Chemometrics and Intelligent Laboratory Systems 146 (2015) 485–493

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

QSAR and molecular mechanism analysis of N-substituted oseltamivir derivatives as potent avian influenza H5N1 neuraminidase inhibitors

Jiaying Sun [a,*], Hu Mei [b]

[a] Department of Chemistry and Chemical Engineering, Sichuan University of Arts and Science, Dazhou, Sichuan 635000, China
[b] College of Bioengineering, Chongqing University, Chongqing 400044, China

ARTICLE INFO

Article history:
Received 24 May 2015
Received in revised form 7 July 2015
Accepted 8 July 2015
Available online 16 July 2015

Keywords:
Avian influenza
QSAR
N-substituted oseltamivir derivatives
Neuraminidase
Inhibitors
Docking

ABSTRACT

To understand chemical–biological interactions and predict activities against avian influenza virus, four QSAR models of 32 N-substituted oseltamivir derivatives neuraminidase inhibitors (NIs) are constructed in this paper. R^2 and Q^2 of PLS, HQSAR, CoMSIA and Almond models are 0.950 and 0.846, 0.789 and 0.751, 0.991 and 0.672, 0.917 and 0.762, respectively. In addition, R^2_{test} and Q^2_{ext} of PLS, HQSAR, CoMSIA and Almond models are 0.897 and 0.787, 0.937 and 0.700, 0.848 and 0.835, 0.932 and 0.851, respectively. Therefore, QSAR models were excellent, robust and had good predictive capability. These models can be further used to evaluate and screen new compounds with a similar structure. Moreover, hydrogen bonds highly contributed to activity, followed by electrostatic and hydrophobic factors. Docking indicated that Arg118, Glu119, Asp151, Arg152, Arg156, Trp178, Glu227, Ser246, Arg292 and Tyr406 were important residues in the active pocket of 2hu4. The main influential factor of interactions between NIs and neuraminidase was the hydrogen bond, followed by electrostatic and hydrophobic factors. Therefore, docking results were unanimous in QSAR results.

1. Introduction

Influenza is a respiratory infection concern with significant morbidity in the general population and mortality in elderly and high-risk patients [1]. H5N1 is a highly pathogenic avian influenza virus, which poses a great threat to humans [2]. Neuraminidase (NA) is one of two surface glycoproteins from the influenza virus and is a potential target for the control of the virus [3]. Neuraminidase inhibitors (NIs) can selectively inhibit NA and prevent progeny virosome from replication and release into the host cell, thus protecting against influenza and alleviating symptoms effectively [4]. At present, both zanamivir and oseltamivir (Table 1) are effective inhibitors for both A and B forms of NA [5,6]. However, zanamivir administered through nebulizer or intravenous routes is highly inconvenient for patients. On the other hand, oseltamivir is absorbed following oral administration, and has been the first choice since 1999 and is widely used up to the present. In addition, peramivir (Table 1) has been authorized for emergent treatment of 2009 H1N1 influenza virus in some countries [7–10]. In 2013, the H7N9 virus, which is a lethal avian influenza virus, evolved to infect humans. However, new influenza viruses will continuously appear. As an effective type of anti-influenza drug, there is a rising trend toward resistance of NIs due to the emergence of new influenza viruses [11]. Therefore, it is necessary to research and develop new NIs against mutated influenza viruses.

In order to understand the chemical–biological interactions and predict their activities toward influenza virus, 32 N-substituted oseltamivir derivatives against influenza H5N1 virus were collected and constructed with QSAR models in this paper. The modeling methods used were PLS (partial least squares), HQSAR (hologram quantitative structure–activity relationship), CoMSIA (comparative similarity indices analysis) and Almond. In addition, Surflex-Dock was employed to investigate the interaction mechanism between NA and NIs.

2. Methods and materials

2.1. Structural characterization

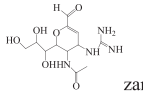
A total of 1081 descriptors including 102 constitutional descriptors, 342 topological (2D) descriptors and 637 conformational (3D) descriptors were obtained with Sarchitect (trial version). Sarchitect is a comprehensive platform for modeling and predicting drug-relevant properties of molecules. Constitutional descriptors contain 76 counts and 26 property descriptors. Topological descriptors include 64 burden eigenvalues, 96 2D autocorrelation indices, 110 topological indices, 41 topological paths/walks and 31 information content descriptors. Conformational descriptors involve 20 randic profiles, 150 radial distribution functions, 160 MoRSE descriptors, 78 WHIM descriptors, 185 getaway descriptors, 38 surface area descriptors, 5 geometrical descriptors and 1 potential energy.

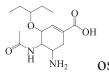
The premise of HQSAR is that the structure of a molecule is encoded within its 2D fingerprint and that structure is the key determinant of all

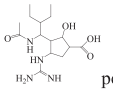
* Corresponding author.
E-mail address: wy7472@126.com (J. Sun).

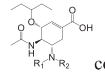
http://dx.doi.org/10.1016/j.chemolab.2015.07.005
0169-7439/

Table 1
Docking scores, experimental and predicted activity of 32 N-substituted oseltamivir derivatives


zanamivir


oseltamivir carboxylate


peramivir


common skeleton

No.	R1	R2	Exp. pIC_{50} (M)	Pred. pIC_{50} (M)				Total Scores	CScore
				PLS	HQSAR	CoMSIA	Almond		
1			6.70	6.49	6.58	6.73	6.41	5.37	2
2			5.93	6.26	6.36	6.02	6.21	5.94	2
3			5.84	5.95	6.15	5.80	6.35	6.37	4
4			6.26	6.16	6.20	6.17	6.73	6.39	5
*5			6.82	6.54	6.23	6.57	6.79	5.59	5
6			6.17	6.36	6.27	6.20	6.18	6.65	5
7			6.64	6.59	6.08	6.53	6.36	6.25	4
8			6.11	6.33	6.09	6.13	6.11	5.77	2
9			5.94	5.72	6.17	5.94	5.93	6.21	5
*10			6.43	6.00	6.03	6.59	6.30	6.13	4
11			6.02	5.83	5.86	6.03	6.23	3.29	3
12			6.10	6.26	6.20	6.08	6.18	5.20	4
13	H	H	7.77	7.61	8.37	7.76	7.76	7.30	4
14		H	8.16	8.01	7.42	8.15	7.30	7.68	3
*15		H	7.57	7.10	7.49	7.72	7.58	7.16	3
16		H	7.72	7.50	7.78	7.55	7.60	7.92	5
17		H	7.11	7.30	7.70	7.22	7.16	6.65	2
18		H	7.82	8.10	8.01	7.78	7.65	6.80	2
19		H	8.15	7.75	7.90	8.14	7.98	5.27	4
*20		H	7.51	6.86	7.76	7.58	7.69	6.06	1
21		H	7.68	7.68	7.75	7.65	7.29	7.38	3
22		H	8.44	8.41	7.96	8.37	7.56	6.83	1
23		H	7.89	7.97	7.95	7.98	7.55	7.08	2
24		H	8.00	8.27	7.95	7.79	7.57	6.20	3
*25		H	7.85	7.85	7.92	7.50	7.54	6.99	2
26		H	8.72	8.72	7.77	8.78	8.28	7.66	3
27		H	7.50	7.80	7.72	7.52	8.06	6.65	5
28		H	7.23	7.36	7.56	7.28	8.03	6.06	4
29		H	7.05	6.81	7.70	7.17	7.80	6.65	3
*30		H	7.96	8.02	7.83	7.67	8.02	6.70	4
31		H	7.37	7.38	7.62	7.55	8.15	5.94	2
32		H	8.68	8.37	7.87	8.70	8.58	7.63	2

molecular properties [12–14]. Information about fragments including atoms, bonds, connections, hydrogen atoms, chirality, hydrogen bond donor and acceptor were considered during HQSAR modeling. Fragment size of default 4–7 and 12 hologram lengths (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 and 401) were used.

GALAHAD (Genetic Algorithm with Linear Assignment of Hypermolecular Alignment of Datasets) uses Tripos' proprietary technology to generate pharmacophore hypotheses and alignments from sets of ligand molecules that bind at a common target site. Operations include two main stages: the ligands are aligned to each other in internal coordinate space, then the conformations produced are aligned in cartesian space. The first stage is fully flexible, whereas the second treats the molecules as rigid bodies. Moreover, it serves to free the user from any need to explicitly or implicitly select a template molecule [15–18]. Therefore, GALAHAD was employed to construct a pharmacophore model.

GRIND (GRid INdependent Descriptors) are based on Molecular Interaction Fields (MIFs). The procedure for obtaining GRIND involves three steps [19]: (1) calculation of MIF, (2) filtration of MIF to extract the most relevant regions that define virtual receptor site (VRS) and (3) encoding the VRS into the GRIND variables. GRIND needs no alignment of compounds. To calculate GRIND descriptors, four different GRID probes can be chosen. In this paper, DRY (hydrophobic probe), O (carbonyl oxygen probe as hydrogen bond acceptor), N1 (amidic nitrogen probe as hydrogen bond donor) and TIP (shape probe) were selected.

2.2. Modeling methods

Variables (descriptors) by Sarchitect were screened using a genetic algorithm (GA). GA was implemented with MATLAB software (version 7.0). GA variable screening parameter establishment was as follows: initial populations were 200, genetic generations were 100, crossover probability was 0.5, mutation probability was 0.01 and the evaluation function was cross-validation correlation coefficient q^2. Based on screening descriptors, the statistical model was constructed by PLS.

Based on the least standard error, the best HQSAR model was selected by PLS. In addition, based on pharmacophores by GALAHAD, the CoMSIA [20,21] model was generated by SAMPLS (sample-distance partial least squares) [22]. ALMOND program version 3.3.0 was used to generate GRIND, and QSAR model was constructed by PLS.

2.3. Model validation

PLS model was evaluated by cross-validation, *Y* random permutation test and external validation. PLS model parameter establishment was given below: confidence level was 95%, and the number of cross-validations was 7. The other QSAR models were validated by leave-one-out cross-validation (LOO CV). R^2_{test} and Q^2_{ext} of the test set were used for external prediction [23]. In addition, the models were validated by different modeling methods.

$$Q^2_{ext} = 1 - \sum_{i=1}^{test} (yi - \hat{y}i)^2 / \sum_{i=1}^{test} (yi - \bar{y}tr)^2$$

2.4. Molecular docking

Surflex-Dock is a classic protein-ligand docking program associated to a specific empirical scoring function and search engine pair [24]. Its usefulness as a drug design tool has already been demonstrated in several cases [25]. Protomol is used to guide molecular docking. In other words, Protomol is a computational representation of the intended binding site to which putative ligands are aligned. Protomol is produced in three processes [26]: (1) automatic: Surflex-Dock finds the largest cavity in the receptor protein; (2) ligand: a ligand in the same coordinate space as the receptor; and (3) residues: specified residues in the receptor. Surflex-Dock scores include hydrophobic, repulsive, entropic and solvation. The strengths of individual scoring functions combine to produce a consensus that is more robust and accurate than any single function for evaluating ligand-receptor interactions. CScore (consensus scores) integrate a number of popular scoring functions for ranking the affinity of ligands bound to the active site of a receptor [27]. CScore provides several functions: D Score [28], PMF Score [29], G Score [30] and CHEM Score [31]. CScores range from 1 to 5, and 5 is the best CScore.

2.4.1. Disposal of receptor (2hu4)

Avian influenza H5N1 virus 2hu4 [32] has eight chains. Before docking, a single chain A was kept, and the other seven chains were deleted. Moreover, all ligands and water molecules were removed and the polar hydrogen atoms were added [33,34].