Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2007) 443–456
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Combining docking, scoring and molecular field analyses to probe influenza neuraminidase–ligand interactions

Areej M. Abu Hammad, Fatma U. Afifi, Mutasem O. Taha *

Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Queen Rania Street, Amman 11942, Jordan

Received 15 November 2006; received in revised form 5 February 2007; accepted 6 February 2007
Available online 11 February 2007

Abstract

In this project, several docking conditions, scoring functions and corresponding protein-aligned molecular field analysis (CoMFA) models were evaluated for a diverse set of neuraminidase (NA) inhibitors. To this end, a group of inhibitors were docked into the active site of NA. The docked structures were utilized to construct a corresponding protein-aligned CoMFA models by employing probe-based (H+, OH, CH3) energy grids and genetic partial least squares (G/PLS) statistical analysis. A total of 16 different docking configurations were evaluated, of which some succeeded in producing self-consistent and predictive CoMFA models. However, the best model coincided with docking the ionized ligands into the hydrated form of the binding site via PLP1 scoring function ($r^2_{LOO} = 0.735$, r^2_{PRESS} against 24 test compounds = 0.828). The highest-ranking CoMFA models were employed to probe NA–ligand interactions. Further validation by comparison with a co-crystallized ligand–NA crystallographic structure was performed. This combination of docking/scoring/CoMFA modeling provided interesting insights into the binding of different NA inhibitors.

Keywords: Neuraminidase inhibitors; Docking; Scoring; Protein-aligned CoMFA; Multiple binding modes

1. Introduction

Influenza is a global health concern responsible for significant morbidity and mortality, particularly to the very young, elderly and immunosuppressed [1–3]. The development of vaccines to prevent or treat influenza infection has been significantly hindered by the high mutability of the virus [4]. Therefore, antiviral agents represent alternative treatment until a pandemic-specific vaccine becomes available [5].

The life cycle of the influenza virus provides several potential molecular targets for drug intervention [1]. Among those, neuraminidase (NA) appears to be particularly attractive [6,7].

NA is one of two glycoproteins expressed at the surface of the flu virus and is responsible for viral release from infected cells and viral transport through the mucus of the respiratory tract. NA also destroys hemaglutinin receptor on host cells, thus allowing the emergence of progeny virus particles from the infected cells [8,9]. Therefore, compounds that inhibit NA should protect the host from viral infection and retard its propagation. Fortunately, the amino acid residues within the active site of NA are conserved in all wild-type influenza viruses [10,11], which promotes NA as an excellent target for broad-spectrum inhibitors [7].

Structure-based drug design methods have played a critical role in the discovery of NA inhibitors [12]. By exploiting the crystal structure of NA–inhibitor complexes, many compounds were successfully modified and optimized based on the charge and shape characters of the binding site. Successful examples include Zanamivir [13] and Oseltamivir [14]. However, the continuous threat of imminent pandemics [15,16] and the emergence of resistant strains to Oseltamivir (Tamiflu®) render the development of new effective NA inhibitors attractive for drug development research [17].

Several series of NA inhibitors originating from a variety of scaffolds have been disclosed using structure-based methods and relying on crystallographic data, including: pyrrolidines [18], cyclopentanes [19], tetrahydrofurans [20] and benzenes [21].

Despite employing a variety of computational approaches towards the discovery of new NA inhibitors [12,22–39], molecular docking remains one of the most popular computa-

* Corresponding author. Tel.: +962 6 5355000x2505; fax: +962 6 5339649.
E-mail address: mutasem@ju.edu.jo (M.O. Taha).

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.02.002

tional methodologies for virtual screening and hit discovery [40,41].

The great recent interest in NA as anti-influenza target [42,43] combined with the inherent inaccuracies of docking studies, i.e., their inability to evaluate the binding free energies correctly [44,45], prompted us to evaluate an interesting approach to identify optimal docking conditions capable of explaining bioactivity variations across a diverse set of NA inhibitors. The central theme of this approach is to employ protein-aligned molecular field analysis (CoMFA) as a probe to validate corresponding docked conformers/poses [46,47]. If a particular docking study succeeded in aligning a group of molecules in such a way to access self-consistent and predictive CoMFA model(s), then one can assume the validity of the corresponding docking/scoring configuration. This is not unreasonable, as CoMFA models are highly sensitive to the way their training compounds are aligned in 3D space [48,49]. In fact, the issue of the 3D alignment is considered the most notorious among other factors that affect CoMFA performance [48,49].

Molecular docking is a conformational search and sampling procedure in which various docked conformations are explored to identify the right one as guided by a proper scoring function [50]. Unfortunately, scoring function(s) are generally incapable of evaluating the binding free energies correctly due to the high complexity of the underlying molecular interactions [45,51,52].

Another factor to be taken into consideration prior to docking is the presence or absence of crystallographically explicit water molecules in the binding site [53–55]. Furthermore, the fact that crystallographic structures lack information on hydrogen atoms means that it should be appropriately assumed whether the ligand's ionizable moieties will exist in their ionized form or not prior to docking [56].

The current study was conducted by docking 121 known NA inhibitors into a selected crystallographic structure of this enzyme (PDB code: 1NNC). It was decided to conduct the docking part utilizing the program LigandFit® [57,58], which was recently reported to be of good overall performance particularly in virtual high-throughput screening (vHTS) [52]. LigandFit® was instructed to select a maximum of ten distinct optimal conformers/poses for each docked inhibitor. Subsequently, four scoring functions (i.e., PLP [59], JAIN [60], LigScore [61] and PMF [62]) were separately employed to rank the optimal docked structures of each inhibitor. The highest-ranking conformers/poses, according to each scoring function, were aligned together to construct corresponding CoMFA models that were subsequently appropriately validated. The cycle of docking, scoring and CoMFA modeling was repeated to cover all possible docking combinations resulting from the presence (or absence) of crystallographically explicit water molecules within the binding site and the ionization states of the ligands (ionized or unionized).

The success in accessing a self-consistent CoMFA model(s) can serve as additional validation for a docking methodology to complement existing validation methods [63]. Generally, docking results are judged from the similarity of the docked conformers/poses to corresponding crystallographic structures or the ability to identify known active compounds from a random pool [64,65]. However, these criteria suffer from the implicit assumption that the crystallographic structures of bound ligands are sufficiently realistic to be used as reference standards. Although crystallographic data are considered reliable structural information for drug design, they are associated with some serious problems such as inadequate resolution [66] and crystallization-related artifacts [67–69]. Furthermore, crystallographic structures generally ignore structural heterogeneity related to protein anisotropic motion and discrete conformational sub-states [70].

2. Methods

2.1. Hardware and software

NA inhibitors were sketched in two-dimensional (2D) format using ChemDraw Ultra (Version 7.01) from Cambridge Soft Corp. (Cambridge, MA, www.cambridgesoft.com) installed on a PC. Docking, scoring and molecular field analysis studies were performed using CERIUS2 suite of programs (Version 4.10) from Accelrys Inc. (San Diego, CA, www.accelrys.com). CERIUS2 was installed on a Silicon Graphics Octane2 desktop workstation equipped with a 600 MHz MIPS R14000 processor (1.0 GB RAM) running the Irix 6.5 operating system.

2.2. Protein-aligned CoMFA modeling

2.2.1. Dataset

A set of 121 diverse neuraminidase inhibitors belonging to the following chemical classes: 5,6-dihydro-4H-pyranes, cyclohexenes, benzoic acid derivatives, pyridines and pyrrolidines [14,18,71–75] (Fig. 1 and Table A (in supplementary data)) were used for modeling. Only achiral inhibitors or those of known absolute chiral configurations were employed. The in vitro bioactivities of most collected inhibitors are expressed as the concentration of the test compound that inhibited NA activity by 50% (IC_{50}). To allow appropriate QSAR analysis, we were obliged to convert and normalize the bioactivities of all inhibitors into K_i format. Therefore, IC_{50} values were converted into K_i values employing the Cheng–Prusoff equation [24,34,76]:

$$K_i = \frac{IC_{50}}{1 + (S/K_m)} \quad (1)$$

where K_m is the Michaelis constant for a particular enzyme and S is the substrate concentration. Fig. 1 and Table A (in supplementary data) illustrate the structures of the collected inhibitors and their corresponding bioactivities.

The logarithm of K_i values (nM) were used in 3D-QSAR modeling, thus correlating the data linearly to the free energy change.

A training subset of 97 molecules was selected. However, since it is essential to access the predictive power of the resulting CoMFA models on an external set of inhibitors, the

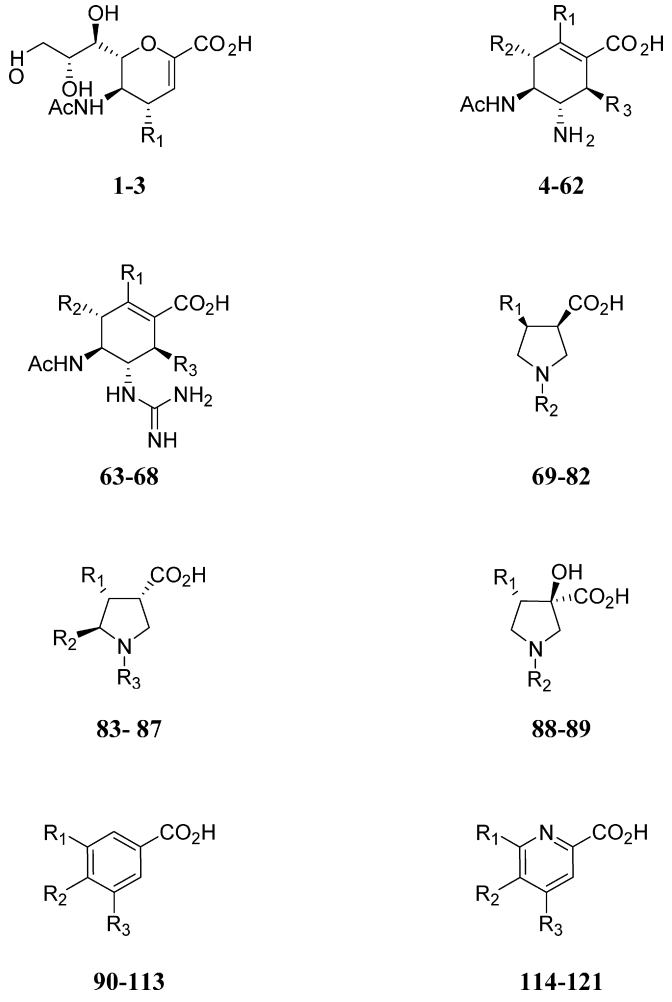

Fig. 1. The chemical scaffolds of different NA inhibitors used in modeling. See Table A (in supplementary data).

remaining 24 molecules (ca. 20% of the dataset) were employed as an external test subset for validating the 3D QSAR models. The test molecules were selected as follows: the inhibitors were ranked according to their K_i values and every fifth compound was selected for the test subset starting from the high-potency end. This selection considers the fact that the test molecules must represent a range of biological activities similar to that of the training set [77]. The following compounds were selected to validate CoMFA models: **15**, **17**, **19**, **22**, **23**, **32**, **33**, **35**, **38**, **39**, **61**, **62**, **64**, **70**, **72**, **75**, **80**, **84**, **92**, **97**, **112**, **117**, **120** and **121** (numbers are as in Fig. 1 and Table A (in supplementary data)).

2.2.2. Preparation of dataset

The 2D structures of the inhibitors (**1**–**121**, Fig. 1 and Table A (in supplementary data)) were sketched using ChemDraw Ultra (Version 7.01) and saved in MDL mol-file format. Subsequently, they were imported into CERIUS2 and converted to corresponding standard 3D structures using the rule based methods implemented in CERIUS2. Two protonation states were assumed for each inhibitor, ionized and unionized. In the ionized forms, the molecules were assumed to exist in neutral conditions (pH 7.0) and therefore carboxylic acids were deprotonated and given formal negative charges, while amino and guanidino groups were protonated and given formal positive charges. The remaining atoms were assigned partial charges using the default Gasteiger method implemented in CERIUS2 [78].

2.2.3. Neuraminidase (NA) crystal structure

The 3D coordinates of the wild type (N9) NA were retrieved from the Protein Data Bank (PDB code 1NNC). This particular structure was selected as it exhibits high 3D resolution (1.80 Å), complexed to potent NA inhibitor (**3**, GG167, see Fig. 1 and Table A (in supplementary data)) and reported to exhibit 87% promiscuity, i.e., a protein structure to which 87% of actives are docked correctly [31,93]. Hydrogen atoms were added to the protein utilizing CERIUS2 templates for protein residues. By default, this template ionizes amino, guanidino and carboxylic acid residues. This assumption is not unreasonable as the binding pocket is relatively shallow and hydrated, which minimize any influential effects on the pK_a values of the ionizable moieties. Gasteiger charges were assigned to the protein atoms as implemented within LigandFit [57,58]. The protein structure was utilized in subsequent docking experiments without energy minimization. Explicit water molecules were either kept or removed according to the required docking conditions, i.e., docking in the presence or absence of explicit water. The binding pocket involves five explicit water molecules of B-factors ranging from 12.45 to 36.74 (see supplementary data for their individual B-factors).

2.2.4. Docking simulations

LigandFit considers the flexibility of the ligand and treats the receptor as rigid. There are two steps implemented in the LigandFit® process [57,78]:

(1) Defining the location(s) of potential binding site(s): in the current docking experiments, the binding site was generated from the co-crystallized ligand (GG167), employing the default settings of LigandFit® [57,78], as shown in Fig. 2.
(2) Docking the ligands in the binding site: in LigandFit®, docking is composed of few major sub-steps [57,78]: (i) conformational search of the flexible ligands employing Monte Carlo randomized process; (ii) pose/conformation selection based on shape similarity with the binding site; (iii) candidate conformers/poses exhibiting low shape discrepancy are further enrolled in calculation of the dock energies by a grid-based interpolation scheme; (iv) each docked conformation/pose is further fitted into the binding pocket through a number of rigid-body minimization (RBM) iterations; (v) dissimilar docked conformers/poses of best-ranking docking energies (DockScores) are selected, further energy minimized via BGFS RBM and saved.

In the current docking experiments, all 121 ligands were docked into the binding sites employing following docking configuration: (i) Monte Carlo search parameters: 20,000 trials and torsion steps with polar hydrogens = 30.0°; (ii) The RMS threshold for ligand-to-binding site shape match was set to 2.0 employing a maximum of 1.0 binding site partitions; (iii)

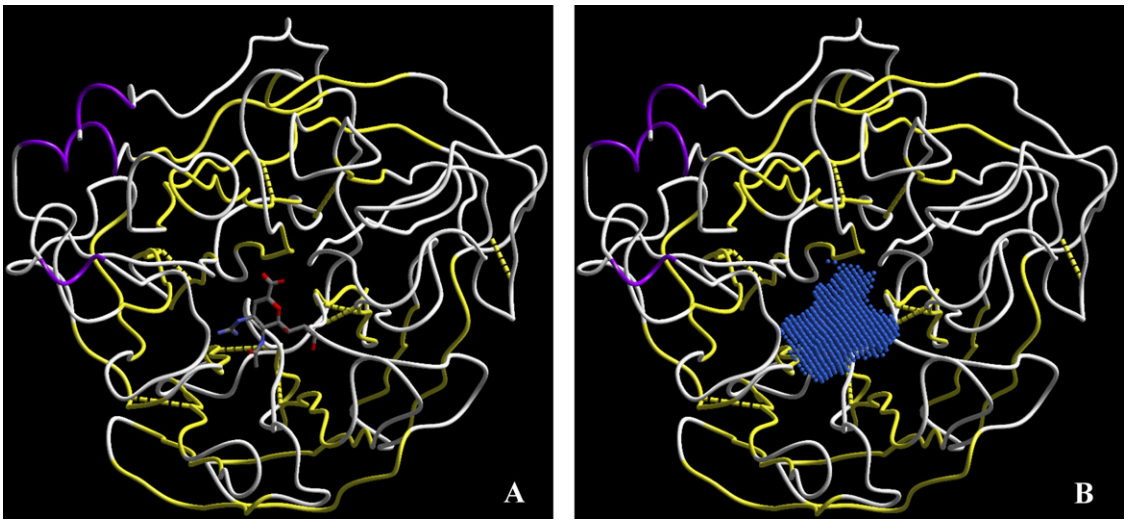

Fig. 2. Perspective cartoon views of NA (PDB code: 1NNC): (A) co-crystallized with compound **3** (Fig. 1, Table A (in supplementary data)) and (B) the corresponding binding pocket (dotted blue area).

interaction energies were assessed employing CFF force field (Version 95 implemented in CERIUS2 4.10) with a non-bonded cutoff distance of 10.0 Å and distance dependent dielectric. An energy grid extending 3.0 Å from the binding site was implemented; (iv) rigid body ligand minimization parameters were: 10 iterations of steepest descent (SD) minimization followed by 20 BFGS iterations applied to every successful orientation of the docked ligand; (v) a maximum of 10 diverse docked conformations/poses of optimal interaction energies were saved. The similarity threshold was set to a DockScore of 20 kcal/mol and an RMS value of 1.5 Å; (vi) the saved conformers/poses were further energy-minimized within the binding site for a maximum of 100 rigid-body iterations.

2.2.5. *Scoring of docked conformers/poses*

For all optimal docked conformers/poses, we computed scores using the scoring functions, PLP1 [59], JAIN [60], LigScore1 [61] and PMF [62]. The first two are empirical first-principle physical interaction terms, while LigScore1 is an empirical regression-based scoring function and PMF is a knowledge-based potential.

Considering each scoring function in turn, the highest scoring docked conformer/pose was selected for each inhibitor for subsequent CoMFA modeling. This resulted in four sets of 121 docked molecules with scores corresponding to each scoring function. LigScore1 scores were calculated employing CFF force field (Version 95) and using grid-based energies with a grid extension of 5.0 Å across the binding site. PMF scores were calculated employing cutoff distances for carbon–carbon interactions and other interactions of 12.0 Å.

2.2.6. *Molecular field analysis*

The molecular field analysis (CoMFA) and G/PLS modules within CERIUS2 were used to perform 3D QSAR analyses [79]. The alignments of different inhibitors came directly from the top-scoring conformers/poses according to each considered scoring function and docking condition (ligands' ionization and binding site hydration state). For each alignment, the interaction fields between the ligands and proton (positively charged), hydrogen-bond donor/acceptor and methyl (neutral) probes were calculated employing a regularly spaced rectangular grid of 2.0 Å spacing. The spatial limits of the molecular field were defined automatically and were extended past the van der Waals volume of all the molecules in the *X*, *Y* and *Z* directions. The ligands were assigned partial charges using the Gasteiger method implemented within CERIUS2. The energy fields were calculated employing the default UNIVERSAL force field (Version 1.02) implemented within CERIUS2 [80] and were truncated to ±50 kcal/mol. The calculation gave around 2187 variables for each compound (ca. 729 variable/probe).

To derive the best possible 3D QSAR statistical model for each alignment, we used genetic partial least squares (G/PLS) analysis to search for optimal regression equations capable of correlating the variations in biological activities of the training compounds with variations in the corresponding interaction fields [81]. G/PLS is derived from two methods: genetic function approximation (GFA) and partial least squares (PLS). GFA techniques rely on the evolutionary operations of "crossover and mutation" to select optimal combinations of descriptors (i.e., chromosomes) capable of explaining bioactivity variation among training compounds from a large pool of possible descriptor combinations (i.e., chromosomes population). Each chromosome is associated with a fitness value that reflects how good it is compared to other solutions. The fitness function employed herein is based on Friedman's "lack-of-fit" (LOF) [81]. G/PLS algorithm uses GFA to select appropriate basis functions to be used to model certain data, and PLS regression as the fitting technique to weigh the basis functions' relative contributions in the final model. Application of G/PLS allows the construction of larger QSAR equations while avoiding over-fitting and eliminating most variables [81]. Preliminary diagnostic trials suggested the following optimal G/PLS parameters: explore linear equations at mating and

mutation probabilities of 50%; population size = 500; number of generations (iterations) = 30,000 and LOF smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (i.e., CoMFA descriptors) and PLS latent variables (i.e., principle components), it was decided to scan and evaluate all possible CoMFA models resulting from all combinations of 1–6 principal components and numbers of CoMFA descriptors ranging from 16 to 22 field variables, i.e., a total of 42 different CoMFA models were evaluated in each case. This evaluation was repeated for all molecular alignments resulting from different scoring functions and docking conditions.

All CoMFA models were validated employing leave one-out cross-validation (r^2_{LOO}), bootstrapping (r^2_{BS}) [82,83] and predictive r^2 (r^2_{PRESS}) calculated from the test subsets. The test molecules were aligned according to the particular docking/scoring configuration, and their bioactivities were predicted by corresponding G/PLS models generated from the training sets and employing the above-mentioned range of combinations of principle components and CoMFA descriptors. The predictive r^2_{PRESS} is defined as [82]:

$$r^2_{PRESS} = \frac{SD - PRESS}{SD} \quad (2)$$

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 sum of squared deviations between predicted and actual bioactivity values for every molecule in the test set.

A particular CoMFA model is defined as successful if it exhibits r^2_{LOO} and π^2_{PRESS} values above 0.60. However, CoMFA models corresponding to the best r^2_{PRESS} values in a particular alignment were further cross-validated through dividing the training set into five groups of approximately the same size in which the objects were assigned randomly. Subsequently, 80% of the training compounds were randomly selected, and a model was generated, which was then used to predict the remaining compounds within the training set (leave-20%-out). This process was repeated 10 times, and the average predictive r^2 ($r^2_{L-20\%-O}$) is determined. This cross-validation technique has been shown to yield better indices for the robustness of a model than the normal LOO procedure [84]. An additional validation was also performed for high-ranking 3D QSAR models to rule out the possibility of chance correlations: all biological activities were randomized 99 times (confidence level 99%) and were subjected to regression analysis, and the mean randomization r^2 was calculated [85].

3. Results

As described above, we evaluated the effects different docking approaches (i.e., scoring functions, ligand ionization states and binding site hydration) on the statistical qualities of the corresponding protein-aligned CoMFA models constructed from diverse NA inhibitors. The compounds were docked into the binding site employing LigandFit®, which was instructed to explore the stabilities of wide variety of potential docked conformers/poses and to tightly minimize promising conformers/poses into the binding pocket. Subsequently, the docked structures were scored employing four different scoring functions and aligned together for molecular field analysis (CoMFA) utilizing genetic partial least squares (G/PLS) for statistical modeling. The inhibitors were divided into two groups: a training subset of 97 compounds and a randomly selected test subset of 24 compounds. The test compounds were selected to represent a range of biological activities similar to that of the training set (see Section 2.2.1). Several G/PLS settings were scanned for each molecular alignment, i.e., CoMFA models resulting from all combinations of 1–6 principal components and numbers of CoMFA descriptors ranging 16–22 were evaluated. The qualities of each CoMFA model were assessed via the following statistical criteria: (i) conventional regression coefficients against the training compounds (r^2_{97}), (ii) leave one-out regression coefficients (r^2_{LOO}), (iii) bootstrapping regression coefficients (r^2_{BS}), (iv) predictive regression coefficients against the external set of 24 test compounds (r^2_{PRESS}) and (v) the sum of squared deviations between predicted and experimental bioactivities for molecules in the test set (PRESS). Table 1 shows the statistical criteria of the resulting CoMFA models.

Despite that traditional cross-validation tests such as r^2_{LOO} are very useful [86], they do not always pick up poor equations [83,87]. Accordingly, we decided to further validate superior models that combine r^2_{LOO} and $r^2_{PRESS} \geq 0.60$ (Table 1) by calculating their leave-20%-out ($r^2_{L\text{-}20\%\text{-}O}$) and randomization correlation coefficients (r^2_{random}). These tests ensure that the generated regression models were not produced by chance [83]. Furthermore, to evaluate the robustness of a particular docking-based alignment, we introduced an additional success criterion: the ability of the alignment to access self-consistent and predictive CoMFA models regardless to variations in the allowed number of principal components or field variables. This criterion is represented in Table 1 by the success rate column, which shows the ratios of satisfactory statistical models compared to the total number of scanned CoMFAs for a particular alignment.

Table 1 shows that several docking conditions accessed satisfactory CoMFA models. However, the best CoMFAs were generated by docking the ionized ligands, regardless to the hydration state of the binding pocket (docking settings **A** as in Table 1). Furthermore, ionizing the ligands allowed access to superior CoMFAs via two scoring functions, i.e., PLP and JAIN, as shown in Table 1.

On the other hand, the unionized ligands accessed satisfactory CoMFA models only upon docking into the hydrated binding pocket and exclusively via PLP1 scoring function. Nevertheless, this configuration illustrated excellent robustness and statistical significance, as evident from Tables 1 and 2.

Figs. 3 (and B and C in supplementary data) illustrate the experimental bioactivities versus fitted (97 training set) and predicted (24 testing compounds) values produced by the best

Table 1
The statistical results of the best CoMFA models obtained via various docking/scoring combinations

Ligand ionization state	Explicit water	Scoring function	Code[a]	LV[b]	CoMFA descriptors[c]	r^2_{97} [d]	r^2_{LOO} [e]	r^2_{BS} [f]	r^2_{PRESS} [g]	PRESS[h]	$r^2_{L\text{-}20\%\text{-}O}$ [i]	r^2_{Random} [j]	Success rate[k]
Ionized	Present	JAIN	A1	2	16	0.763	0.704	0.763	0.676	24.167	0.698	0.150	5/42
		LigScore1	A2	1	17	0.696	0.645	0.696	0.621	28.267	0.572	0.160	1/42
		PLP1[l]	**A3**	**4**	**20**	**0.827**	**0.735**	**0.811**	**0.828**	**12.820**	**0.690**	**0.186**	**11/42**
		PMF	A4	4	18	0.847	0.692	0.802	0.606	29.344	0.754	0.181	1/42
	Absent	**JAIN**[l]	**A5**	**5**	**16**	**0.804**	**0.650**	**0.747**	**0.754**	**18.339**	**0.669**	**0.158**	**11/42**
		LigScore1	A6	1	19	0.665	0.612	0.665	0.711	21.518	0.613	0.182	5/42
		PLP1	A7	2	22	0.776	0.707	0.776	0.521	35.690	–	0.212	0/42
		PMF	A8	1	19	0.692	0.628	0.692	0.661	25.301	0.572	0.179	2/42
Unionized	Present	JAIN	B1	1	21	0.647	0.572	0.647	0.613	28.811	–	0.209	0/42
		LigScore1	B2	4	22	0.825	0.641	0.795	0.553	33.317	–	0.221	0/42
		PLP1[l]	**B3**	**1**	**21**	**0.746**	**0.677**	**0.746**	**0.729**	**20.230**	**0.655**	**0.205**	**11/42**
		PMF	B4	1	22	0.658	0.563	0.658	0.544	33.990	–	0.215	0/42
	Absent	JAIN	B5	2	21	0.748	0.530	0.671	0.552	33.366	–	0.199	0/42
		LigScore1	B6	1	19	0.656	0.568	0.656	0.626	27.862	–	0.179	0/42
		PLP1	B7	3	21	0.695	0.504	0.627	0.709	21.651	–	0.206	0/42
		PMF	B8	2	19	0.738	0.619	0.738	0.596	30.087	0.599	0.178	1/42

[a] These codes are used in the captions of tables and figures and within the text to indicate the corresponding docking/scoring conditions.
[b] Optimal number of latent variables at 30,000 generations of G/PLS.
[c] Number of CoMFA descriptors (molecular interaction field descriptors) in the optimal 3D-QSAR equation.
[d] Non-cross-validated correlation coefficients for 97 training compounds.
[e] Cross-validation correlation coefficients determined by the leave-one-out technique.
[f] Bootstrapping correlation coefficients.
[g] Predictive r^2 determined for the 24 test compounds.
[h] The sum of squared deviations between predicted and actual activity values for every molecule in the test set of 24 compounds.
[i] Cross-validation correlation coefficients determined by the leave-20%-out technique.
[j] Average randomization correlation coefficients: the biological activities were randomized 99 times and the mean randomization r^2 was calculated.
[k] Number of successful CoMFA models (of r^2_{LOO} and $r^2_{PRESS} \geq 0.6$) per total number of evaluated QSAR equations (42 trials).
[l] Bold statistical parameters correspond to the best docking/scoring combinations.

CoMFA models (**A3**, **A5** and **B3** conditions, Tables 1 and 2). The robustness of the resulting alignments is clearly obvious as they maintained high-quality CoMFA models despite variations in the number of principal components and field descriptors, as shown in Table 2. Undoubtedly, successful protein-aligned CoMFA models reflect realistic corresponding docked conformers/poses. Figs. 4 (and D in supplementary data) show the alignments of the inhibitors into the binding pocket of NA as proposed by the best docking/scoring configurations.

Notably, knowledge-based PMF performed significantly less than empirical PLP1 and JAIN, which hints to certain uniqueness in how NA binds to its respective ligands, i.e.,

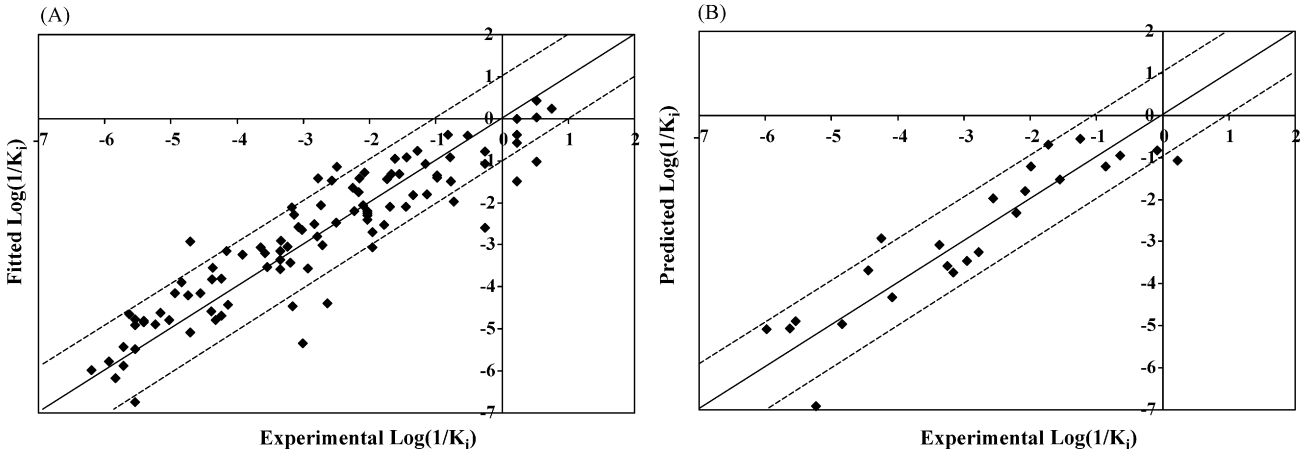

Fig. 3. Experimental vs. fitted (A, **97** compounds, $r^2_{LOO} = 0.735$) and predicted (B, 24 compounds, $r^2_{PRESS} = 0.828$) bioactivities calculated from the best CoMFA model obtained after 30,000 iterations of G/PLS performed on the training compounds as aligned by **A3** docking/scoring conditions (code as in Table 1, hydrated binding pocket, ionized ligands and PLP1 scoring function). The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the ±1.0 log point error margins.

Table 2
Effects of varying the number of principle components and field descriptors on the statistical criteria of the 3D-QSAR equations generated from successful docking-based alignments

Docking conditions	Scoring function	LV	Number of terms	r^2_{97}	r^2_{LOO}	r^2_{BS}	r^2_{PRESS}	PRESS
Ionized-hydrated	JAIN	2	16	0.763	0.704	0.763	0.676	24.167
		3	17	0.821	0.731	0.813	0.663	25.082
		3	16	0.812	0.694	0.774	0.655	25.735
		6	22	0.883	0.708	0.811	0.625	26.936
		5	16	0.835	0.729	0.803	0.604	29.498
	PLP1	4	20	0.827	0.735	0.811	0.828	12.831
		3	18	0.782	0.682	0.754	0.783	16.148
		1	19	0.684	0.613	0.683	0.730	20.835
		6	17	0.843	0.731	0.810	0.709	21.681
		1	17	0.688	0.617	0.688	0.690	23.070
Ionized un-hydrated	JAIN	6	19	0.832	0.649	0.771	0.777	16.652
		5	16	0.804	0.650	0.747	0.754	18.339
		5	22	0.830	0.652	0.741	0.717	21.096
		3	18	0.793	0.685	0.775	0.683	23.597
		1	21	0.656	0.589	0.656	0.682	23.662
	PLP1	1	19	0.665	0.612	0.665	0.711	21.518
		1	16	0.697	0.621	0.697	0.708	21.765
		1	21	0.685	0.610	0.684	0.666	24.854
		1	17	0.702	0.641	0.701	0.665	24.939
		1	20	0.699	0.621	0.699	0.645	26.443
Unionized-hydrated	PLP1	1	21	0.746	0.677	0.746	0.729	20.230
		1	16	0.711	0.655	0.711	0.713	21.358
		1	17	0.737	0.680	0.737	0.701	22.304
		3	21	0.812	0.727	0.813	0.691	23.015
		3	19	0.834	0.590	0.680	0.671	24.491

compared to other ligand–protein complexes in the protein databank. Still, PLP1 was generally reported to yield superior docking accuracies when compared to other docking/scoring combinations [52,88].

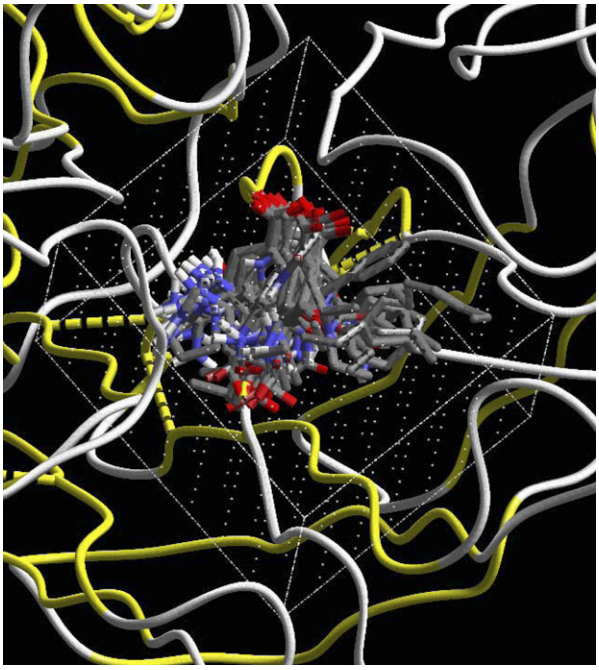

Fig. 4. Alignment of some docked inhibitors within the binding pocket of NA as proposed by **A3** docking scoring configurations (codes as in Table 1). The figure also shows the CoMFA grid box.

4. Discussion

By evaluating the binding interactions proposed by the highest-performing docking approaches, one can envisage the binding compartment of NA as a shallow pouch comprised of four primary regions involved in binding with docked ligands (Figs. 6–14 and Figs. E–L in supplementary data).

The first region (R1) is comprised of the positively charged guanidino groups of arginines 118, 292 and 371. Confronting R1 is a mainly hydrogen-bond forming region (R2) consisting of the side chain of Arg152 and the amidic carbonyl of Trp178 and Asp151. The third binding region (R3) is a negatively charged inner arch nearly perpendicular to the chord connecting R1 and R2 and composed of Glu119, Asp151 and Glu227, Glu276, Glu277 and Tyr406. This site contains three explicit water molecules: one is adjacent to Asp 151 (B-factor = 32.94) while the other two lay between Glu227 and Glu277 (B-factors 12.45 and 16.26, respectively). The fourth region (R4) is outer hydrophobic and situated adjacent to R2 and comprised of the side chains of Ala246, Ile222, Trp178 and the hydrophobic propylene linker of Arg224. Two more explicit water molecules (B-factors 33.80 and 36.74) are located within R4 between Ala246 and Ile222 and hydrogen bonded to the amidic carbonyl of Asn221. Fig. 5 shows a schematic representation of the binding pocket.

All binding regions were implicated in substrate recognition, however, Asp151, Glu277 and Tyr406 are believed to play a critical role in the catalytic activity of NA [89,90].

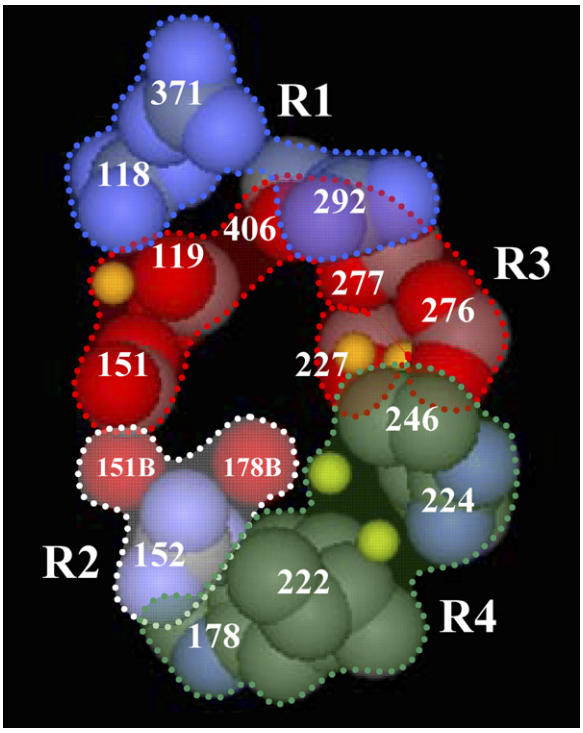

Fig. 5. A presentation of binding pocket of NA showing the different binding regions. The positively charged arginine triad (**R1**) is shown in blue, the hydrogen-bonding area (**R2**) is shown in white (151B and 178B represent the backbone carbonyls of amino acids 151 and 178, respectively), at a deeper level, the negative zone (R3) is shown in red and the hydrophobic region is represented in green (R4). The crater of the binding pocket is the middle gap. Water molecules are shown as yellow balls.

4.1. Ligand–NA interactions proposed by the best docking/scoring conditions

Emergence of three high-quality docking/scoring configurations (i.e., **A3**, **A5** and **B3**) is quite interesting. The validity of certain docking approach can be judged upon based on how close the corresponding CoMFA-based estimate (fitted or predicted bioactivity) to the experimental counterpart. However, if several docking/scoring conditions succeeded in estimating the bioactivity of certain ligand(s), then this probably encodes for multiple binding modes assumed by the ligand within the binding site.

We selected several diverse NA inhibitors (from Fig. 1 and Table A (in supplementary data)) to probe and discuss their binding interactions within NA binding pocket.

The fact that **A3** conditions accessed satisfactory bioactivity estimates for **31** (IC_{50} = 0.30 nM), **67** (IC_{50} = 0.50 nM) and **68** (IC_{50} = 0.50 nM) (Table 3) validate the corresponding docked conformers/poses (Figs. 6–8, respectively). **A3** conditions directed the carboxylate of **67** (Fig. 7) towards R1 to form ionic and hydrogen bonding interactions with the corresponding arginine triad, while the opposite acetamido carbonyl was allowed to form a network of hydrogen bonds involving Asp151 and a water molecule bound to Glu227. On the other hand, the guanidino group of **67** was directed towards the negative zone R3 to interact with Glu276 and Glu277. The hydrophobic moiety of **67** (2-butyl) was directed towards the hydrophobic R4.

Table 3
Experimentally determined bioactivities vs. their estimated counterparts for some NA inhibitors

Inhibitors[a]	Experimental bioactivities, $\log(1/K_i)$[b]	Estimated bioactivities, $\log(1/K_i)$[c]		
		A3[d]	**A5**[d]	**B3**[d]
3	0.223	−0.382	−1.338	−1.620
14	−0.254	−2.601	−0.391	−1.297
24	−0.254	−1.069	−1.530	−0.664
31	0.746	0.236	−0.366	−0.020
32	−0.856	−2.930	−3.345	−3.544
67	0.524	0.430	−0.150	−1.297
68	0.524	0.034	−1.173	−0.243
88	−2.787	−2.802	−2.783	−2.956
118	−2.839	−2.516	−2.817	−3.331

[a] The structures are as in Fig. 1 and Table A (in supplementary data).
[b] The bioactivities are normalized in K_i format (as mentioned in Section 2.2.1).
[c] Estimated bioactivities by the best docking/scoring/CoMFA models (see Tables 1 and 2).
[d] The codes correspond to docking/scoring conditions as in Table 1.

Similarly, the carboxylate of **68** (Fig. 8) was projected by **A3** conditions towards R1 to form hydrogen-bond reinforced ionic interactions with the guanidino of Arg292. However, unlike **67**, **A3** conditions directed the acetamido group of **68** towards R2 to allow mutual hydrogen-bonding interactions with Arg152. Unsurprisingly, the guanidino and 3-pentyl moieties of **68** were directed by **A3** towards R3 and R4, respectively. The guanidine is electrostatically bound to Asp151, Glu227 and the amidic carbonyl of Trp178.

However, **A3** flips some cyclohexene inhibitors, particularly those of bulky hydrophobic substituents, e.g., **31** (Fig. 6), in

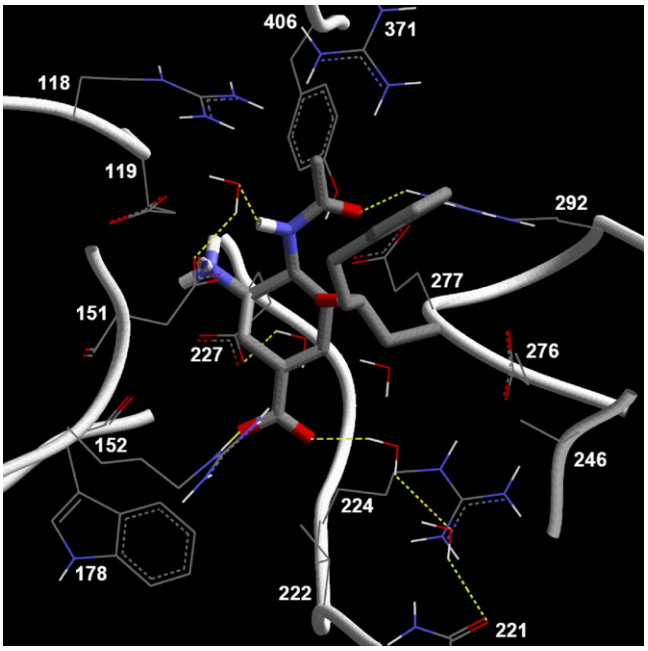

Fig. 6. The docked conformer/pose of inhibitor **31** (IC_{50} = 0.30 nM) as generated by **A3** conditions. A wire-frame detailed view showing the major amino acid residues involved in the binding interactions.

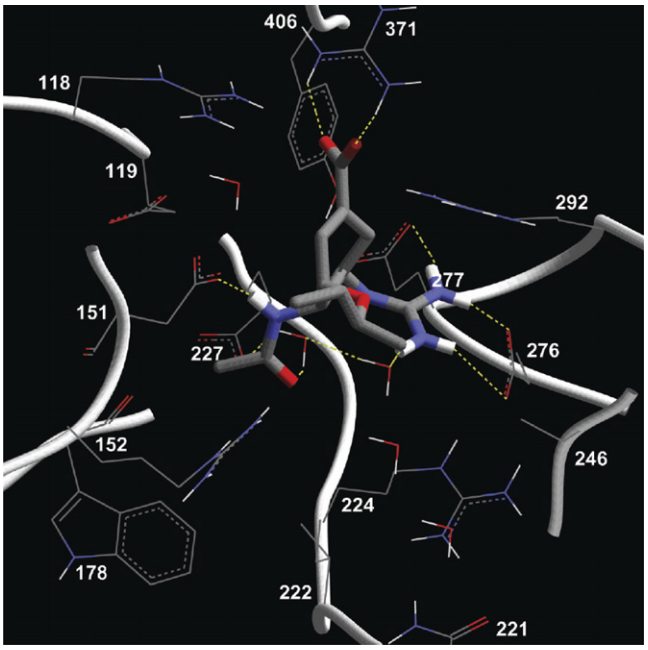

Fig. 7. The docked conformer/pose of inhibitor **67** (IC_{50} = 0.50 nM) as generated by **A3** conditions. A wire-frame detailed view showing the major amino acid residues involved in the binding interactions.

such a way that their carboxylates are bonded to Arg152 in R2 while their bulky hydrophobic fragments are allowed to protrude outside the binding pocket. However, this docking approach directs the acetamido of **31** to be involved in a hydrogen-bond network with Arg292 (in R1) and a water molecule bound to Glu119, while the amino moiety interacts with Asp151 (in R3).

The fact that **A5** conditions (docking the ionized ligands after dehydrating the binding pocket, Table 1) succeeded in estimating the bioactivities of a significant number of inhibitors (e.g., **14**, IC_{50} = 3.0 nM; **32**, IC_{50} = 12.0 nM, Table 3) indicates that these inhibitors eject certain water molecule(s) upon binding. For example, **A5** conditions forced the carboxylate group of **14** (Fig. 9) into R1, and directed the adjacent fluorine substituent to form hydrogen-bond with Arg292. The amine substituent was directed towards Asp151 and Glu227 in R3, while the acetamido moiety was forced deep into the binding pocket where it replaces a water molecule (B-factor 16.26) embedded in R3. The 3-pentyl side chain of **14** protrudes outside the binding pocket, albeit towards the hydrophobic region R4. Similarly, inhibitor **32** (Fig. 10) is anchored by **A5** conditions in such a way that the carboxylate and the acetamido moieties are directed towards R1 (Arg118 and Arg371) and R2 (Arg152), respectively. The amino group is embedded within R3 via electrostatic interactions with Glu119, while the hydrophobic aryl alkyl ether is directed into R4 where it displaces the same water molecule ejected by **14**.

The repeated displacement of the same static explicit water molecule by several potent inhibitors supports the proposition that binding of some NA inhibitors requires ejecting water molecules from the binding site. Emergence of **B3** (i.e., docking the unionized ligands into the hydrated form of the binding pocket) as successful docking/scoring condition (Table 3) is quite surprising, particularly since the binding pocket is shallow and populated with strongly ionizable residues [91]. Masking the ligands' ionized moieties in this docking configuration leaves hydrogen-bonding as the major determinant of the docked conformations/poses. Therefore, the ability of this approach to achieve self-consistent and predictive

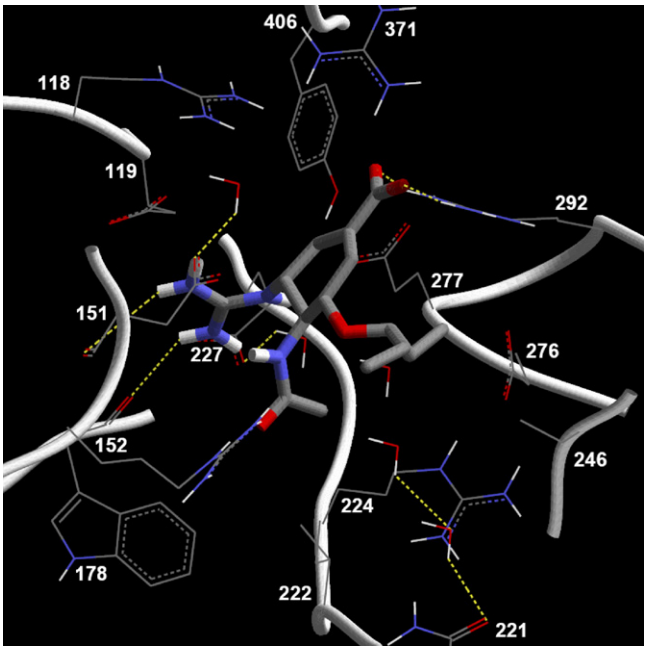

Fig. 8. The docked conformer/pose of inhibitor **68** (IC_{50} = 0.50 nM) as generated by **A3** conditions. A wire-frame detailed view showing the major amino acid residues involved in the binding interactions.

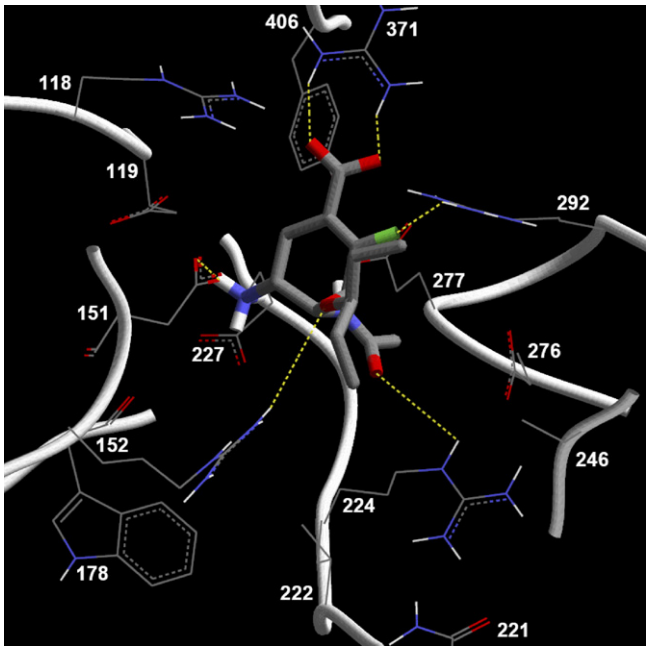

Fig. 9. The docked conformer/pose of inhibitor **14** (IC_{50} = 3.0 nM) as generated by **A5** conditions. A wire-frame detailed view showing the major amino acid residues involved in the binding interactions.

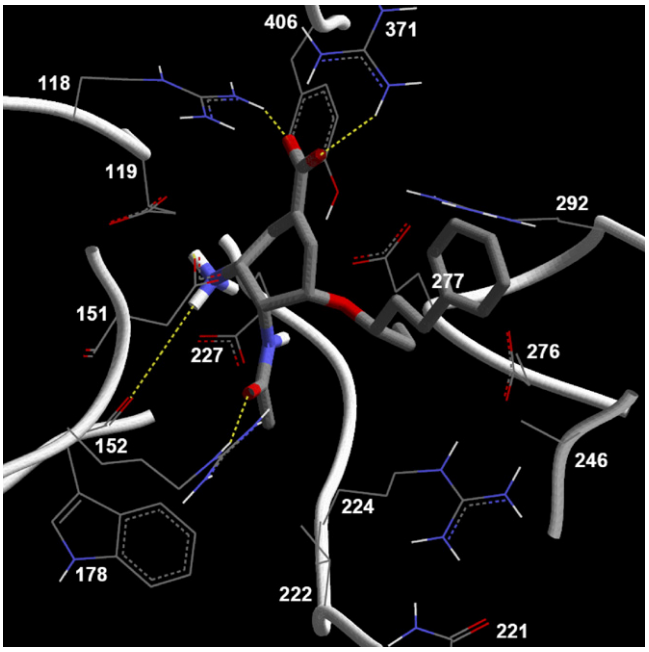

Fig. 10. The docked conformer/pose of inhibitor **32** (IC_{50} = 12.0 nM) as generated by **A5** conditions. A wire-frame detailed view showing the major amino acid residues involved in the binding interactions.

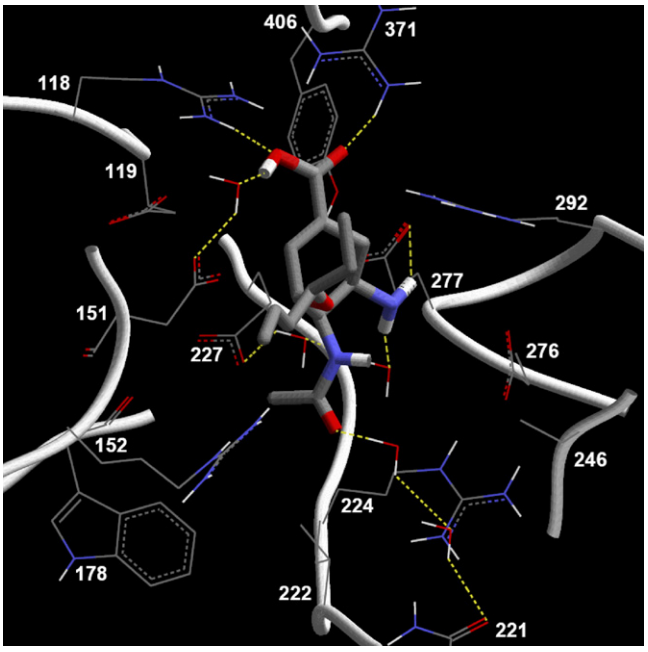

Fig. 11. The docked conformer/pose of inhibitor **24** (IC_{50} = 3.0 nM) as generated by **B3** conditions. A wire-frame detailed view showing the major amino acid residues involved in the binding interactions.

CoMFA models highlights the importance of hydrogen-bonding in ligand–NA interactions. Careful examination of successful conformers/poses produced by this docking configuration indicates that hydrogen-bonding gains significance in ligand–NA affinity whenever oppositely charged electrostatic moieties of the ligand(s) are placed at closer proximities. Apparently, oppositely charged adjacent groups favor intramolecular over intermolecular electrostatic attraction, and therefore magnify the role of hydrogen bonding in ligand–NA binding. This trend is particularly evident in amino-substituted ligands compared to those substituted with guanidino groups. This can be exemplified by inhibitor **24** (IC_{50} = 3.0 nM), which had its bioactivity closely estimated by **B3**/CoMFA modeling (Fig. 11; Table 3). The carboxylic acid group of **24** is oriented towards R1 and involved in a network of hydrogen bonds with Agr371, Arg118 and a water molecule bound to Asp151. On the other hand, the opposite acetamido moiety is directed towards R2 where it is involved in hydrogen-bonding to a string of two water molecules attached to Asn221. Similarly, the amino group of **24** is hydrogen-bonded directly to Glu277 and to Glu227 via two water molecules. The hydrophobic fragment of **24** is forced by **B3** conditions outside the binding pocket.

In some cases, all three docking configurations accessed close bioactivity estimates to the experimental values pointing to the validity of all corresponding conformers/poses. This conduct hints to the existence of multiple binding modes whereby a particular ligand keeps perturbing from one conformer/pose to another. The pyrrolidine derivative **88** (Fig. 12; Table 3; K_i = 360 nM) and the benzoic acid analogue **118** (Fig. 13; Table 3; IC_{50} = 4.0 μM) exemplify this behavior. Inhibitor **88** is

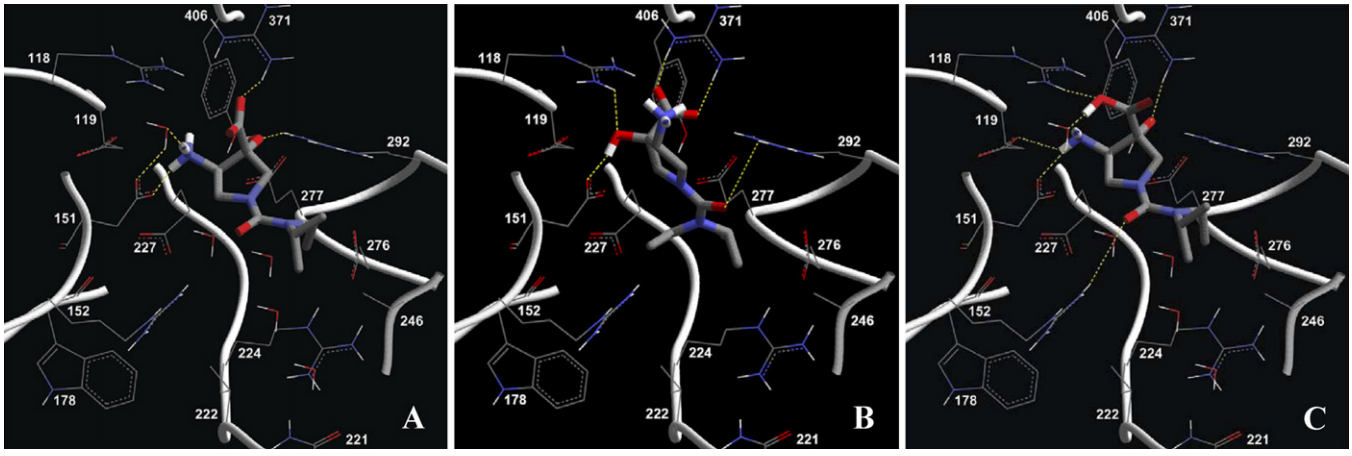

Fig. 12. The docked conformers/poses of inhibitor **88** (K_i = 360 nM) as generated by: (A) **A3**, (B) **A5** and (C) **B3** docking/scoring conditions.

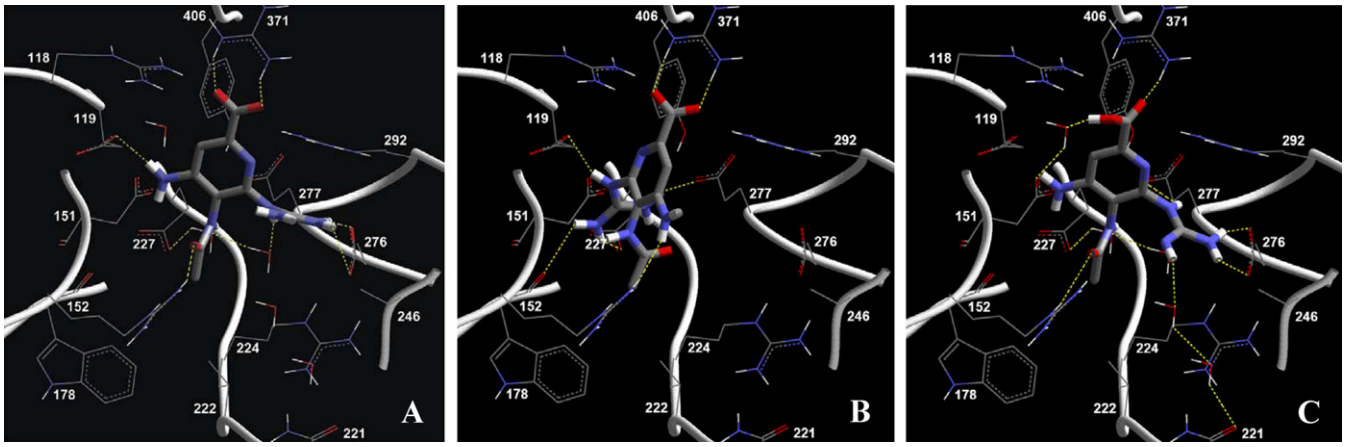

Fig. 13. The docked conformers/poses of inhibitor **118** (IC_{50} = 4.0 μM) as generated by: (A) **A3**, (B) **A5** and (C) **B3** docking/scoring conditions.

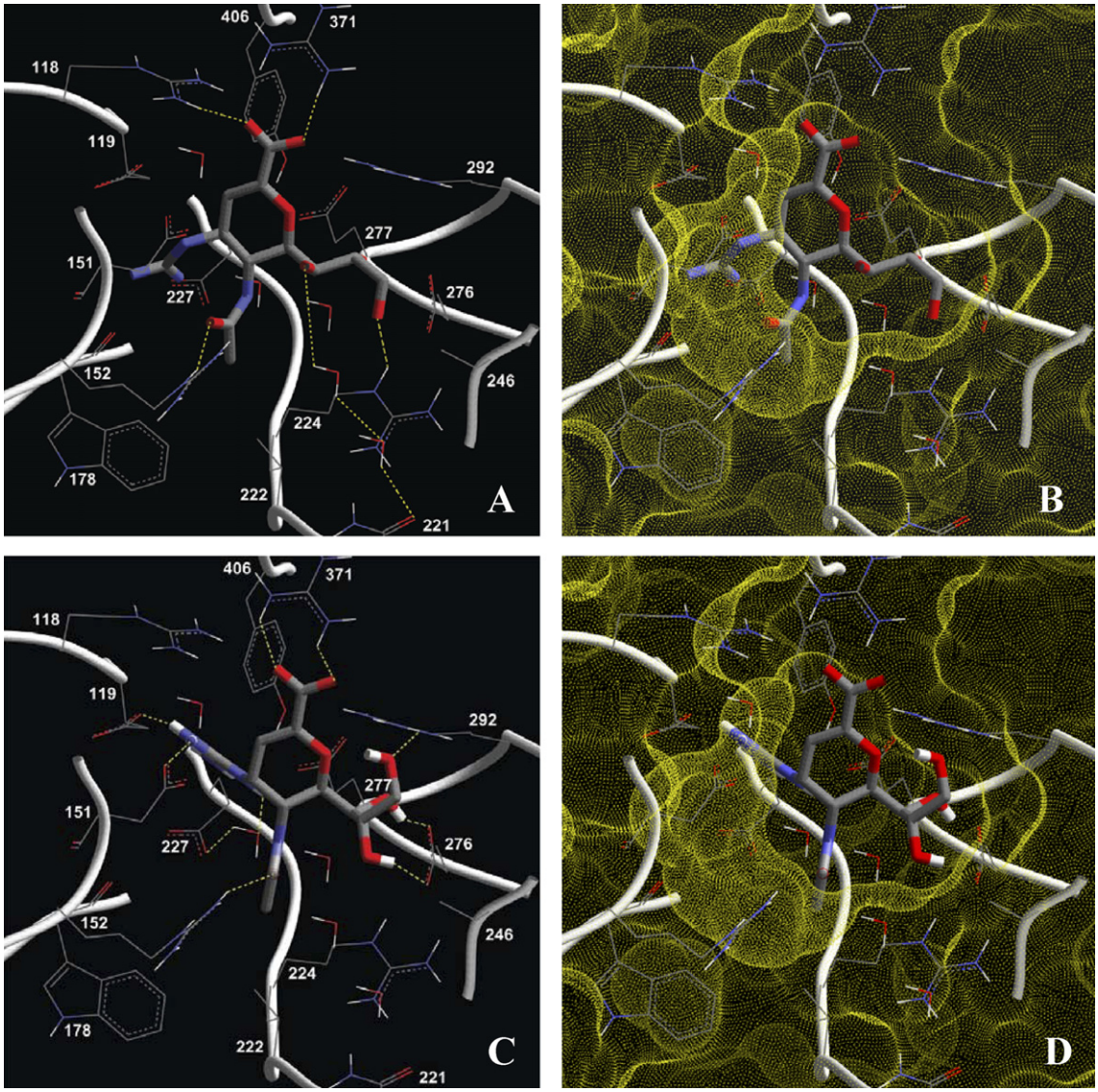

Fig. 14. (A and B) The crystallographic structure of inhibitor **3** (GG167, IC_{50} = 1.0 nM) complexed within the binding pocket of NA (PDB code 1NNC). (C and D) The docked conformer/pose of inhibitor **3** as generated by **A3** docking/scoring conditions.

fitted by the three methods in such a way so that the carboxyl group is placed within R1, while the opposite urea carbonyl is oriented towards R2 to form weak hydrogen-bonding interactions with Arg152. However, its amino substituent has the freedom to shift from R3 (where it interacts with Asp 151) and to protrude outside the binding pocket. The hydrophobic fragment of **88** is allowed to extend towards R4 in all three docking configurations. Interestingly, **A3** and **B3** (Fig. 12a and c) conditions produced nearly identical conformers/poses for **88** suggesting minimal electrostatic contribution to the binding interaction (see discussion about **B3**). Apparently, the close proximity between the anionic carboxylate and cationic ammonium groups within **88** favors intra-molecular electrostatic attraction over intermolecular ionic affinity to the receptor. Furthermore, the presence of an α-hydroxyl (to the carboxyl group) creates a triangle of substituents (OH, NH_2 and COOH) involved in significant intra-molecular hydrogen-bonding and therefore disfavors intermolecular hydrogen-bonding with the binding pocket, which is responsible for the poor bioactivity of this inhibitor.

Comparably, all three docking configurations anchored the carboxyl group of inhibitor **118** within R1 and its *p*-acetamido towards R2 (Fig. 13). However, the positions of the *m*-guanidino and *m*-amino groups vary across the three docking conditions giving the impression that the guanidino is allowed to freely slide within the negatives R3 zone.

4.2. Comparison of the best docking/scoring conditions with corresponding crystallographic structure

The availability of crystallographic image for NA complexed to **3** (PDB code: 1NNC), prompted us compare the conformer/pose produced by the best docking/scoring approach with that seen in the crystallographic structure. Apparently, **A3** conditions gave the best bioactivity estimate for this ligand suggesting a realistic corresponding docked conformer/pose (Table 3). Fig. 14 shows the docked and co-crystallized structures of **3**. Clearly from the figure, the two structures share very similar binding interactions: the carboxylate group is oriented towards R1 to form electrostatic and hydrogen bonding interactions within the arginine triad, while the acetamido carbonyl is hydrogen-bonded to the guanidino of Arg152 in R2. The glycerol side chain in both structures is projected between R3 and R4 where it forms some hydrogen-bonding and hydrophobic interactions. However, a noticeable difference can be seen in the position of the guanidine side chain, still, it remained within R3 in both cases. In the co-crystallized structure, the guanidine group forms electrostatic and hydrogen bonding interactions with Asp151, Glu227 and the backbone carbonyl of Trp178. While it is shifted by **A3** conditions towards the carboxylates of Glu119 and Glu277. This difference is consistent with our observations regarding the perturbation of guanidino groups in some NA inhibitors within R3 (as discussed with **118** and **88**). The close similarity between the docked conformer/pose of **3** and the corresponding co-crystallized structure validates our overall docking/scoring/CoMFA approach as a probe for ligand–NA interactions.

5. Conclusions

This effort included elaborate evaluation of several docking conditions and scoring functions and their corresponding protein-aligned CoMFA models constructed from diverse influenza NA inhibitors. A total of 16 different docking configurations were evaluated, of which some succeeded in producing self-consistent and predictive CoMFA models. However, the best 3D-QSAR model coincided with docking the ionized ligands into the hydrated form of the binding pocket via PLP1 scoring function. The best CoMFA models were employed as tools to identify docked conformers/poses most relevant to the observed bioactivities.

Unsurprisingly, this approach revealed different binding modes for NA inhibitors, which can be attributed to the symmetrical nature of the binding pocket [18,92]. Furthermore, we noticed that minor variations in the chemical structures of the inhibitors yielded significant variations in the docked poses, which explains the apparent complexity of SAR within NA inhibitors. Moreover, several inhibitors seem to eject water molecules upon binding.

The validity of our docking/scoring/CoMFA approach was highlighted by comparing the crystallographic structure of a co-crystallized ligand–NA complex with the corresponding docked conformer/pose suggested by this method.

This modeling combination of docking/scoring and CoMFA should help in discovering new NA inhibitors through high throughput docking of virtual molecular libraries employing optimal docking configurations, followed by prioritizing hit compounds according to their CoMFA-predicted bioactivities. Furthermore, the results of this study can be used to discard odd looking docking-based conformers/poses of screened virtual molecules. We are currently in the process of employing this modeling approach for the discovery of new potent NA inhibitors.

Acknowledgments

This project was partially sponsored by the Faculty of Graduate Studies (M.Sc. Thesis of Areej M. Abu Hammad). The authors wish to thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for providing funds towards purchasing O2 and Octane2 Sgi workstations and CATALYST and CERIUS2 software packages.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2007.02.002.

References

[1] K.G. Nicholson, R.G. Webster, A.J. Hay, Textbook of Influenza, Blackwell Science, Oxford, 1998.
[2] D.B. Lewis, Avian flu to human influenza, Annu. Rev. Med. 57 (2006) 139–154.

[3] G. Laver, E. Garman, Pandemic influenza: its origin and control, Microbes Infect. 4 (2002) 1309–1316.
[4] J.M. Katz, S. Garg, S. Sambhara, Influenza vaccines: current and future strategies, in: Influenza Virology, Caister Academic Press, Wymondham, UK, 2006, pp. 203–228.
[5] J.S. Oxford, P. Novelli, A. Sefton, R. Lambkin, New millennium antivirals against pandemic and epidemic influenza: the neuraminidase inhibitors, Antivir. Chem. Chemother. 13 (2002) 205–217.
[6] P.M. Colman, Influenza virus neuraminidase: structure, antibodies, and inhibitors, Protein Sci. 3 (1994) 1687–1696.
[7] R.C. Bethell, P.W. Smith, Sialidase as a target for inhibitors of influenza virus replication, Expert Opin. Investig. Drugs 6 (10) (1997) 1501–1509.
[8] E. Garman, G. Laver, The structure, function, and inhibition of influenza virus neuraminidase, Protein Rev. 1 (2005) 247–267.
[9] M. Ohuchi, N. Asaoka, T. Sakai, R. Ohuchi, Roles of neuraminidase in the initial stage of influenza virus infection, Microbes Infect. 8 (5) (2006) 1287–1293.
[10] P.M. Colman, J.N. Varghese, W.G. Laver, Structure of the catalytic and antigenic sites in influenza virus neuraminidase, Nature 303 (1983) 41.
[11] G.M. Air, W.G. Laver, The neuraminidase of influenza virus, Proteins 6 (1989) 341–356.
[12] R.C. Wade, 'Flu' and structure-based drug design, Structure 5 (1997) 1139–1145.
[13] M. Itzstein, W.Y. Wu, G.B. Kok, Rational design of potent salidase-based inhibitors of influenza virus replication, Nature 363 (1993) 418–423.
[14] C. Kim, W. Lew, M. Williams, H. Liu, L. Zhang, S. Swaminathan, N. Bischofberger, M. Chen, D. Mendel, C. Tai, W. Laver, R. Stevens, 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. 119 (1997) 681–690.
[15] S. Arnold, M.D. Monto, The threat of an avian influenza pandemic, New Engl. J. Med. 352 (4) (2005) 323–325.
[16] A.S. Monto, Vaccines and antiviral drugs in pandemic preparedness, Emerg. Infect. Dis. 12 (1) (2006) 55–60.
[17] Q.M. Le, M. Kiso, K. Someya, Y.T. Sakai, T.H. Nguyen, K.H.L. Nguyen, N.D. Pham, H.H. Ngyen, S. Yamada, Y. Muramoto, T. Horimoto, A. Takada, H. Goto, T. Suzuki, Y. Suzuki, Y. Kawaoka, Isolation of drug-resistant H5N1 virus, Nature 437 (2005) 1108.
[18] G. Wang, Y. Chen, S. Wang, R. Gentles, T. Sowin, W. Kati, S. Muchmore, V. Giranda, K. Stewart, H. Sham, D. Kempf, G. Laver, Design synthesis, and structural analysis of influenza neuraminidase inhibitors containing pyrrolidine cores, J. Med. Chem. 44 (2001) 1192–1201.
[19] Y. Babu, P. Chand, S. Bantia, P. Kotian, A. Dehghani, Y. El-Kattan, T. Lin, T. Hutchison, A. Elliott, C. Parker, S. Ananth, L. Horn, G. Laver, J. Montgomery, 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. 43 (2000) 3482–3486.
[20] G. Wang, S. Wang, R. Gentles, T. Sowin, C. Maring, D. Kempf, W. Kati, V. Stoll, K. Stewarta, W. Laver, Design, synthesis, and structural analysis of inhibitors of influenza neuraminidase containing a 2,3-disubstituted tetrahydrofuran-5-carboxylic acid core, Bioorg. Med. Chem. Lett. 15 (2005) 125–128.
[21] S. Singh, M. Jedrzejas, G. Air, M. Luo, W. Laver, W. Brouillette, Structure-based inhibitors of influenza virus sialidase. A benzoic acid lead with novel interaction, J. Med. Chem. 38 (1995) 3217–3225.
[22] P. Bonnet, R. Bryce, Scoring binding affinity of multiple ligands using implicit solvent and a single molecular dynamics trajectory: application to influenza neuraminidase, J. Mol. Graph. Model. 24 (2005) 147–156.
[23] K. Masukawa, P. Kollman, I. Kuntz, Investigation of neuraminidase-substrate recognition using molecular dynamics and free energy calculations, J. Med. Chem. 46 (2003) 5628–5637.
[24] I. Wall, A. Leach, D. Salt, M. Ford, J. Essex, Binding constants of neuraminidase inhibitors: an investigation of the linear interaction energy method, J. Med. Chem. 42 (1999) 5142–5152.
[25] B. Smith, P. Colman, M. Itzstein, B. Danylec, J. Varghese, Analysis of inhibitor binding in influenza virus neuraminidase, Protein Sci. 10 (2001) 689–696.
[26] M. Fornabaio, P. Cozzini, A. Mozzarelli, D. Abraham, G. Kellogg, Simple intuitive calculations of free energy of binding for protein–ligand complexes: 2. Computational titration and pH effects in molecular models of neuraminidase–inhibitor complexes, J. Med. Chem. 46 (2003) 4487–4500.
[27] N. Taylor, M. Itzstein, Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis, J. Med. Chem. 37 (1994) 616–624.
[28] J.C. Woods, M.A. King, J.W. Essex, The configurational dependence of binding free energies: a Poisson–Boltzmann study of neuraminidase inhibitors, J. Comput. Aided Mol. Des. 15 (2001) 129–144.
[29] A. Parrill, P. Ramamoorthy, J. Gervay, Molecular dynamics docking and relative binding affinity comparison of potential neuraminidase inhibitors, Book of Abstracts, vol. 215, American Chemical Society, Washington, 1998, pp. 702–712.
[30] S. Sabesan, S. Neira, Z. Wasserman, Structural and functional-group tuning in the design of neuraminidase inhibitors, Carbohydr. Res. 267 (1995) 239–261.
[31] L. Birch, C.W. Murray, M.J. Hartshorn, J. Ian, I.J. Tickle, M.L. Verdonk, Sensitivity of molecular docking to induced-fit effects in influenza virus neuraminidase, J. Comput. Aided Mol. Des. 16 (2002) 855–869.
[32] I. Muegge, The effect of small changes in protein structure on predicted binding modes of known inhibitors of influenza virus neuraminidase: PMF-scoring in DOCK4, Med. Chem. Res. 9 (7/8) (2002) 490–500.
[33] K.A. Armstrong, B. Tidor, A.C. Cheng, Optimal charges in lead progression: a structure-based neuraminidase case study, J. Med. Chem. 49 (2006) 2470–2477.
[34] T. Wang, R. Wade, Comparative binding energy (COMBINE) analysis of influenza neuraminidase–inhibitor complexes, J. Med. Chem. 44 (2001) 961–971.
[35] T. Steindl, T. Langer, 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. 44 (2003) 1849–1856.
[36] M. Itzstein, J. Dyason, S. Oliver, H. White, W. Wu, G. Kok, M. Pegg, A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs, J. Med. Chem. 39 (1996) 383–391.
[37] R. Verma, C. Hansch, A QSAR study on influenza neuraminidase inhibitors, Bioorg. Med. Chem. 14 (2006) 982–996.
[38] X. Yi, Z. Guo, F. Chu, Study on molecular mechanism and 3D-QSAR of influenza neuraminidase inhibitors, Bioorg. Med. Chem. 11 (2003) 1465–1474.
[39] J. Zhang, K. Yu, W. Zhua, H. Jiang, Neuraminidase pharmacophore model derived from diverse classes of inhibitors, Bioorg. Med. Chem. Lett. 16 (2006) 3009–3014.
[40] J. Mestres, R.M.A. Knegtel, Similarity versus docking in 3D virtual screening, Perspect. Drug Discov. 20 (2000) 191–207.
[41] M. Kontoyianni, G.S. Sokol, L.M. Mcclellan, Evaluation of library ranking efficacy in virtual screening, J. Comput. Chem. 26 (2004) 11–22.
[42] G.M. Air, A.A. Ghate, S.J. Stray, Influenza neuraminidase as target for antivirals, Adv. Virus Res. 54 (1999) 375–402.
[43] P. Ward, I. Small, J. Smith, P. Suter, R. Dutkowski, Oseltamivir (Tamiflu) and its potential for use in the event of an influenza pandemic, J. Antimicrob. Chemother. 55 (S1) (2005) i5–i21.
[44] H. Chen, P.D. Lyne, F. Giordanetto, T. Lovell, J. Li, On evaluating molecular-docking methods for pose prediction and enrichment factors, J. Chem. Inf. Model. 46 (2006) 401–415.
[45] T. Schulz-Gasch, M. Stahl, Scoring functions for protein–ligand interactions: a critical perspective, Drug Discov. Today 1 (3) (2004) 231–239.
[46] M.O. Taha, M.A. AlDamen, Effects of variable docking conditions and scoring functions on corresponding protein-aligned comparative molecular field analysis models constructed from diverse human protein tyrosine phosphatase 1B inhibitors, J. Med. Chem. 10 (2005) 8016–8034.
[47] T. Tuccinardi, V. Calderone, S. Rapposelli, A. Martinelli, Proposal of a new binding orientation for non-peptide AT1 antagonists: homology modeling. Docking and three-dimensional quantitative structure–activity relationship analysis, J. Med. Chem. 49 (2006) 4305–4316.

[48] Y.C. Martin, 3D QSAR: current state, scope, and limitations, Perspect. Drug Discov. 12/13/14 (1998) 3–23.
[49] A. Tropsha, Recent trends in quantitative structure–activity relationships, in: D.J. Abraham (Ed.), Burger's Medicinal Chemistry and Drug Discovery: Drug Discovery, John Wiley & Sons, Inc., USA, 2003, pp. 49–72.
[50] G.M. Morris, A.J. Olson, D.S. Goodsell, Protein–ligand docking methods, Princ. Med. Chem. 8 (2000) 31–48.
[51] J.R.H. Tame, Scoring functions: a view from the bench, J. Comput. Aided Mol. Des. 13 (1999) 99–108.
[52] E.M. Krovat, T. Langer, Impact of scoring functions on enrichment in docking-based virtual screening: an application study on renin inhibitors, J. Chem. Inf. Comput. Sci. 44 (2004) 1123–1129.
[53] M. Pastor, G. Cruciani, K. Watson, A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure–activity relationship analysis, J. Med. Chem. 40 (1997) 4089–4102.
[54] C.S. Poornima, P.M. Dean, Hydration in drug design: 1. Multiple hydrogen-bonding features of water molecules in mediating protein–ligand interactions, J. Comput. Aided Mol. Des. 9 (1995) 500–512.
[55] M.L. Verdonk, G. Chessari, J.C. Cole, M.J. Hartshorn, C.W. Murray, J.W.M. Nissink, R.D. Taylor, R. Taylor, Modeling water molecules in protein–ligand docking using GOLD, J. Med. Chem. 48 (2005) 6504–6515.
[56] K.F. Koehler, S.N. Rao, J.P. Snyder, in: N.C. Cohen (Ed.), Guidebook on Molecular Modeling in Drug Design, Academic Press, San Diego, CA, 1996, pp. 253–255.
[57] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites, J. Mol. Graph. Model. 21 (2003) 289–307.
[58] CERIUS2 LigandFit User Manual, Accelrys Inc., San Diego, CA, 2000, pp. 3–48.
[59] D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.B. Fogel, L.J. Fogel, S.T. Freer, Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming, Chem. Biol. 2 (1995) 317–324.
[60] A.N. Jain, Scoring non-covalent protein–ligand interactions: a continuous differentiable function tuned to compute binding affinities, J. Comput. Aided Mol. Des. 10 (1996) 427–440.
[61] A. Krammer, P.D. Kirchhoff, X. Jiang, C.M. Venkatachalam, M. Waldman, LigScore: a novel scoring function for predicting binding affinities, J. Mol. Graph. Model. 23 (2005) 395–407.
[62] I. Muegge, A knowledge-based scoring function for protein–ligand interactions: probing the reference state, Perspect. Drug Discov. 20 (2000) 99–114.
[63] X. Pan, N. Tan, G. Zeng, H. Hana, H. Huanga, 3D-QSAR and docking studies of aldehyde inhibitors of human cathepsin K, Bioorg. Med. Chem. 14 (2006) 2771–2778.
[64] M. Stahl, M. Rarey, Detailed analysis of scoring functions for virtual screening, J. Med. Chem. 44 (2001) 1035–1042.
[65] G.E. Terp, B.N. Johansen, I.T. Christensen, F.S. Jorgensen, A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein–ligand binding affinities, J. Med. Chem. 44 (2001) 2333–2343.
[66] N.R.A. Beeley, C. Sage, GPCRs: an update on structural approaches to drug discovery, Targets 2 (2003) 19–25.
[67] G. Klebe, Virtual ligand screening: strategies, perspectives and limitations, Drug Discov. Today 11 (13/14) (2006) 580–594.
[68] H. Steuber, M. Zentgraf, C. Gerlach, C.A. Sotriffer, A. Heine, G. Klebe, Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying conditions, J. Mol. Biol. 363 (1) (2006) 174–187.
[69] M.T. Stubbs, S. Reyda, F. Dullweber, M. Moller, G. Klebe, D. Dorsch, W.W.K.R. Mederski, H. Wurziger, pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design, Chembiochem 3 (2002) 246–249.
[70] M.A. DePristo, P.I.W. de Bakker, T.L. Blundell, Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography, Structure 12 (2004) 831–838.
[71] C. Kim, W. Lew, M. Williams, H. Wu, L. Zhang, X. Chen, P. Escarpe, D. Mendel, W. Laver, R. Stevens, Structure–activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors, J. Med. Chem. 41 (1998) 2451–2460.
[72] W. Lew, H. Wu, D. Mendel, P. Escarpe, X. Chen, W. Laver, B. Graves, C. Kim, A new series of C3-Aza carbocyclic influenza neuraminidase inhibitors: synthesis and inhibitory activity, Bioorg. Med. Chem. Lett. 8 (1998) 3321–3324.
[73] W. Lew, H. Wu, X. Chen, B. Graves, P. Escarpe, H. MacArthur, D. Mendel, C. Kim, Carbocyclic influenza neuraminidase inhibitors possessing a C3-cyclic amine side chain: synthesis inhibitory activity, Bioorg. Med. Chem. Lett. 10 (2000) 1257–1260.
[74] W. Kati, D. Montgomery, C. Maring, V. Stoll, V. Giranda, X. Chen, W. Laver, W. Kohlbrenner, D. Norbeck, Novel α- and β-amino acid inhibitors of influenza virus neuraminidase, Antimicrob. Agents Chemother. 45 (9) (2001) 2563–2570.
[75] P. Chand, P. Kotian, P. Morris, S. Bantia, D. Walsh, Y. Babu, Synthesis and inhibitory activity of benzoic acid and pyridine derivatives on influenza neuraminidase, Bioorg. Med. Chem. 13 (2005) 2665–2678.
[76] Y. Cheng, W. Prusoff, Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 percent inhibition (I50) of an enzymatic reaction, Biochem. Pharmacol. 22 (1973) 3099–3108.
[77] J.T. Leonard, K. Roy, On selection of training and test sets for the development of predictive QSAR models, QSAR Comb. Sci. 25 (3) (2006) 235–251.
[78] J. Gasteiger, M. Marsili, A new model for calculating atomic charges in molecules, Tetrahedron Lett. 34 (1978) 3181–3184.
[79] CERIUS2 4.10 QSAR, Accelrys Inc., San Diego, CA, 2005, pp. 177–187.
[80] CERIUS2 OFF, Accelrys Inc., San Diego, CA, 1997, pp. 5–109.
[81] CERIUS2 4.10 QSAR, Accelrys Inc., San Diego, CA, 2005, pp. 237–250.
[82] CERIUS2 4.8.1 QSAR, Accelrys Inc., San Diego, CA, 2003, pp. 210–235.
[83] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, Quant. Struct.-Act. Relat. 22 (2003) 69–77.
[84] W. Sippl, Receptor-based 3D QSAR analysis of estrogen receptor ligands-merging the accuracy of receptor-based alignments with the computational efficiency of ligand-based methods, J. Comput. Aided Mol. Des. 14 (2000) 559–572.
[85] M. Clark, R.D. Cramer, The probability of chance correlation using partial least squares (Pls), Quant. Struct.-Act. Relat. 12 (1993) 137–145.
[86] R.D. Cramer III, J.D. Bunce, D.E. Patterson, Crossvalidation, bootstrapping and PLS compared with multiple linear regression in conventional QSAR studies, Quant. Struct.-Act. Relat. 7 (1988) 18–28.
[87] M.G.B. Drew, N.R. Lumley, N.R. Price, R.W. Watkins, in: K. Gundertofte, F.S. Fergensen (Eds.), Proceedings of the 12th European Symposium on Quantitative Structure–Activity Relationships, Molecular Modeling and Prediction of Bioactivity, Kluwer Academic/Olenum Publishers, New York, 1998, pp. 453–454.
[88] R. Wang, Y. Lu, S. Wang, Comparative evaluation of 11 scoring functions for molecular docking, J. Med. Chem. 46 (2003) 2287–2303.
[89] J. Zhang, W. Xu, Recent advances in anti-influenza agents with neuraminidase as target, Mini Rev. Med. Chem. 6 (2006) 429–448.
[90] P. Bossart-Whitaker, M. Carson, Y. Babu, C. Smith, W. Laver, G. Air, Three-dimensional structure of influenza A N9 neuraminidase and its complex with the inhibitor 2-deoxy 2,3-dehydro-*N*-acetyl neumarinic acid, J. Mol. Biol. 232 (1993) 1069–1083.
[91] W.M. Kati, D. Montgomery, R. Carrick, L. Gubareva, C. Maring, K. McDaniel, K. Steffy, A. Molla, F. Hayden, D. Kempf, W. Kohlbrenner, In vitro characterization of A-315675, a highly potent inhibitor of A and B strain influenza virus neuraminidases and influenza virus replication, Antimicrob. Agents Chemother. 46 (2002) 1014–1021.
[92] G. Wang, Recent advances in the discovery and development of anti-influenza drugs, Expert Opin. Ther. Pat. 12 (6) (2002) 845–861.
[93] M.L. Verdonk, V. Berdini, M.J. Hartshorn, W.T.M. Mooij, C.W. Murray, R.D. Taylor, P. Watson, Virtual screening using protein–ligand docking: avoiding artificial enrichment, J. Chem. Inf. Comput. Sci. 44 (2004) 793–806.