor features. (B) Inhibitor 136 (structure C, $IC_{50} = 0.3$ nM) fitted against B-2-6 and combined with the corresponding molecular shape.

corresponding IC_{50} values (Table 8). Clearly, from Table 8 and Figure 8, our active hits belonged to diverse chemical scaffolds. In particular, glycosidic aromatic and heteroaromatic scaffolds are heavily represented in the active list, e.g.,

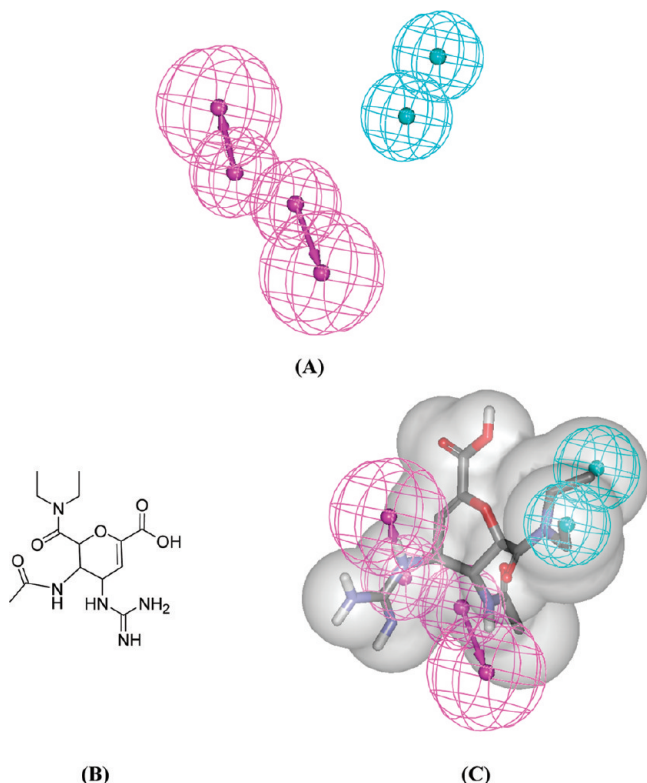


Figure 7. (A) The pharmacophoric features of **C-1-2**. Light blue spheres represent hydrophobic features, and magenta vectored spheres encode for hydrogen-bond donor. (B) Inhibitor **37** (structure **C**, $IC_{50} = 0.2$ nM) fitted against **C-1-2** and combined with the corresponding molecular shape.

300, 308, 316, 318, 319, 320, 322, 323, 324, 326, and **328**. This is not unexpected since a natural NA substrate, i.e., sialic acid, as well as some recently disclosed NA inhibitors exhibit noticeable glycosidic properties.⁶⁶ Similarly, the related thiolated mannoses **307** and **315** showed appreciable inhibitory potentials.

In our view, the most interesting active hits are **289** and **296** (IC_{50} values = 7.34 μ M and 12.52 μ M, respectively, Table 8 and Figure 8) as they represent completely novel anti-NA scaffolds amenable for optimization.

2.5. Comparison of QSAR-Selected Pharmacophores with NA Binding Site. Despite the problems of crystallographic structures (see the Introduction),¹⁷⁻²¹ pharmacophore features obtained by pharmacophore/QSAR modeling can be compared with the structure of the NA binding site to identify probable residues important for ligand binding and inhibition. Therefore, several training and hit compounds were fitted against our QSAR-selected pharmacophores, and the resulting mapped conformers were compared with docked poses of these compounds into the NA binding site (PDB code: 1NNC, resolution 1.8 Å). Moreover, we compared the crystallographically resolved bound poses of zanamivir and oseltamivir (i.e., to NA) with the way these established NA inhibitors map our QSAR-selected pharmacophores.

The docking experiments were performed employing LIGANDFIT docking engine and through default docking parameters and consensus scoring function.⁶⁷ However, we confined the docking simulation to rigid docking of pharmacophore-fitted conformers to avoid unnecessary exploration of irrelevant docked conformers, which should focus

the attention on probable binding residues corresponding to binding features within our QSAR-selected pharmacophores.

The features in **A-5-5** as well as the alignment of training compound **6** ($IC_{50} = 0.5$ nM) as proposed by **A-5-5** were compared with the way this compound docks into the binding pocket of NA, as in Figure 9. A marked similarity was observed between the features proposed by the pharmacophore model and the ligand binding features in the docked structure.

In one of the high-ranking docked poses of **6** (Figure 9a) the carboxylic acid group was oriented toward the guanidino triad ARG292, ARG371, and ARG118 of the binding pocket. This interaction corresponds to a NegIon feature in **A-5-5** mapping the carboxylate group of **6**. Similarly, the docking experiment suggests that three hydrogen bond-forming groups in **6**, namely, the hydroxyl of the tetrahydropyran, the oxygen of central dihydropyran ring, and the guanidino side chain, interact with ARG152, H₂O126 (fixed by hydrogen-bonding to ASP151 and ARG118), and ARG292, respectively, which seem to agree with mapping these groups with three HBA features in **A-5-5** (Figure 9a). Comparably, the docked pose of **99** ($IC_{50} = 0.5$ nM) corresponds to fitting **99** against binding model **A-8-1** (Figure 9b). The carboxylate moiety of **99** seems to electrostatically interact with the guanidine of ARG371, which corresponds to mapping this moiety with the NegIon feature in **A-8-1**. Similarly, the guanidino side chain of **99** is directed by the docking engine toward ASP151 suggesting significant mutual electrostatic attraction, as in Figure 9b. This agrees with mapping the guanidino of **99** onto the PosIon feature in **A-8-1**. The two hydroxyls of **99** seem to form hydrogen bonds with H₂O126 and H₂O285 (fixed by hydrogen-bonding to ARG224) in the docked pose, which agrees nicely with mapping the same hydroxyls by HBD and HBA features in model **A-8-1**. Finally, the dihydropyran ring of **99** was placed by the docking engine at close proximity to the aromatic ring of TYR406 suggesting the existence of mutual hydrophobic interactions corresponding to mapping the dihydropyran ring with hydrophobic feature in **A-8-1**.

Interestingly, docking **136** ($IC_{50} = 0.30$ nM) into the binding pocket of NA accessed two high-ranking poses corresponding to mapping **136** against **B-2-6** and **B-3-2** (as in Figure 9c and d, respectively). Apparently, in one of the poses (Figure 9c), LigandFit directed the amino group of **136** toward the acidic residue ASP151 suggesting possible mutual electrostatic attraction, which agrees with fitting this group against a NegIon feature in **B-2-6**. Similarly, mapping the acetamido carbonyl of **136** against a HBA feature in **B-2-6** corresponds to a hydrogen-bonding interaction tying this group with the phenolic hydroxyl of TYR406 (Figure 9c). On the other hand, the bulky hydrophobic arylalkyl moiety of **136** was directed, in this docked pose, toward a hydrophobic pocket comprised of the side chains of ALA246, ILE222, and the propylene linker of ARG224, which correlates nicely with mapping the ethyl and phenylethyl groups of **136** against two hydrophobic features in **B-2-6**, as in Figure 9c. However, the docking engine suggested an alternative flipped pose for **136** corresponding to **B-3-2**: The carboxylic acid and amino groups of **136** were directed toward the side chains of ARG152 and GLU277 to form ionic and hydrogen-bonding interactions, respectively. Furthermore, the amido NH of **136** seems to get involved in a

Table 5. Performance of Different Pharmacophoric Hypotheses Generated for Neuraminidase Inhibitors Employing Different Parameters (As in Table 2)

hypothesis ^a	pharmacophoric features ^b						hypothesis scores				
	HBA	HBD	Hydrophobic	NegIon	PosIon	total	total cost	correl. ^c	error cost	weight cost	confidence (%)
A-5-5	3	---	---	1	---	4	129.7	0.90	107.7	2.9	95
A-8-1	1	1	1	1	1	5	122.0	0.90	99.9	2.6	95
B-2-6	1	2	---	---	1	4	111.5	0.80	89.3	5.9	90
B-3-2	---	2	2	1	---	5	104.8	0.80	86.5	4.1	90
C-1-2	---	2	2	---	---	4	98.6	0.90	78.8	1.7	90

^a The pharmacophore hypotheses are coded as follows: training subset-trial number-hypothesis serial rank, for example, **A-5-5** encodes for the 5th pharmacophore hypothesis from trial 5 performed on subset A (as in Tables 1 and 2), while **B-3-2** encodes for the 2nd hypothesis from the 3rd trial performed on subset B, and so on. ^b Feature coordinates, weight and tolerances are shown in Table 6. ^c The correlation coefficient of pharmacophore fit values calculated for each corresponding training list of compounds against their respective bioactivities.

Table 6. Pharmacophoric Features and Corresponding Weights, Tolerances, and 3D Coordinates of Successful Models

model	definitions	chemical features									
		HBA			HBA			HBA		NegIon	
A-5-5	weights		2.57			2.57			2.57		2.57
	tolerances		1.60	2.20	1.60	2.20	1.60	2.20	1.60	2.20	1.60
	coordinates	X	1.60	4.57	-1.36	-1.19	-1.94	-2.66	-3.70		
		Y	-2.56	-2.02	-1.80	-4.35	3.67	5.88	-3.00		
	Z	-2.74	-3.10	1.70	3.26	-1.25	0.66	0.51			

model	definitions	chemical features								
		HBA		HBD		Hydrophobic		NegIon	PosIon	
A-8-1	weights		2.52		2.99		1.10		2.99	2.99
	tolerances		1.60	2.20	1.60	2.20	1.60	2.20	1.60	2.20
	coordinates	X	1.87	4.47	-1.59	0.55	1.52	3.57	1.19	
		Y	-3.30	-2.77	-4.44	-6.48	0.64	0.42	2.17	
	Z	0.63	2.02	0.52	0.04	1.90	-1.62	5.21		

model	definitions	chemical features								
		HBA		Hydrophobic		Hydrophobic		PosIon		
B-2-6	weights		2.93		2.93		2.93		2.93	
	tolerances		1.60	2.20	1.60	2.20	1.60	2.20	1.60	
	coordinates	X	3.96	6.26	-3.15	0.54	0.45			
		Y	-1.68	-3.23	-1.35	-0.48	-1.60			
	Z	-0.53	-1.69	2.96	-2.92	3.09				

model	definitions	chemical features								
		HBD		HBD		Hydrophobic		Hydrophobic		NegIon
B-3-2	weights		2.74		2.74		2.74		2.74	
	tolerances		1.60	2.20	1.60	2.20	1.60	2.20	1.60	
	coordinates	X	1.36	2.79	0.90	3.67	-1.34	-0.24	3.76	
		Y	1.12	1.58	2.65	3.84	-1.74	-2.34	-1.31	
	Z	2.91	5.51	-0.53	-0.47	3.72	-2.59	0.50		

model	definitions	chemical features								
		HBD		HBD		Hydrophobic		Hydrophobic		
C-1-2	weights		2.32		2.32		2.32		2.32	
	tolerances		1.60	2.20	1.60	2.20	1.60	2.20	1.60	
	coordinates	X	4.34	6.63	3.79	1.87	1.13	-1.39		
		Y	1.34	-0.52	2.87	2.85	-4.58	-3.82		
	Z	4.30	3.66	2.23	-0.07	-0.45	2.51			

hydrogen-bonding interaction with H₂O126 in this pose, which agrees with mapping amido NH by a HBD feature in **B-3-2**, as in Figure 9d. Similarly, the arylalkyl side chain of **136** was oriented in this pose toward a hydrophobic pocket comprised of ALA246, ILE222, and the aliphatic linkers of ARG224 and GLU276, which correlates with mapping the arylalkyl side chain against two hydrophobic features in **B-3-2** (Figure 9d).

Finally, mapping **37** (IC₅₀ = 0.20 nM) against **C-1-2** correlates nicely with a docked pose that orients the

compound's guanidino and acetamido groups in such a way to form hydrogen bonds with GLU276 and H₂O285, respectively, corresponding to two HBD features mapping the same groups, as in Figure 9e. The hydrophobic diethyl substituents of **37** were oriented by the docking engine toward the hydrophobic side chain of TRP178 and aliphatic linker of ASP151 explaining mapping the diethyl groups by two hydrophobic features in **C-1-2**.

To further emphasize the validity of our pharmacophore/QSAR modeling approach, we compared the crystallographic

Table 7. Performance of Shape-Complemented Pharmacophore Hypotheses as 3D Search Queries against the National Cancer Institute (NCI) List of Compounds

pharmacophore	fitted compound for ligand shape development ^a	shape tolerance ^b	number of <i>in silico</i> hits	number of tested hits	number of active hits
A-5-5	6	94-106	34	16 (225-240 ^c)	7
	99	90-110	48	5 (241-245 ^c)	2
A-8-1	6	---	---	---	---
	99	82-118	28	5 (246-250 ^c)	5
B-2-6	136	96-104	71	38 (251-288 ^c)	36
B-3-2	136	84-116	28	13 (289-301 ^c)	8
C-1-2	37	90-110	81	29 (302-330 ^c)	26

^a The most potent compound in the corresponding subset was fitted against the respective pharmacophore and used to complement the respective pharmacophore hypotheses with ligand shapes. ^b Shape tolerance extent percentage ranges. ^c Compounds' codes as in Figure 1 and Table A in the Supporting Information.

structures of two NA/ligand complexes, namely, zanamivir and oseltamivir (PDB codes: 1NNC and 2QWK, respectively, resolution 1.8 Å), with our QSAR-selected pharmacophore models. Interestingly, the crystallographic structures correlated tightly with A-8-1 and B-3-2, respectively, as shown in Figure 10.

Furthermore, the docked poses of the most potent hits were compared with the way they map our QSAR-selected pharmacophores. Intriguingly, the three most potent hits, i.e., 289, 296, and 300, dock into the binding pocket in comparable poses/conformations to the way they fit B-3-2, as shown in Figure 11.

3. CONCLUSION

This work includes elaborate pharmacophore exploration of NA inhibitors utilizing CATALYST-HYPOGEN. QSAR analysis was employed to select the best combination of molecular descriptors and pharmacophore models capable of explaining bioactivity variation across an informative list of training compounds. The QSAR-selected binding pharmacophores were complemented with shape constraints corresponding to potent template inhibitors and used as 3D search queries to screen the NCI database for new NA inhibitors. One hundred six high-ranking hits were acquired, out of which 83 compounds exhibited mediocre to potent inhibitory percentages against NA at 10 μM levels. Three hits were further evaluated and were found to possess micromolar IC₅₀ values. Interestingly, the pharmacophoric features of our QSAR-selected binding models correlated well with binding features apparent in crystallographic ligand-NA complexes and binding interactions proposed by docking studies. The fact that our pharmacophore models coincided concurrently with docked poses and experimentally resolved crystallographic poses supports both: our pharmacophore-QSAR exploration as well as our docking experiments.

The combination of pharmacophore modeling, QSAR analysis, and shape constraints proved very successful in enriching the hit-list despite the fact that our pharmacophore modeling relied on TSA inhibitors.

4. EXPERIMENTAL SECTION

4.1. Molecular Modeling. 4.1.1. Software and Hardware.

The following software packages were utilized in the present research.

- CATALYST (Version 4.11), Accelrys Inc. (www.accelrys.com), USA.

- CERIOUS2 (Version 4.10), Accelrys Inc. (www.accelrys.com), USA.

- CS ChemDraw Ultra 7.01, Cambridge Soft Corp. (http://www.cambridgesoft.com), USA.

Pharmacophore modeling, QSAR analysis, and docking studies were performed using CATALYST (HYPOGEN module) and CERIOUS2 software suites from Accelrys Inc. (San Diego, CA, www.accelrys.com) installed on a Silicon Graphics Octane2 desktop workstation equipped with a 600 MHz MIPS R14000 processor (0.5 GB RAM) running the Irix 6.5 operating system.

4.1.2. Data Set. The structures of 224 influenza A neuraminidase enzyme inhibitors belonging to the following chemical classes were used for modeling: dihydropyrans,³⁸⁻⁴⁷ cyclohexenes,⁴⁸⁻⁵³ cyclopentanes,^{54,55} and benzene-based⁵⁶⁻⁵⁹ (Figure 1, Table A in the Supporting Information). Only achiral inhibitors or those of known absolute chiral configurations were employed. The *in vitro* bioactivities of all collected inhibitors are expressed as the concentration (in nanomolar) of the test compound that inhibited NA activity by 50% (IC₅₀). The bioactivities of the collected inhibitors were normalized to allow appropriate statistical QSAR correlation: Since oseltamivir and zanamivir frequented in all our selected literature references,³⁸⁻⁵⁹ we employed their reported bioactivities according to Kim et al.¹³ (IC₅₀ = 1.0 nM) as reference points to normalize the bioactivities of other collected inhibitors. For example, since the IC₅₀ value of zanamivir in ref 40 was reported to be 5 nM, the bioactivities of all compounds collected from this reference (compounds 18-30, Table A in the Supporting Information) were divided by 5.

The logarithm of measured IC₅₀ (nM) values were used in pharmacophore modeling and QSAR analysis, thus correlating the data linear to the free energy change.

The two-dimensional (2D) chemical structures of the inhibitors were sketched using ChemDraw Ultra and saved in MDL-mol file format. Subsequently, they were imported into CATALYST, converted into corresponding standard 3D structures, and energy minimized to the closest local minimum using the molecular mechanics CHARMM force field implemented in CATALYST. The resulting 3D structures were utilized as starting conformers for conformational analysis.

4.1.3. Conformational Analysis. Molecular flexibility was taken into account by considering each compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a

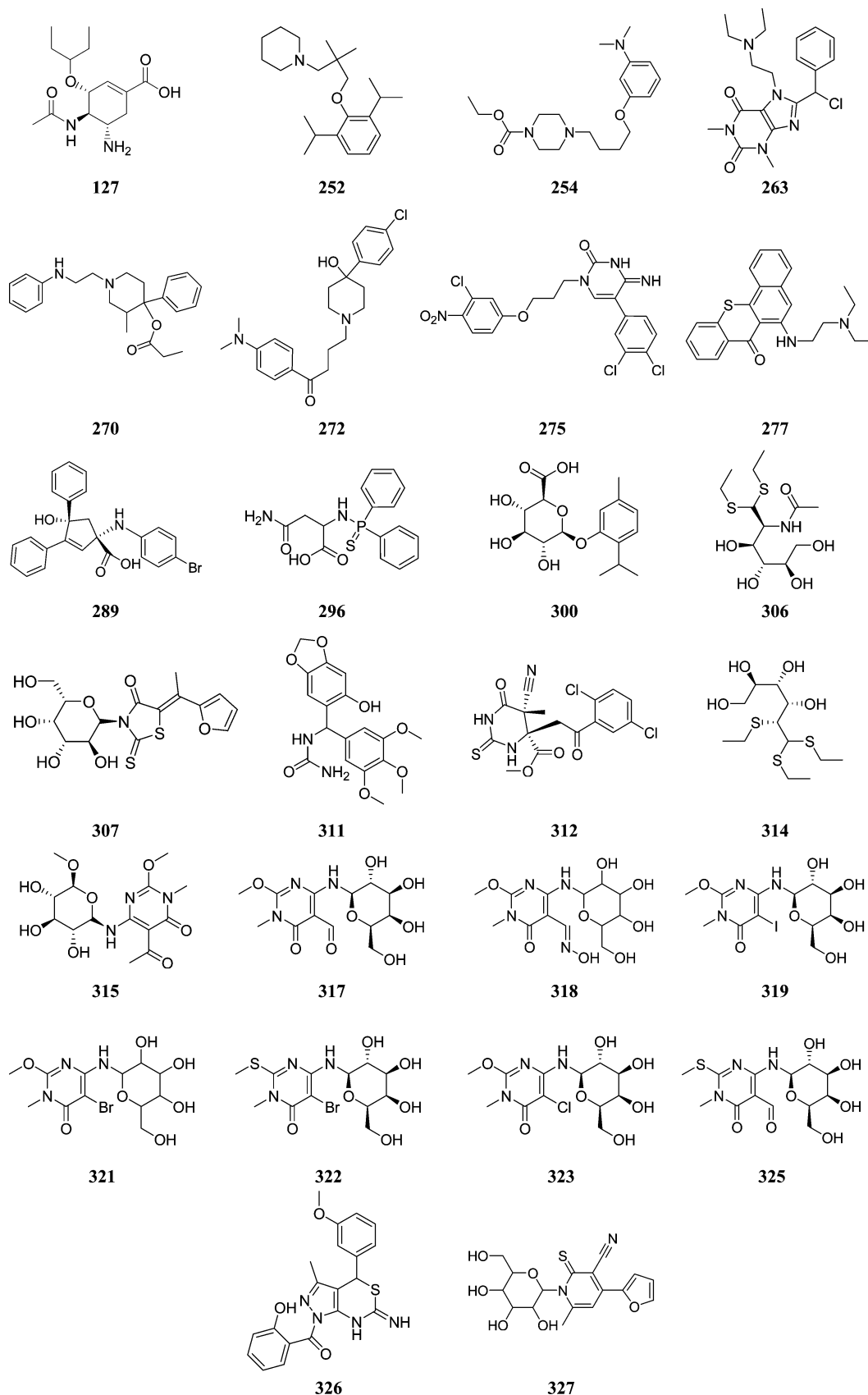


Figure 8. The active hits captured by our QSAR-selected pharmacophore models. The corresponding inhibitory activities are shown in Table 8.

given energy range. Accordingly, the conformational space of each inhibitor (1–224, Figure 1 and Table A in the Supporting Information) was explored adopting the “best conformer generation” option within CATALYST.^{30,34–36}

Conformational ensembles were generated with an energy threshold of 20 kcal/mol from the local minimized structures and maximum limit of 250 conformers per molecule. This search procedure will probably identify the best three-

dimensional arrangement of chemical functionalities explaining the activity variations among the training set.^{30,34–36}

4.1.4. Exploration of the Pharmacophoric Space of NA. All collected molecules with their associated conformational models were regrouped into a spreadsheet. The biological data of the inhibitors were reported with an “Uncertainty” value of 3, which means that the actual bioactivity of a particular inhibitor is assumed to be situated somewhere in an interval ranging from one-third to three times the reported bioactivity value of that inhibitor.^{30,34–36} Subsequently, three carefully selected diverse subsets (visually assessed and selected) from the collected inhibitors were employed as training sets for pharmacophore modeling (Table 1). Typically, CATALYST requires informative training sets that include at least 16 compounds of evenly spread bioactivities over at least 4 orders of magnitude.^{30,34–36}

Each training subset was utilized to conduct several modeling runs to explore the pharmacophoric space of NA inhibitors, as shown in Table 2. Different hypotheses were generated by altering interfeature spacing parameter, number and type of input features, total number of allowed features in the resulting pharmacophores, and features’ weights (Table 2).

Pharmacophore modeling employing CATALYST proceeds through three successive phases: the constructive phase, the subtractive phase, and the optimization phase.^{30,34–36} During the constructive phase, CATALYST generates common conformational alignments among the most-active training compounds. Only molecular alignments based on a maximum of five chemical features are considered. The program identifies a particular compound as being within the most active category if it satisfies eq 1^{30,34–36}

$$(\text{MAct} \times \text{UncMAct}) - (\text{Act}/\text{UncAct}) > 0.0 \quad (1)$$

where “MAct” is the activity of the most active compound in the training set, “Unc” is the uncertainty of the compounds, and “Act” is the activity of the training compounds in question. Table 1 shows the most active categories within the selected training subsets. In the subsequent subtractive phase, CATALYST eliminates some hypotheses that fit inactive training compounds. A particular training compound is defined as being inactive if it satisfies eq 2^{30,34–36}

$$\log(\text{Act}) - \log(\text{MAct}) > \text{BS} \quad (2)$$

where “BS” is the bioactivity spread (equals 3.5 by default). Table 1 shows the inactive categories within the selected training compounds.

In the final optimization phase, CATALYST applies fine perturbations in the form of vectored feature rotation, adding a new feature and/or removing a feature, to selected hypotheses that survived the subtractive phase to find new models of enhanced bioactivity-to-compound mapping correlations. Eventually, CATALYST selects the highest-ranking models (10 by default) and presents them as the optimal pharmacophore hypotheses resulting from the particular automatic modeling run.³⁴

4.1.5. Assessment of the Generated Hypotheses. When generating hypotheses, CATALYST attempts to minimize a cost function consisting of three terms: weight cost, error cost, and configuration cost.^{30,34,64} Weight cost is a value that increases as the feature weight in a model deviates from

an ideal value of 2. The deviation between the estimated activities of the training set and their experimentally determined values adds to the error cost and is reflected in the correlation coefficient r (see Table 3).^{30,34,64} The activity of any compound can be estimated from a particular hypothesis through eq 3³⁴

$$\log(\text{Estimated activity}) = I + \text{Fit} \quad (3)$$

where I = the intercept of the regression line obtained by plotting the log of the biological activity of the training set compounds against the Fit values of the training compounds. The Fit value for any compound is obtained automatically employing eq 4^{30,34,35}

$$\text{Fit} = \sum \text{mapped hypothesis features} \times W[1 - \Sigma(\text{disp}/\text{tol})^2] \quad (4)$$

where Σ mapped hypothesis features represents the number of pharmacophore features that successfully superimpose (i.e., map or overlap with) corresponding chemical moieties within the fitted compound, and W is the weight of the corresponding hypothesis feature spheres. This value is fixed to 1.0 in CATALYST-generated models. disp is the distance between the center of a particular pharmacophoric sphere (feature centroid) and the center of the corresponding superimposed chemical moiety of the fitted compound; tol is the radius of the pharmacophoric feature sphere (known as tolerance, equal to 1.6 Å by default). $\Sigma(\text{disp}/\text{tol})^2$ is the summation of $(\text{disp}/\text{tol})^2$ values for all pharmacophoric features that successfully superimpose corresponding chemical functionalities in the fitted compound.^{30,34,64}

The third term, i.e., the configuration cost, penalizes the complexity of the hypothesis, i.e. the configuration cost. This is a fixed cost, which is equal to the entropy of the hypothesis space. The more the numbers of features (a maximum of five) in a generated hypothesis, the higher is the entropy with a subsequent increase in this cost. The overall cost (total cost) of a hypothesis is calculated by summing over the three cost factors. However, error cost is the main contributor to total cost. CATALYST calculates a fixed cost, which estimates the cost of the ideal hypothesis. CATALYST also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost and the closer the difference from the fixed cost (Table 3), the more likely that the hypothesis does not reflect a chance correlation.^{30,34,64} In a successful automatic modeling run, CATALYST ranks the generated models according to their total costs.^{30,34–36}

An additional approach to assess the quality of CATALYST-HYPOGEN pharmacophores is to cross-validate them using the Cat-Scramble algorithm implemented in CATALYST.^{62,63} This validation procedure is based on Fischer’s randomization test.⁶³ We implemented a 95% confidence level in this validation procedure, which instructs CATALYST to generate 19 random spreadsheets by the Cat-Scramble command. Subsequently, CATALYST-HYPOGEN is challenged to use these random spreadsheets to generate hypotheses using exactly the same features and parameters used in generating the initial unscrambled hypotheses.⁶⁸ Success in generating pharmacophores of comparable cost

Table 8. Tested Hits and Their Corresponding Fit Values, QSAR Predictions, and Experimental Bioactivities

hits ^a	NCI code	best fit values against ^b					QSAR estimates pIC ₅₀ ^c					experimental activity ^d		
		A-5-5	A-8-1	B-2-6	B-3-2	C-1-2	I	II	III	IV	V	% inhibiton at 10 μM	IC ₅₀ (μM)	pIC ₅₀
252	16365	0.00	0.00	7.96	0.00	0.00	-0.01	-1.70	2.02	-12.16	-5.29	32.00	---	---
254	22177	0.00	0.00	11.51	0.00	0.00	-1.47	-2.30	1.08	-15.02	-6.82	30.60	---	---
263	71896	0.00	0.00	10.69	0.00	0.00	-3.12	-6.52	-1.14	-13.50	-7.89	32.10	---	---
270	167770	0.00	0.00	11.36	0.00	0.00	-0.69	-1.56	1.43	-12.74	-6.45	29.40	---	---
272	170476	0.00	0.00	10.97	0.00	0.00	-2.15	-3.39	0.13	-13.69	-7.50	36.80	---	---
275	211290	0.00	0.00	10.79	0.00	3.77	-4.91	-6.61	-2.59	-11.57	-9.35	29.90	---	---
277	292448	0.00	0.00	10.92	0.00	0.00	-2.43	-3.35	-0.14	-12.67	-6.80	32.90	---	---
289	43112	0.00	0.00	0.00	9.80	0.00	-0.86	-5.50	0.72	-10.35	-8.04	58.10	7.34 (0.99)	-0.87
296	289755	6.44	0.00	0.00	8.62	0.30	0.57	-3.61	2.13	-10.08	-6.73	50.00	12.52 (0.92)	-1.10
300	404789	6.22	0.00	0.00	11.70	9.10	0.33	-2.48	1.97	-10.93	-6.71	86.40	1.84 (0.99)	-0.26
306	170116	0.00	0.00	0.00	0.00	9.14	-1.67	-2.24	0.57	-10.77	-6.85	29.30	---	---
307	284180	0.00	0.00	0.00	0.00	8.49	-2.88	-5.12	-0.52	-10.56	-7.85	30.40	---	---
311	369685	0.00	0.00	0.00	0.00	8.20	-0.14	-1.47	1.95	-9.84	-5.46	39.40	---	---
312	371477	0.00	0.00	0.00	0.00	7.49	-2.37	-4.73	-0.51	-8.63	-8.00	34.60	---	---
314	401405	0.00	0.00	0.00	0.00	9.14	-0.33	-1.67	1.88	-9.68	-5.92	46.10	---	---
315	606255	0.00	0.00	5.29	0.00	9.05	-2.64	-3.71	-0.52	-11.71	-7.73	38.20	---	---
317	607157	0.00	0.00	5.42	0.00	8.96	-2.94	-3.69	-0.75	-11.30	-7.81	32.60	---	---
318	607158	0.00	0.00	5.06	0.00	8.93	-3.32	-2.72	-0.99	-10.97	-6.84	42.90	---	---
319	608643	0.00	0.00	9.03	0.00	8.91	-1.98	-2.40	0.26	-6.01	-6.62	43.10	---	---
321	608650	0.00	0.00	8.63	0.00	8.92	-1.73	-2.26	0.51	-7.55	-6.51	32.30	---	---
322	608654	0.00	0.00	8.73	0.00	8.97	-2.11	-2.45	0.16	-7.28	-6.65	41.00	---	---
323	609241	0.00	0.00	7.23	0.00	9.02	-1.35	-2.13	0.84	-9.18	-6.49	45.50	---	---
325	609520	0.00	0.00	6.34	0.00	8.99	-3.32	-3.93	-1.10	-10.91	-7.84	34.70	---	---
326	657994	0.00	0.00	7.90	0.00	6.88	-1.75	-4.44	0.45	-10.90	-6.44	42.75	---	---
327	666331	0.00	0.00	0.00	0.00	8.17	-1.03	-2.70	1.42	-9.21	-5.42	31.30	---	---
127	oseltamivir	0.00	0.00	9.85	12.64	8.34	-1.30	-1.65	-1.32	-1.91	-1.70	100.00	0.16 nM (0.99)	0.80

^a Hit compounds are shown in Figure 8. ^b Best-fit values against each pharmacophore as calculated by eq 4 (section 4.1.5 Assessment of the Generated Hypotheses). ^c Table 4. ^d *In vitro* enzyme inhibition.

criteria to those produced by the original unscrambled data reduces the confidence in the nonrandomized training compounds and their pharmacophore models.⁶⁴

Table 3 shows the success ranges of different pharmacophore modeling trials, including their cost ranges (configuration, total, and fixed costs) as well as the costs of the corresponding null hypotheses. The table also shows the corresponding Fischer (Cat.scramble) confidence ranges for each modeling attempt. Based on Fischer randomization criteria, only 178 pharmacophores exceeded the 85% significance threshold for subsequent processing (clustering and QSAR analyses).

4.1.6. Clustering of the Generated Pharmacophore Hypotheses. Successful models (Cat.scramble \geq 85%) from each automatic pharmacophore modeling run were clustered into three groups, i.e., the total number of successful pharmacophores (178) were categorized into 57 groups, utilizing the hierarchical average linkage method available in CATALYST. Subsequently, the highest-ranking representatives, as judged based on their total costs, were selected to represent their corresponding clusters in subsequent processing. The best 57 representatives were directly employed in QSAR analysis to yield models **I** and **III**, in Table 4 and Table C in Supporting Information. Subsequently, 57 representatives were enrolled in a second hierarchical average linkage clustering cycle to select 12 pharmacophoric subgroups. Their best representative 12 models were further employed in a second round of QSAR analysis that culminated in models **II**, **IV**, and **V**, as in Table 4 and Table C in Supporting Information. Detailed information about all explored pharmacophores, including the best representatives that emerged in both clustering cycles, is shown in Table B in the Supporting Information. The table also shows the corresponding Cat.scramble confidence levels for each representative pharmacophore.

4.1.7. QSAR Modeling. 181 compounds were employed as a training set for QSAR modeling (see section 2.3 QSAR Modeling in the Results and Discussion). However, since it is essential to access the predictive power of the resulting QSAR models on an external set of inhibitors, the remaining 43 molecules (*ca.* 20% of the data set) were employed as an external test subset for validating the QSAR models (indicated with asterisks in Table A in the Supporting Information). The selection of the test set considers the fact that the molecules must represent a range of biological activities similar to that of the training set.

The chemical structures of the inhibitors were imported into CERIU2 as standard 3D single conformer representations in SD format. The selected conformers are of the lowest energy within the conformational ensemble generated by CATALYST. Upon importing the structures into CERIU2, we switched off the “Clean and Minimize” options in CERIU2 to preserve the conformers generated by CATALYST. Subsequently, different descriptor groups were calculated for each compound employing the C2.DESRIPTOR module of CERIU2. The calculated descriptors included various simple and valence connectivity indices, electrotopological state indices, single point quantum-mechanical descriptors (via the AM1 model), and other molecular descriptors (e.g., logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, etc.).⁷⁶

Training compounds were fitted against the best representative pharmacophores as clustered based on the scheme described earlier (see section 4.1.6 Clustering of the Generated Pharmacophore Hypotheses). Subsequently, the fit values, generated by the best-fit command within CATALYST (eq 4), were added as additional molecular descriptors.

Thereafter, genetic function approximation (GFA) was employed to search for the best possible QSAR regression

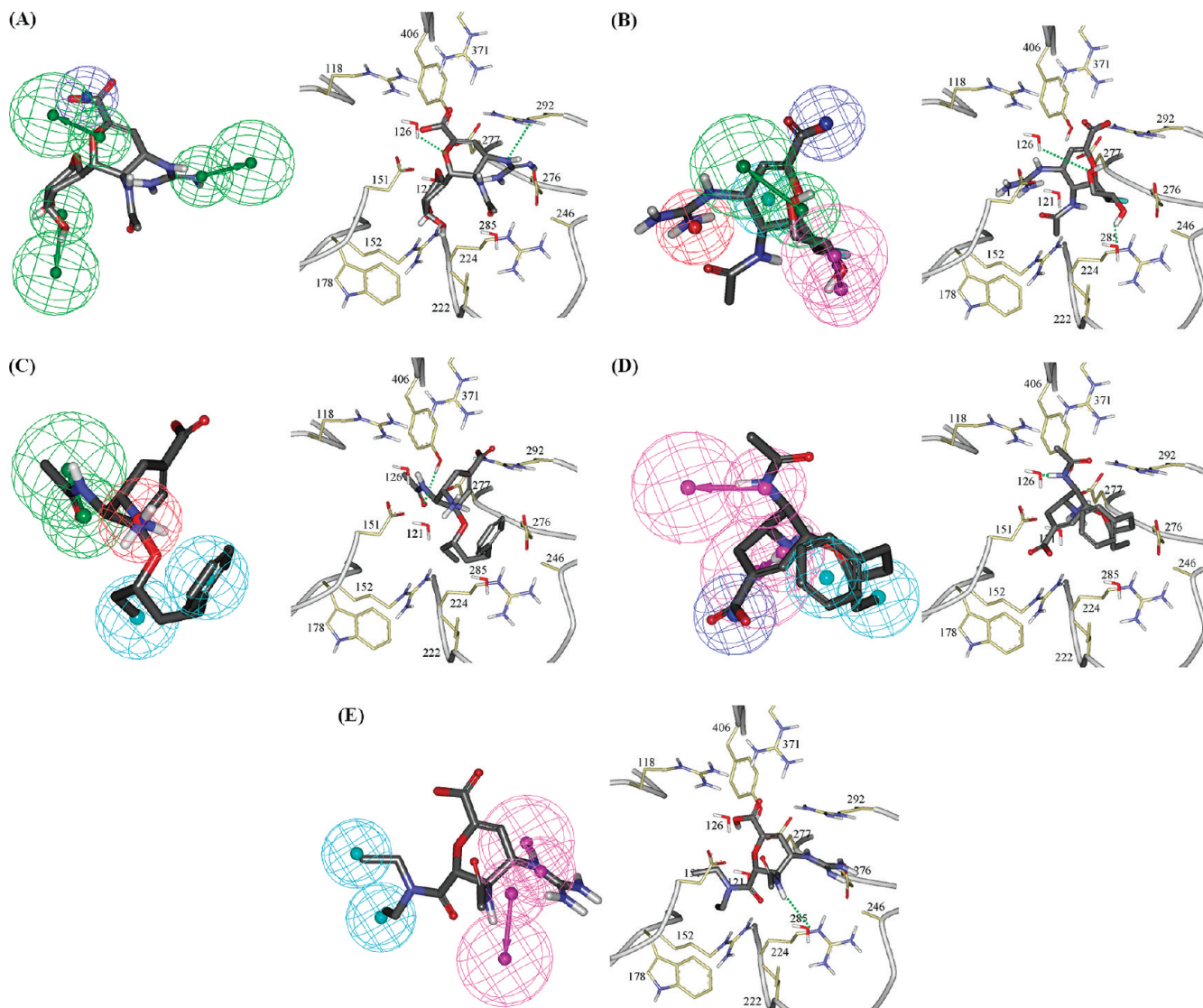


Figure 9. Mapping our QSAR-selected pharmacophore models against some potent training compounds compared to their corresponding docked poses into NA (protein databank code: 1NNC, resolution 1.8 Å). (A) A-5-5 and inhibitor 6, (B) A-8-1 and 99, (C) B-2-6 and 136, (D) B-3-2 and 136, and (E) C-1-2 and compound 37.

equation capable of correlating the variations in biological activities of the training compounds with variations in the generated descriptors, i.e., multiple linear regression modeling (MLR). The fitness function employed herein is based on Friedman's 'lack-of-fit' (LOF).^{75,76} Our preliminary diagnostic trials suggested the following optimal GFA parameters: explore linear equations at mating and mutation probabilities of 50%; population size = 500; number of genetic iterations = 30000, and LOF smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (QSAR descriptors), it was decided to scan and evaluate all possible QSAR models resulting from 10 to 20 explanatory terms.

All QSAR models were validated employing leave one-out cross-validation (r^2_{LOO}), bootstrapping (r^2_{BS}),^{69,76} and predictive r^2 (r^2_{PRESS}) calculated from the test subset. The predictive r^2_{PRESS} is defined as^{69,76}

$$r^2_{\text{PRESS}} = \text{SD-PRESS}/\text{SD} \quad (5)$$

where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of

the training set molecules, and PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set. Our GFA-MLR-based QSAR scanning procedure identified 5 successful QSAR models as the best equations. Table 4 shows the statistical performance of the successful QSAR models, while Figure 2 shows the plots of experimental versus fitted (training set) and predicted (testing set) bioactivities as calculated from the successful QSAR models. The corresponding statistical equations are shown in Table C in the Supporting Information. Five pharmacophore hypotheses emerged in the successful QSAR models as shown in Table 4. Table 5 shows the statistical and cost criteria of the emerged models, while Table 6 shows the 3D coordinates, weight, and tolerances of their pharmacophoric features.

4.1.8. In Silico Screening of the NCI Database for New Neuraminidase Inhibitors. The QSAR-based successful pharmacophoric hypotheses were complemented with ligand-shapes (see Figures 3-7 and Tables 5 and 6) and employed as 3D search queries against the NCI structural library. The shape component of the query was built using the CatShape module within Catalyst.³² To generate merged pharmacoph-

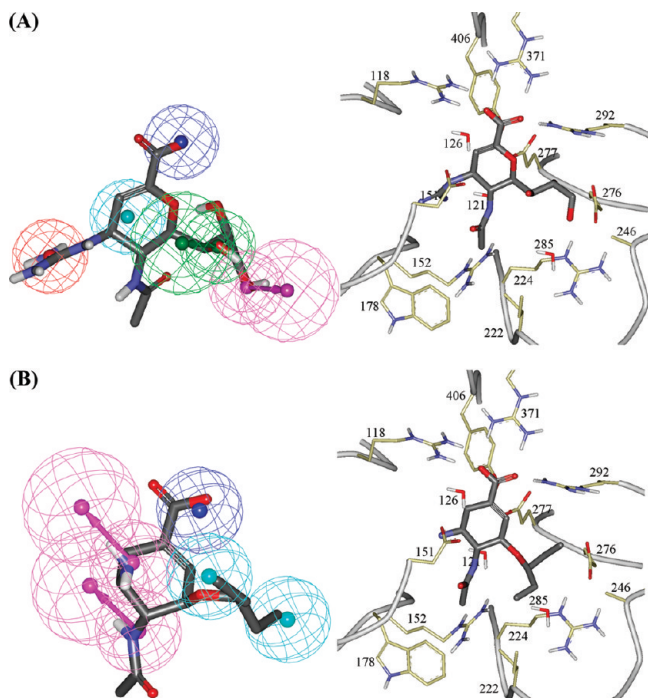


Figure 10. Mapping pharmacophores **A-8-1** and **B-3-2** against compounds **1** (zanamivir) **(A)** and **127** (oseltamivir) **(B)**, respectively, compared to the corresponding complexed crystallographic poses (protein databank codes: 1NNC and 2QWK, respectively, resolution 1.8 Å).

ore-shape queries, each successful pharmacophore was mapped against the most potent training inhibitor(s) from the corresponding training subset (i.e., from which the respective pharmacophore model was generated), and the best fitted conformer was used to generate shape constraints that were subsequently merged with the pharmacophore. In an attempt to generate the tightest possible shape queries, several shape-tolerance values were evaluated for each pharmacophore-shape model. Shape tolerance values that allowed the capture of a reasonable number of hits (more than 20 and less than 100 hits, see Table 7) were used for virtual screening.³⁰

The generated shape-pharmacophore models were employed as 3D search queries against the NCI database (around 230,000 compounds) using the “Best Flexible Database Search” option implemented within CATALYST. The resulting hits were requested from the NCI, and available compounds were screened for their anti-NA activities.

4.1.9. Docking Studies. The ionized forms of compounds **6**, **37**, **99**, **136**, **289**, **296**, and **300** were first fitted against the particular QSAR-selected pharmacophore(s) (i.e., **A-5-5**, **C-1-2**, **A-8-1**, **B-3-2**, or **B-2-6**), and the resulting fitted conformers were docked into the NA binding site (PDB code: 1NNC, resolution 1.8 Å) employing rigid body minimization. The docking experiments were performed employing LIGANDFIT docking engine and through default docking parameters and consensus scoring based on PLP1, PLP2, Ligsocre1, ligscore2, JAIN, and PMF scoring function.⁶⁷

4.2. Neuraminidase Enzyme Inhibition Assays. All reagents used were of fluorescence/analytical grade. Pure N9 NA crystals were generously provided by Dr. Graeme Laver (Australia) and Dr. Elspeth Garman (Oxford, U.K.).

NA enzyme activity and hits inhibition assays were conducted based on the method of Potier et al.⁷⁰ with slight

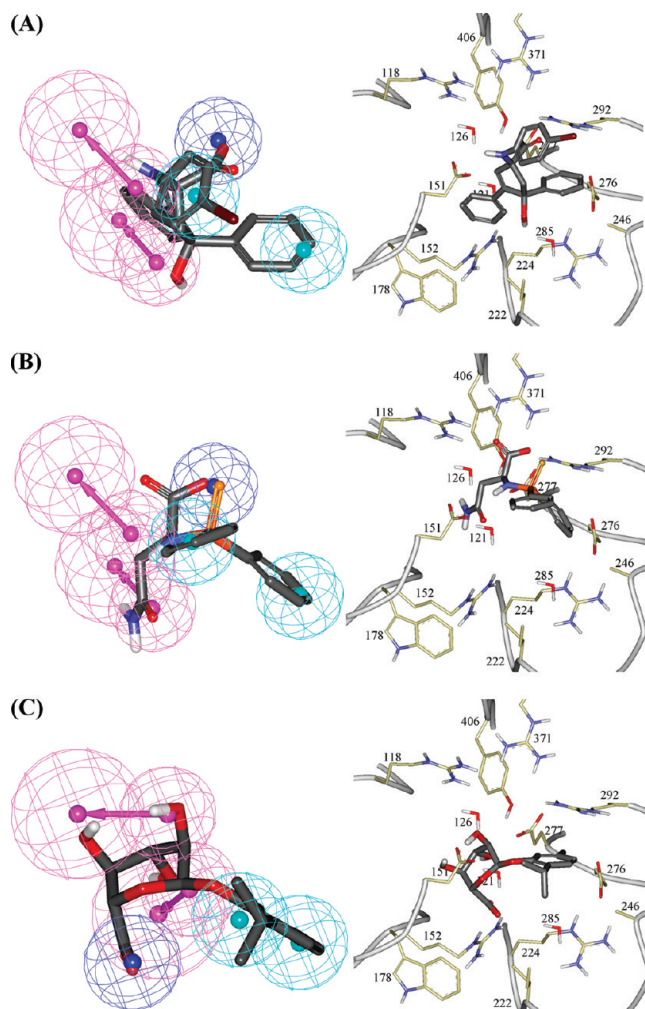


Figure 11. Mapping **B-3-2** against the most potent hits and the corresponding docked poses into NA (protein databank code: 1NNC, resolution 1.8 Å): **(A)** **239** ($IC_{50} = 7.6 \mu M$), **(B)** **296** ($IN_{50} = 12.51 \mu M$), and **(C)** **300** ($IC_{50} = 1.80 \mu M$).

modifications. Briefly, 10 μL of NA inhibitors and 10 μL of diluted enzyme were mixed and preincubated for 1 h at 37 °C in black 96-well plate (BD Falcon, U.S.A.). The final hit concentrations in the assay ranged from 10000 nM to 1 nM in serial half-log dilutions in either 0.02% DMSO in Tris buffer (pH = 7.5) or pure DMSO according to hit solubility. The reaction was started by adding 30 μL of the substrate 4-methylumbelliferyl-*N*-acetylneuraminic acid (MUNANA, Fluka) in 20 mM Tris-Cl buffer (pH = 7.5) containing 10 mM $CaCl_2$ to get the final substrate concentration of 75 μM . After 1 h at 37 °C, the reaction was stopped by the addition of 150 μL of freshly prepared 0.14 M NaOH in 83% ethanol. Fluorometric determinations were quantified with FLx800 BTI Microplate Fluorescence Reader (Bio-Tek Instruments, U.S.A.) by using an excitation wavelength of 360 ± 20 nm and an emission wavelength of 460 ± 20 nm. Substrate blanks were subtracted from each sample reading. Fluorescent values for uninhibited reactions were generally between 40,000 and 60,000 fluorescent units and were within the linear range of the assay. The fluorescence intensity of duplicate reaction with no inhibitor was used to determine the maximum (100%) neuraminidase activity.

The relationship between the concentration of inhibitor and the percent fluorescence inhibition was determined by plotting the percent neuraminidase activity relative to the

activity of an uninhibited control versus logarithm the inhibitor concentration, and the IC₅₀ values were obtained by extrapolation of data. All reactions were carried out in duplicate; the mean values were used in the analysis of the data.

ACKNOWLEDGMENT

The authors wish to thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds. The authors are also indebted to the National Cancer Institute for freely providing hit molecules for evaluation and to Prof. Graeme Laver (Australia) and Dr. Elspeth Garman (Oxford, U.K.) for kindly providing NA crystals for *in vitro* assay. We also thank Dr. Warren M. Kati, Dr. Pooran Chand, and Prof. Gillian Air for their useful information.

Supporting Information Available: (i) Chemical structures, biological activities, and bibliographical sources of the training compounds (224 structures); (ii) features and performances of all explored neuraminidase pharmacophore hypotheses; (iii) successful QSAR models, their statistical performances, and definitions of their descriptors; (iv) tested hits and their corresponding fit values, QSAR predictions, and experimental activities; and (v) chemical structures of the tested hit. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- Potter, C. W. Chronicle of influenza pandemics. In *Textbook of influenza*; Nicholson, K. G., Webster, R. G., Hay, A. J., Eds.; Blackwell Science: Oxford, England, 1998; pp 3–18.
- Lewis, D. B. Avian Flu to Human Influenza. *Annu. Rev. Med.* **2006**, *57*, 139–54.
- Laver, G.; Garman, E. Pandemic influenza: its origin and control. *Microbes Infect.* **2002**, *4*, 1309–1316.
- Oxford, J. S.; Novelli, P.; Sefton, A.; Lambkin, R. New Millennium Antivirals against Pandemic and Epidemic Influenza: The Neuraminidase Inhibitors. *Antiviral Chem. Chemother.* **2002**, *13*, 205–217.
- Bethell, R. C.; Smith, P. W. Sialidase as a target for inhibitors of influenza virus replication. *Expert Opin. Invest. Drugs* **1997**, *6*, 1501–1509.
- Garman, E.; Laver, G. The structure, function, and inhibition of influenza virus neuraminidase. *Protein Rev.* **2005**, *1*, 247–267.
- Ohuchi, M.; Asaoka, N.; Sakai, T.; Ohuchi, R. Roles of neuraminidase in the initial stage of influenza virus infection. *Microbes Infect.* **2006**, *8*, 1287–1293.
- Colman, P. M.; Varghese, J. N.; Laver, W. G. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature (London)* **1983**, *303*, 41.
- Air, G. M.; Laver, W. G. The Neuraminidase of Influenza Virus. *Proteins* **1989**, *6*, 341–356.
- Wade, R. C. 'Flu' and structure-based drug design. *Structure* **1997**, *5*, 1139–1145.
- Von, I. M.; Wu, W. Y.; Kok, G. B. Rational Design of Potent Sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363*, 418–423.
- Stoll, V.; Stewart, K. D.; Maring, C. J.; Muchmore, S.; Giranda, V.; Gu, Y.; Wang, G.; Chen, Y.; Sun, M.; Zhao, C.; Kennedy, A. L.; Madigan, D. L.; Xu, Y.; Saldivar, A.; Kati, W.; Laver, G.; Sowin, T.; Sham, H. L.; Greer, J.; Kempf, D. Influenza Neuraminidase Inhibitors: Structure-Based Design of a Novel Inhibitor Series. *Biochemistry* **2003**, *42*, 718–727.
- Kim, C.; Lew, W.; Williams, M.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M.; Mendel, D.; Tai, C.; Laver, W.; Stevens, R. Influenza Neuraminidase Inhibitors Possessing a Novel Hydrophobic Interaction in the Enzyme Active Site: Design, Synthesis, and Structural Analysis of Carbocyclic Sialic Acid Analogues with Potent Anti-Influenza Activity. *J. Am. Chem. Soc.* **1997**, *119*, 681–690.
- Arnold, S.; Monto, M. D. The Threat of an Avian Influenza Pandemic. *N. Engl. J. Med.* **2005**, *4*, 323–325, 352.
- Monto, A. S. Vaccines and Antiviral Drugs in Pandemic Preparedness. *Emerging Infect. Dis.* **2006**, *12*, 55–60.
- Le, Q. M.; Kiso, M.; Someya, K.; Sakai, Y. T.; Nguyen, T. H.; Nguyen, K. H. L.; Pham, N. D.; Ngyen, H. H.; Yamada, S.; Muramoto, Y.; Horimoto, T.; Takada, A.; Goto, H.; Suzuki, T.; Suzuki, Y.; Kawaoka, Y. Isolation of drug-resistant H5N1 virus. *Nature* **2005**, *437*, 1108.
- Beeley, N. R. A.; Sage, C. GPCRs: an update on structural approaches to drug discovery. *Targets* **2003**, *2*, 19–25.
- Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.
- Steuber, H.; Zentgraf, M.; Gerlach, C.; Sotriffer, C. A.; Heine, A.; Klebe, G. Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying conditions. *J. Mol. Biol.* **2006**, *363*, 174–187.
- Stubbs, M. T.; Reyda, S.; Dullweber, F.; Moller, M.; Klebe, G.; Dorsch, D.; Mederski, W.; Wurziger, H. pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. *ChemBioChem* **2002**, *3*, 246–249.
- DePristo, M. A.; de Bakker, P. I. W.; Blundell, T. L. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* **2004**, *12*, 831–838.
- Colman, P. Influenza virus neuraminidase: Structure, antibodies, and inhibitors. *Protein Sci.* **1994**, *3*, 1687–1696.
- Wilson, J.; von Itzstein, M. Recent Strategies in the Search for New Anti-Influenza Therapies. *Curr. Drug Targets* **2003**, *4*, 389–408.
- Schramm, V. Enzymatic transition state poise and transition state analogues. *Acc. Chem. Res.* **2003**, *36*, 588–596.
- Schramm, V. Enzymatic transition states and transition state analogues. *Curr. Opin. Struct. Biol.* **2005**, *15*, 604–613.
- Amyes, T.; Richard, J. Rational Design of Transition-State Analogues as Potent Enzyme Inhibitors with Therapeutic Applications. *ACS Chem. Biol.* **2007**, *2*, 711–714.
- Sutherland, J.; O'Brien, L.; Weaver, D. Pruned receptor surface models and pharmacophores for three-dimensional database searching. *J. Med. Chem.* **2004**, *47*, 3777–3787.
- Zhang, J.; Yu, K.; Zhua, W.; Jiang, H. Neuraminidase pharmacophore model derived from diverse classes of inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3009–3014.
- Steindl, T.; Langer, T. Influenza Virus Neuraminidase Inhibitors: Generation and Comparison of Structure-Based and Common Feature Pharmacophore Hypotheses and Their Application in Virtual Screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1849–1856.
- Catalyst User Guide, version 4.11*; Accelrys Software Inc.: San Diego, CA, 2005.
- Abu Hammad, A. M.; Affi, F. U.; Taha, M. O. Combining docking, scoring and molecular field analyses to probe influenza neuraminidase-ligand interactions. *J. Mol. Graphics* **2007**, *26*, 443–456.
- Hahn, M. Three-dimensional shape-based searching of conformationally flexible compounds. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 80–86.
- Moffat, K.; Gillet, V. J.; Whittle, M.; Bravi, G.; Leach, A. R. A Comparison of Field-Based Similarity Searching Methods: CatShape, FBSS, and ROCS. *J. Chem. Inf. Model.* **2008**, *48*, 719–729.
- Sprague, P. W.; Hoffmann, R. CATALYST Pharmacophore Models and Their Utility As Queries for Searching 3D Databases. In *Computer Assisted Lead Finding and Optimization*; Van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; Verlag Helvetica Chimica Acta: Zürich, Switzerland, 1997; pp 223–240.
- Smellie, A.; Teig, S.; Towbin, P. Poling: promoting conformational variation. *J. Comput. Chem.* **1995**, *16*, 171.
- Sutter, J.; Güner, O.; Hoffmann, R.; Li, H.; Waldman, M. Effect of Variable Weights and Tolerances on Predictive Model Generation. In *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: CA, 2000; pp 501–511.
- Singh, J.; Chuaqui, C. E.; Boriack-Sjodin, P. A.; Lee, W.; Pontz, T.; Corbley, M. J.; Cheung, H.; Arduini, R. M.; Mead, J. N.; Newman, M. N.; Papadatos, J. L.; Bowes, S.; Josiah, S.; Ling, L. E. Successful Shape-Based Virtual Screening: The Discovery of a Potent Inhibitor of the Type I TGF β Receptor Kinase (T β RI). *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4355–4359.
- Masuda, T.; Shibuya, S.; Arai, M.; Yoshida, S.; Tomozawa, T.; Ohno, A.; Yamashita, M.; Honda, T. Synthesis and Anti-Influenza Evaluation of Orally Active Bicyclic Ether Derivatives Related to Zanamivir. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 669–673.
- Wyatt, P. G.; Coomber, B. A.; Evans, D. N.; Jack, T.; Fulton, H. E.; Wonacott, A. J.; Colman, P.; Varghese, J. Sialidase inhibitors related to zanamivir further SAR studies of 4-amino-4H-pyran-2-carboxylic acid-6-propylamides. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 669–673.

- (40) Andrews, D. M.; Cherry, P.; Humber, D.; Jones, P.; Keeling, S.; Martin, P.; Shaw, C.; Swansom, S. Synthesis and influenza virus sialidase inhibitory activity of analogues of 4-guanidino-neu5ac2en-(zanamivir) modified in the glycerol side-chain. *Eur. J. Med. Chem.* **1999**, *34*, 563–574.
- (41) Taylor, N. R.; Cleasby, A.; Singh, O.; Skarzynski, T.; Wonacott, A. J.; Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Bethell, R.; Colman, P.; Varghese, J. Dihydropyranocarboxamides Related to Zanamivir: A New Series of Inhibitors of Influenza Virus Sialidases. 2. Crystallographic and Molecular Modeling Study of Complexes of 4-Amino-4H-pyran-6-carboxamides and Sialidase from Influenza Virus Types A and B. *J. Med. Chem.* **1998**, *41*, 798–807.
- (42) Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Starkey, I. D.; Copley, K. N.; Weston, H.; Scicinski, J.; Merritt, A.; Whittington, A.; Wyatt, P.; Taylor, N.; Green, D.; Bethell, R.; Madar, S.; Fenton, R. J.; Morley, P. J.; Ateman, T.; Beresford, A. Dihydropyranocarboxamides Related to Zanamivir: A New Series of Inhibitors of Influenza Virus Sialidases. 1. Discovery, Synthesis, Biological Activity, and Structure-Activity Relationships of 4-Guanidino- and 4-Amino-4H-pyran-6-carboxamides. *J. Med. Chem.* **1998**, *41*, 787–797.
- (43) Smith, P. W.; Starkey, I. D.; Howes, P. D.; Sollis, S. L.; Keeling, S. P.; Cherry, P. C.; von-Itzstein, M.; Wu, W. Y.; Jin, B. Synthesis and influenza virus sialidase inhibitory activity of analogues of 4-guanidino-Neu5Ac2en (GG167) with modified substituents. *Eur. J. Med. Chem.* **1996**, *31*, 143–150.
- (44) Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Copley, K. N.; Taylor, H.; Whittington, A. R.; Scicinski, J.; Bethell, R. C.; Taylor, N.; Skarzynski, T.; Cleasby, A.; Singh, O.; Wonacott, A.; Varghese, J.; Colman, P. Novel inhibitors of influenza sialidases related to GG167 Structure-Activity, Crystallographic and Molecular Dynamics Studies with 4H-Pyran-2-Carboxylic Acid 6-Carboxamides. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2931–2936.
- (45) Smith, P. W.; Whittington, A. R.; Sollis, S. L.; Howes, P. D.; Taylor, N. R. Novel inhibitors of influenza sialidases related to zanamivir. Heterocyclic replacements of the glycerol sidechain. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2239–2242.
- (46) Honda, T.; Masuda, T.; Yoshida, S.; Arai, M.; Kobayashita, Y.; Yamashita, M. Synthesis and Anti-Influenza Virus Activity of 4-Guanidino-7-substituted Neu5Ac2en Derivatives. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1921–1924.
- (47) Honda, T.; Masuda, T.; Yoshida, S.; Arai, M.; Kanekoa, S.; Yamashita, M. Synthesis and Anti-Influenza Virus Activity of 7-O-Alkylated Derivatives Related to Zanamivir. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1925–1928.
- (48) Kerrigan, S. A.; Pritchard, R. G.; Smith, P. W.; Stoodley, R. J. Synthesis of (3R*,4R*)-4-acetylamino-3-(pent-3-oxy)cyclohex-1-ene-1-carboxylic acid and its (3S*,4R*)-isomer and their inhibitory action against influenza virus sialidases. *Tetrahedron Lett.* **2001**, *42*, 7687–7690.
- (49) Kim, C. U.; Lew, W.; Williams, M. A.; Wu, H.; Zhang, L.; Chen, X.; Escarpe, P. A.; Mendel, D. B.; Laver, W. G.; Stevens, R. C. Structure-Activity Relationship Studies of Novel Carbocyclic Influenza Neuraminidase Inhibitors. *J. Med. Chem.* **1998**, *41*, 2451–2460.
- (50) Lew, W.; Wu, H.; Chen, X.; Graves, B. J.; Escarpe, P. A.; MacArthur, H. L.; Mendel, D. B.; Kim, C. U. Carbocyclic Influenza Neuraminidase Inhibitors Possessing a C3-Cyclic Amine Side Chain: Synthesis and Inhibitory Activity. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1257–1260.
- (51) Lew, W.; Wu, H.; Mendel, D. B.; Escarpe, P. A.; Chen, X.; Laver, W. G.; Graves, B. J.; Kim, C. U. A new series of C3-aza carbocyclic influenza neuraminidase inhibitors: synthesis and inhibitory activity. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3321–3324.
- (52) Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Stevens, R. C. Influenza Neuraminidase Inhibitors Possessing a Novel Hydrophobic Interaction in the Enzyme Active Site: Design, Synthesis, and Structural Analysis of Carbocyclic Sialic Acid Analogues with Potent Anti-Influenza Activity. *J. Am. Chem. Soc.* **1997**, *119*, 681–690.
- (53) Williams, M. A.; Lew, W.; Mendel, D. B.; Tai, C. Y.; Escarpe, P. A.; Laver, W. G.; Stevens, R. C.; Kim, C. U. Structure-activity relationships of carbocyclic influenza neuraminidase inhibitors. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 1837–1842.
- (54) Babu, Y. S.; Chand, P.; Bantia, S.; Kotian, P.; Dehghani, A.; El-Kattan, Y.; Lin, T.; Hutchison, T. L.; Elliott, A. J.; Parker, C. D.; Ananth, S. L.; Horn, L. L.; Laver, W. G.; Montgomery, J. A. BCX-1812 (RWJ-270201): Discovery of a Novel, Highly Potent, Orally Active, and Selective Influenza Neuraminidase Inhibitor through Structure-Based Drug Design. *J. Med. Chem.* **2000**, *43*, 3482–3486.
- (55) Chand, P.; Kotian, P. L.; Dehghani, A.; El-Kattan, Y.; Lin, T.; Hutchison, T. L.; Babu, Y. S.; Bantia, S.; Elliott, A. J.; Montgomery, J. A. Systematic Structure-Based Design and Stereoselective Synthesis of Novel Multisubstituted Cyclopentane Derivatives with Potent Antiinfluenza Activity. *J. Med. Chem.* **2001**, *44*, 4379–4392.
- (56) Chand, P.; Babu, Y. S.; Bantia, S.; Chu, N.; Cole, L. B.; Kotian, P. L.; Laver, W. G.; Montgomery, J. A.; Pathak, V. P.; Petty, S. L.; Shrouf, D. P.; Walsh, D. A.; Walsh, G. M. Design and Synthesis of Benzoic Acid Derivatives as Influenza Neuraminidase Inhibitors Using Structure-Based Drug Design. *J. Med. Chem.* **1997**, *40*, 4030–4052.
- (57) Atigadda, V. R.; Brouillette, W. J.; Duarte, F.; Babu, Y. S.; Bantia, S.; Chand, P.; Chu, N.; Montgomery, J. A.; Walsh, D. A.; Sudbeck, E.; Finley, J.; Air, G. M.; Luo, M.; Laver, G. W. Hydrophobic Benzoic Acids as Inhibitors of Influenza Neuraminidase. *Bioorg. Med. Chem.* **1999**, *7*, 2487–2497.
- (58) Howes, P.; Cleasby, A.; Evans, D.; Feilden, H.; Smith, P.; Sollis, S.; Taylor, N.; Wonacott, A. 4-Acetylamino-3-(imidazol-1-yl)-benzoic acid as novel inhibitors of influenza sialidase. *Eur. J. Med. Chem.* **1999**, *34*, 225–234.
- (59) 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. W. Potent Inhibition of Influenza Sialidase by a Benzoic Acid Containing a 2-Pyrrolidinone Substituent. *J. Med. Chem.* **1999**, *42*, 2332–2343.
- (60) Sheridan, R. P.; Kearsley, S. K. Why do we need so many chemical similarity search methods. *Drug Discovery Today* **2002**, *7*, 903.
- (61) Taha, M. O.; Qandil, A. M.; Zaki, D. D.; AlDamen, M. A. Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling. *Eur. J. Med. Chem.* **2005**, *40*, 701–727.
- (62) Poptodorov, K.; Luu, T.; Hoffmann R. Pharmacophore Model Generation Software Tools. In *Methods and Principles in Medicinal Chemistry: Pharmacophores and Pharmacophores Searches*; Langer, T., Hoffmann, R. D., Eds.; Wiley-VCH: Weinheim, Germany, 2006; Vol. 2, pp 17–44.
- (63) *The Principle of Experimentation Illustrated by a PsychoPhysical Experiment*, 8th ed.; Fischer, R., Ed.; Hafner Publishing Co.: New York, 1966.
- (64) Bersuker, I. B.; Bahçeci, S.; Boggs, J. E. The Electron-Conformational Method of Identification of Pharmacophore and Anti-Pharmacophore Shielding. In *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: CA, 2000; pp 457–473.
- (65) Taha, M. O.; Bustanji, Y.; Al-Ghoussein, M.; Mohammad, M.; Zalloum, H.; Al-Masri, I. M.; Atallah, N. Pharmacophore modeling, quantitative structure-activity relationship analysis, and in-silico screening reveal potent glycogen synthase kinase-3beta inhibitory activities for cimetidine, hydroxychloroquine, and gemifloxacin. *J. Med. Chem.* **2008**, *51*, 2062–77.
- (66) Mann, M. C.; Islam, T.; Dyason, J. C.; Florio, P.; Trower, C. J.; Thomson, R. J.; Von Itzstein, M. Unsaturated N-acetyl-D-glucosaminuronic acid glycosides as inhibitors of influenza virus sialidase. *Glycoconjugate J.* **2006**, *23*, 127–133.
- (67) *CERIUS2 4.10 LigandFit User Manual*; Accelrys Inc.: San Diego, CA, 2000. Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289–307.
- (68) Krovat, E. M.; Langer, T. Nonpeptide Angiotensin II Receptor Antagonists: Chemical Feature Based Pharmacophore Identification. *J. Med. Chem.* **2003**, *46*, 716–726.
- (69) Tropsha, A.; Gramatica, P.; Gombar, V. K. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003**, *22*, 69–76.
- (70) Potier, M.; Mamel, L.; Béliisle, M.; Melançon, S. B. Fluorometric Assay of Neuraminidase with a Sodium (4-Methylumbelliferyl- α -D-N-Acetylneuraminic) Substrate. *Anal. Biochem.* **1979**, *94*, 287–296.
- (71) Guha, R. On the interpretation and interpretability of quantitative structure-activity relationship models. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 857–871.
- (72) Taha, M. O.; Dahabiyeh, L. A.; Bustanji, Y.; Zalloum, H.; Saleh, S. Combining Ligand-Based Pharmacophore Modeling, QSAR Analysis and In-Silico Screening for the Discovery of New Potent Hormone Sensitive Lipase Inhibitors. *J. Med. Chem.* **2008**, *51*, 6478–6494.
- (73) Taha, M. O.; Atallah, N.; Al-Bakri, A. G.; Paradis-Bleau, C.; Zalloum, H.; Younis, K.; Levesque, R. C. Discovery of New MufR Inhibitors via Pharmacophore Modeling and QSAR Analysis followed by in silico screening. *Bioorg. Med. Chem.* **2008**, *16*, 1218–1235.

- (74) Taha, M. O.; Bustanji, Y.; Al-Bakri, A. G.; Yousef, M.; Zalloum, W. A.; Al-Masri, I. M.; Atallah, N. Discovery of new potent human protein tyrosine phosphatase inhibitors via pharmacophore and QSAR analysis followed by in silico screening. *J. Mol. Graphics Modell.* **2007**, *25*, 870–884.
- (75) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure Activity Relationships and Quantitative Structure Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
- (76) *CERIUS2 4.10 QSAR User Manual, version 4.10*; Accelrys Inc.: San Diego, CA, 2000.

CI8003682