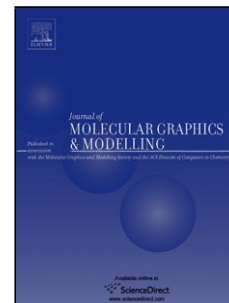


Accepted Manuscript

Title: Discovery of New Human Epidermal Growth Factor Receptor-2 (HER2) Inhibitors for Potential Use as Anticancer Agents via Ligand-Based Pharmacophore Modeling

Author: Hiba Zalloum Rabab Tayyem Basha'er Abu-Irmaileh Yasser Bustanji Malek Zihlif Mohammad Mohammad Talal Abu Rjai Mohammad S. Mubarak



PII: S1093-3263(15)30016-4
DOI: <http://dx.doi.org/doi:10.1016/j.jmgm.2015.06.008>
Reference: JMG 6562

To appear in: *Journal of Molecular Graphics and Modelling*

Received date: 8-3-2015
Revised date: 18-5-2015
Accepted date: 20-6-2015

Please cite this article as: Hiba Zalloum, Rabab Tayyem, Basha'er Abu-Irmaileh, Yasser Bustanji, Malek Zihlif, Mohammad Mohammad, Talal Abu Rjai, Mohammad S.Mubarak, Discovery of New Human Epidermal Growth Factor Receptor-2 (HER2) Inhibitors for Potential Use as Anticancer Agents via Ligand-Based Pharmacophore Modeling, *Journal of Molecular Graphics and Modelling* <http://dx.doi.org/10.1016/j.jmgm.2015.06.008>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Discovery of New Human Epidermal Growth Factor Receptor-2 (HER2) Inhibitors for Potential Use as Anticancer Agents via Ligand-Based Pharmacophore Modeling

Hiba Zalloum^{a*}, Rabab Tayyem^b, Basha'er Abu- Irmaileh^a, Yasser Bustanji^c, Malek Zihlif^d, Mohammad Mohammad^b, Talal Abu Rjai^e, Mohammad S. Mubarak^{f*}

^a Hamdi Mango Research Center for Scientific Research, The University of Jordan, Amman 11942, Jordan

^b ACDIMA BioCenter, Amman 11190, Jordan

^c Department of Biopharmaceutics and clinical pharmacy, Faculty of Pharmacy, University of Jordan,

^d Department of Pharmacology, Faculty of Medicine, University of Jordan,

^e Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan

^f Department of Chemistry, Faculty of Science, The University of Jordan, Amman 11942, Jordan
Amman 11942, Jordan.

Abstract

To discover potential antitumor agents directed toward human epidermal growth factor receptor-2/HER2/ErbB2 overexpression in cancer, we have explored the pharmacophoric space of 115 HER2/ErbB2 inhibitors. This identified 240 pharmacophores which were subsequently clustered into 20 groups and cluster centers were used as 3D-pharmacophoric descriptors in QSAR analysis with 2D-physicochemical descriptors to select the optimal combination. We were obliged to use ligand efficiency as the response variable because the logarithmic transformation of bioactivities failed to access self-consistent QSAR models. Two binding pharmacophore models emerged in the optimal QSAR equation, suggesting the existence of distinct binding modes accessible to ligands within the HER2/ErbB2 binding pocket. The QSAR equation and its associated pharmacophore models were employed to screen the National Cancer Institute (NCI) and Drug Bank databases to search for new, promising, and structurally diverse HER2 inhibitory leads. Inhibitory activities were tested against HER2-overexpressing SKOV3 Ovarian cancer cell line and MCF-7 which express low levels of HER2. *In silico* mining identified 80 inhibitors out of which four HER2 selective compounds inhibited the growth of SKOV3 cells with IC₅₀ values < 5 μM and with virtually no effect in MCF-7 cells. These lead compounds are excellent candidates for further optimization.

Keywords: Human epidermal growth factor receptor-2 (HER2); pharmacophore modeling; Quantitative structure–activity relationship (QSAR); *in silico* screening.

*Corresponding authors. Tel. (Mobile): +962 795902908; Fax: +962 6 5300238(H. Zalloum);
Tel. (Mobile): +962 799950383; Fax: +962 6 5300253
(M. S. Mubarak).

E-mail addresses: hibazalloum@gmail.com (H. M.Zalloum); mmubarak@ju.edu.ju (M. S. Mubarak).

Graphical abstract

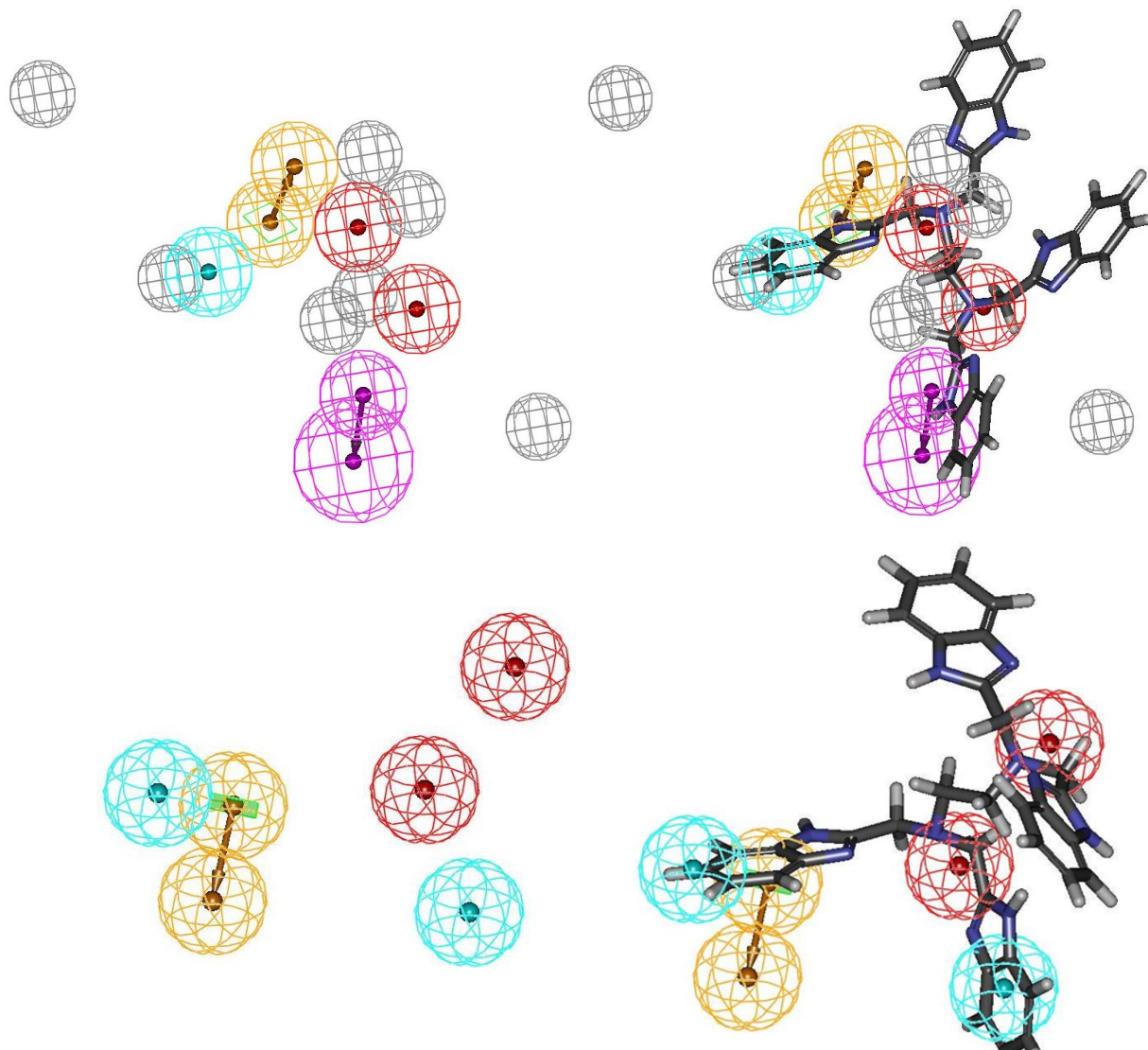
fx1

Pharmacophoric features of the binding models Hypo 4/15 and Hypo9/12 (HBD as violet vectored spheres and Hbic as blue spheres, PosIon as red spheres, RingArom as orange vectored spheres, and exclusion volumes as grey spheres), fitted against the most active hit 120 ($IC_{50} = 1.43 \mu M$, Table 4).

Highlights

- Pharmacophoric space of 115 (HER2/ErbB2) inhibitors was explored.
- Two pharmacophore models, Hypo4/15 and Hypo9/12, illustrated the best performance.
- *in silico* screening of NCI and DrugBank databases identified 80 potential HER2 inhibitors.
- 4 of investigated hits selectively inhibited the growth of SKOV3 ovarian cells with $IC_{50} < 5 \mu M$.

Graphical Abstract



Pharmacophoric features of the binding models **Hypo 4/15** and **Hypo9/12** (HBD as violet vectored spheres and Hbic as blue spheres, PosIon as red spheres, RingArom as orange vectored spheres, and exclusion volumes as grey spheres), fitted against the most active hit 120 ($IC_{50} = 1.43\mu M$, Table 4)

Highlights

- Pharmacophoric space of 115 (HER2/ErbB2) inhibitors was explored.
- Two pharmacophore models, Hypo4/15 and Hypo9/12, illustrated the best performance.
- *in silico* screening of NCI and DrugBank databases identified 80 **potential** HER2 inhibitors.
- 4 of investigated hits **selectively** inhibited the growth of SKOV3 ovarian cells with **$IC_{50} < 5\mu M$** .

1. Introduction

Many human cancers are characterized by elevated levels of proteins that regulate cell cycle progression and proliferation. The HER tyrosine kinase receptor family comprises four homologous epidermal growth factor (EGF) receptors: EGFR/ErbB1/HER1, HER2/*neu*/ErbB2, HER3/ErbB3, and HER4/ErbB4 [1]. In addition, the HER family of receptors are among the most studied cell signaling families in cancer biology [2]. Deregulation of growth-factor signaling due to hyperactivation of the ErbB receptors is seen in a wide variety of cancers (breast, bladder, prostate, pancreatic, colon, ovary and non-small cell lung cancers). [3,4,5] Importantly, overexpression of HER receptors is associated with poor disease prognosis and reduced survival. The human epidermal growth factor receptor-2 (HER2/*neu*) protooncogene encodes a 185 kDa glycoprotein that plays a key role in breast cancer. Approximately, 30% of breast cancers, as well as ovarian tumors, have an amplification of the *HER2/neu* gene or overexpression of its protein product and this is associated with increased disease metastasis, recurrence, aggressive tumors, and poor prognosis. [6,7,8,9,10,11] HER-2 Overexpression occurs primarily as a result of gene amplification, it occurs in other cancers such as ovarian cancer [12], stomach cancer, lung cancer (especially lung adenocarcinomas) [13,14,15,16] prostate cancer [16,17,18] and biologically aggressive forms of uterine cancer, such as uterine serous endometrial carcinoma.[19] However, in lung and prostate cancer, HER-2 expression levels are generally lower than those seen in breast and ovary cancer, gene amplification is rare, and the prognostic significance of overexpression is uncertain. [13,14,17,20] Survival rates and tumor aggressiveness can be directly correlated to the level of HER2 expression [21]. In this regard, regulating the cellular signal transduction *via* inhibition of HER2 has been considered a promising way of controlling malignant tumors. HER2 is known to form clusters which might play a role in tumorigenesis. [22,23]

Drugs targeting ErbB/HER family fall into three main categories depending on the receptor region targeted: extracellular, intracellular, and nuclear. A humanized monoclonal antibody, trastuzumab (Herceptin, Genentech), directed against the extracellular domain of HER2 has proven to be an effective therapy for cancer patients who overexpress HER2 [21,24] Lapatinib (Tykerb, GlaxoSmithKline) is a tyrosine kinase inhibitor that targets the HER2 tyrosine kinase domain [25]; it is a second-line treatment for breast cancer patients who

are refractory to trastuzumab and chemotherapy [26]. Trastuzumab and lapatinib have been studied in cell and animal models of ovarian cancer, demonstrating the potential use of these therapies to treat patients. In SKOV3 cells, a HER2 over-expressing cell line derived from the ascites of an ovarian adenocarcinoma, trastuzumab has been shown to reduce the activation of HER2 signaling pathways. [27] Moreover, both trastuzumab and lapatinib have been shown to reduce the ability of SKOV3 cells to form spheres in a dose-dependent manner, suggesting that both agents have an effect on the growth or viability of ovarian cancer cells. [28,29] Several related protein ligands interact with the ErbB/HER receptors with varying affinity and overlapping specificity [30,31,32]. However, HER-2 does not have a corresponding ligand; instead, it is the preferred heterodimerization partner for all other family members and functions as a co-receptor [33]. The absence of a HER2 ligand is consistent with structural studies, showing a fixed, open configuration for HER2 that resembles the ligand-activated state, suggesting that the receptor is poised to interact with other family members in the absence of direct ligand binding [34].

The identification of molecular cancer drug targets and the development of specific agents directed to these targets have raised expectations of disease outcomes in cancer patients. HER2 has generated an enormous amount of interest in the development of drugs that may have therapeutic potential for the treatment of this disease. This continued interest in the development of new selective HER2 ligands, combined with the lack of adequate ligand-based computer-aided modeling and drug discovery efforts in this area, prompted us to explore the possibility of developing ligand-based three-dimensional (3D) pharmacophore(s) integrated within selfconsistent QSAR model(s) [35]. The pharmacophore model(s) can be used as 3D search query (ies) to mine 3D libraries for new HER2 ligands, while the QSAR model helps to predict the biological activities of the captured compounds. We have recently, employed two pharmacophore models (Hypo1/7 and Hypo2-1/2-6) to screen a National Cancer Institute (NCI) list of compounds to search for promising fructose-1,6-bisphosphatase (FBPase) inhibitory leads. *In silico* mining identified 18 FBPase inhibitors out of which six demonstrated submicromolar IC₅₀ values. [36] Recently, *in silico* design targeting HER2 revealed some oligopeptides and peptidomimetic for treatment of breast cancer.[37,38] In this study, we present the pharmacophore and QSAR modeling, followed by *in silico* screening to

discover a set of novel HER2 inhibitory leads with activity in HER2-overexpressing ovarian SKOV-3 cells.

2. Methods and Materials

2.1. Materials and equipment

All chemicals used were obtained from commercial sources and were used as received without further purification. NCI samples were kindly provided by the National Cancer Institute and were of purity >95%. Although most of the NCI samples have CAS numbers, we decided to do characterization of some of them using ^1H and ^{13}C NMR spectroscopy and high resolution mass spectrometry. Details of the spectral data of these compounds are given in supplementary material. ^1H and ^{13}C NMR spectra were recorded with a Bruker-Avance III 500 MHz spectrometer with DMSO-*d*₆ as solvents and TMS as an internal standard. Chemical shifts are expressed in δ units; coupling constants (*J*-values) for ^1H - ^1H , are given in Hertz. High resolution mass spectra (HRMS) were acquired by electrospray ionization (ESI) technique with the aid of Bruker APEX-2 instrument. The samples were dissolved in acetonitrile, diluted in spray solution (methanol / water 1:1 v/v + 0.1 % formic acid) and infused using a syringe pump with a flow rate of 2 μL / min. External calibration was conducted using arginine cluster in a mass range *m/z* 175-871. These data, detailed in supplementary material, are consistent with the suggested structures. The mass spectra display the correct molecular ion peaks for which the measured high resolution (HRMS) data are in good agreement with the calculated values. DEPT experiments were employed to differentiate secondary and quaternary carbons from primary and tertiary ones.

2.2. Molecular Modeling

2.2.1 Software and Hardware

We have utilized the following software packages in the present research. The software packages: Discovery Studio 2.5.5, Accelrys Inc., USA (www.accelrys.com) and CS ChemDraw Ultra 6.0, Cambridge Soft Corp., USA (<http://www.cambridgesoft.com>), were employed in this investigation. Pharmacophore and QSAR modeling studies were conducted

with the aid of Discovery Studio from Accelrys Inc. (San Diego, California, USA, www.accelrys.com) installed on a hp Z420 PC workstation equipped with a 6 Core Inter Xeon Processor (E5-1650), 12 M Cache and 3.2 GHz and 32.0 GB RAM running Windows 7 operating system. Structure drawing was performed using ChemDraw Ultra 6.0 installed on the same machine.

2.2.2 Data Set of Human Epidermal growth factor Receptor2 (HER2, ErbB2) Inhibitors

The structures of 115 HER2 inhibitors (**1-115**, Table 1 and **Figure 1**) were collected from recently published literature.[39,40,41,42,43] Although the *in vitro* bioactivities of the collected inhibitors were gathered from five separate literature procedures, they were, however, carried out by employing the same bioassay methodologies. The bioactivities were expressed as the concentrations of the test compounds that inhibited the activity of HER2 by 50% (IC_{50}). The logarithm of measured IC_{50} (nM) values were used in the three-dimensional quantitative structure activity analysis (3D-QSAR), **thus correlating the data in a linear fashion to the free energy change**. In cases where the IC_{50} values of some compounds were expressed as being higher than 1000 nM, we assumed that their IC_{50} values are assumed to be equal to 50,000 nM. This assumption is necessary to allow 3.5 log **units** separation between the most potent training compounds (i.e., **2, 3, 8, 10, 11, 12, 19, 23, 24, 103, 110, 111 and 113**) and the least active training compounds (i.e., **29, 30, 100 and 105**). This assumption was applied for pharmacophore modeling purposes, since poorly active compounds are utilized by the HYPOGEN algorithm (implemented in Discovery Studio 3D QSAR Pharmacophore Generation) in the subtractive phase to exclude non-discriminating pharmacophores (i.e., those that lack the ability to differentiate active training compounds from inactive ones). This is necessary to allow statistical correlation and QSAR analysis. The logarithmic transformation of IC_{50} values should minimize any potential errors resulting from this assumption.

The two-dimensional (2D) chemical structures of the inhibitors were drawn using ChemDraw Ultra, converted into the corresponding standard 3D structures by Discovery Studio.

2.2.3 Conformational Analysis

The conformational space of each inhibitor (**1-115**, Figure 1 and Table 1) was explored by adopting the “best conformer generation” option within pharmacophore generation protocol in the discovery studio [44]. Default parameters were employed in the conformation generation procedure of training compounds and screened libraries (NCI and Drug Bank), i.e., a conformational ensemble was generated with an energy threshold of 20 kcal/mol from the structure which has the lowest energy and a maximum of 250 conformers per molecule [44,45,46,47].

2.2.4 Pharmacophoric Hypotheses Generation

All of 115 compounds included along with their associated conformational models were regrouped into a spreadsheet. Data pertaining to the biological activity of the HER2 inhibitors were reported with an “Uncertainty” value of 3; 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.^[46,47] Therefore, three structurally diverse training subsets: sets **I**, **II** and **III** in (Table A in supplementary material), respectively, were carefully selected from the collection for pharmacophore modeling. Typically, HYPOGEN pharmacophore generation requires informative training sets that include at least 16 compounds of evenly spread bioactivities over at least three and a half logarithmic cycles; smaller training sets could lead to chance correlation and thus faulty models. The selected training sets were then utilized to conduct 24 modeling runs to explore the pharmacophoric space of HER2 inhibitors (Table A in supplementary material). The exploration process included altering interfeature spacing parameter (100 and 300 picometers) and the maximum number of allowed features in the resulting pharmacophore hypotheses, i.e., they were allowed to vary from 4 to 5 for the first and second runs and from 5 to 5 for the third and fourth runs of each training set, as shown in (Table B in supplementary material). The other four runs for each set were similar to previous runs with 10 excluded volumes added. Pharmacophore modeling employing HYPOGEN proceeds through three successive phases: The constructive phase, subtractive phase, and optimization phase, in addition to the refinement of produced pharmacophores with excluded volumes, using data from inactive ligands when these excluded volumes added to the run.[44,45,46,47,48,49] Eventually, our

pharmacophore exploration efforts (24 automatic runs, Table 1) culminated in 240 pharmacophore models of variable qualities.

2.2.5 Assessment of the *Hypotheses Generated*

When generating hypotheses, the software attempts to minimize a cost function consisting of three terms: Weight cost, Error cost and Configuration cost.[44,45,46,47] An additional approach to assess the significance of the pharmacophores is to cross-validate them. This validation procedure is based on a statistical test called Fischer's randomization test. [44,45,46,47,50,51] In this validation test, we selected a 99% confidence level. Based on Fischer randomization criteria, all of our pharmacophores passed the 99% significance threshold and **were used** for subsequent processing (clustering).

2.2.6 Clustering of the Generated Pharmacophore Hypotheses

The successful models (240) were clustered into 20 groups using the hierarchical average linkage method available in the software, it compares all the input pharmacophores and creates a distance matrix, which is clustered and presented as a dendrogram. Based on their significance F-values(**Calculated through LINEST function in Excel**), the highest-ranking candidates were selected to represent their corresponding clusters in subsequent modeling. Table 2 presents information about representative pharmacophores.

2.2.7 QSAR Modeling

A subset of 91 compounds from the total list of inhibitors (1–115) was utilized as a training set for QSAR modeling. However, since it is essential to access the predictive power of the resulting QSAR models on an external set of inhibitors, the remaining 24 molecules (ca. 20% of the dataset) were employed as an external test subset for validating the QSAR models. The test molecules were selected as follows: the collected inhibitors (1–115, Table1, Figure 1) were ranked according to their IC_{50} values, and then every fifth compound was selected for the test set 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. **The selected test inhibitors are marked with double asterisks in Table1.**

The chemical structures of the inhibitors were imported into Discovery Studio as standard 3D single conformer representations in SD format. Subsequently, different descriptor groups were calculated for each compound. The calculated descriptors included various simple and valence connectivity indices, electro-topological state indices and other molecular descriptors. Furthermore, the training compounds were fitted (using the Best-fit option in ligand pharmacophore mapping implemented in Discovery Studio Software) against the representative pharmacophores (20 models, Table B in supplementary material), and their fit values were added as additional descriptors.

Genetic function approximation (GFA) was employed to search for the best possible QSAR regression equation capable of correlating the variations in biological activities of the training compounds with variations in the generated descriptors, that is, multiple linear regression modeling (MLR). In this work, we have employed the fitness function that is based on Friedman's 'lack-of-fit' (LOF). [52] We were obliged to normalize the potencies of the training compounds *via* division by their corresponding molecular weights i.e., ligand efficiency ($\log(1/IC_{50}) / Mwt$) [53,54], to achieve reasonable self-consistent QSAR models. The Calculate Ligand Efficiency protocol in Discovery Studio normalizes a selected set of numeric molecular properties based on ligand size. [55,56]

Our preliminary diagnostic trials suggested the following optimal GFA parameters: explore linear and quadratic equations at mating and mutation probabilities of 50%; population size = 500; number of genetic iterations = 50,000 and a scoring method set to Friedman lack-of-fit (LOF). However, to determine the optimal number of explanatory terms (QSAR descriptors), we decided to scan and evaluate all possible QSAR models resulting from 3 to 15 explanatory terms. All QSAR models were validated by employing leave one-out crossvalidation (r^2_{LOO}), (r^2_{adj}) which is adjusted for the number of terms in the model and predictive r^2 (r^2_{PRESS}) calculated from the test subsets. The predictive r^2 PRESS is defined as:

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

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, PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set.

2.2.8 *In Silico Screening for New HER2 Inhibitors*

Hypo4/15 and Hypo 9/12 were employed as 3D search queries to screen two 3D flexible structural databases, NCI and DrugBank. The screening was performed by employing the "Fast Flexible Database Search" option implemented within Discovery Studio. NCI hits were filtered according to Lipinski's and Veber's [57,58] rules. The remaining hits were fitted against the pharmacophores using the "best fit" option within Ligand Pharmacophore Mapping protocol.

2.3 *In-vitro* and Cell Culture Studies

2.3.1 *Materials and Methods*

HER2 overexpressing human ovarian cancer cell line SKOV-3 and MCF-7 breast cancer cell line were obtained from American Type Culture Collection ATCC (Manassas, VA) and were used for cytotoxicity assays. Dulbecco's phosphate buffer saline PBS (1X) without Ca/Mg, an MTT assay kit, McCoy's 5 α medium, L- Glutamine, Trypsin (1X), penicillin/streptomycin, and Gentamycin were obtained from PAA (Austria).

2.3.2 *Preparation of hit compounds for assay*

The tested compounds were provided as dry powders in variable quantities. They were initially dissolved in DMSO cell culture grade to give stock solutions. Subsequently, they were diluted to the required concentrations with the proper medium for proliferation assays.

2.3.3 *Cell culture*

SKOV3 human ovarian carcinoma cell lines were grown in McCoy's 5 α medium while MCF-7 cells were grown in RPMI-1640 medium. Media were supplemented with 10% heat-inactivated fetal bovine serum, 10 mM L-glutamine, 100 units /mL penicillin, and 100 mg/mL streptomycin. All cells were incubated at 37 °C in a humidified 5% CO₂ atmosphere. Media was changed every 48 to 72 h, and cells were about 80% confluent at the time of passaging.

2.3.4 *Cell Proliferation Assay*

Cells were seeded in 96-well plates at a density of 1×10^5 cells per well in medium and allowed to attach overnight, then drugs were screened for cytotoxicity at 20 μM final concentration. For the IC_{50} determination, cells were treated with increasing concentrations of the tested drugs ranging from (0.098-100 μM). All drugs were dissolved in DMSO before the addition to cell cultures and equal amounts of the solvent were added to control cells and were incubated for 72 hrs, the final concentration of DMSO did not exceed 1%. After 72 hrs, cell viability was assessed; 10 μL of tetrazolium dye 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) was added to each well. After the cells were incubated for 2 hrs in the presence of MTT, the medium was removed, and 100 μL of solubilization stop solution mix was added to each well to solubilize the dark violet formazan crystals. The optical densities at 579 nm were then measured using a microplate reader. IC_{50} concentrations were obtained from the dose-response curves using Graph Pad Prism Software 5 (San Diego, CA, USA, www.graphpad.com), and trastuzumab (Herceptin) as positive control. Figure A, under Supplementary data, illustrates the dose/response plots of these hits against sigmoidal inhibition mode.

3. Results and Discussion

3.1 Data Mining and Conformational Coverage

Discovery Studio-HYPOGEN allows automatic pharmacophore construction by using a collection of molecules with activities ranging over a number of orders of magnitude. Pharmacophores (hypotheses) explain the variability of bioactivity with respect to the geometric localization of the chemical features present in the molecules used to build it. The pharmacophore model consists of a collection of features necessary for the biological activity of the ligands arranged in 3D space (e.g., hydrogen bond acceptors and donors, hydrophobic regions, etc...). Different hypotheses were generated for a series of HER2 inhibitors. The literature was surveyed to collect as many reported structurally diverse HER2 inhibitors as possible, A total of 115 compounds were collected and were used in this investigation (Table 1, Figure 1) [39,40,41,42,43] and conformational coverage was modelled by employing the “Best” module to ensure extensive sampling of conformational space. Efficient conformational coverage guarantees minimal conformation-related noise during pharmacophore generation and validation stages.[59] Three training subsets were selected

from the collection (Table A in supplementary material) and each subset consisted of inhibitors with structural diversity that appear to follow certain 3D structure–activity relationship (SAR) rules. The biological activity in the training subsets spanned from 3.5 to 4 orders of magnitude. Genetic algorithm and multiple linear regression statistical analyses were subsequently employed to select an optimal combination of complementary pharmacophores capable of explaining bioactivity variations among all inhibitors.

3.2 Exploration of HER2 Pharmacophoric Space

HYPOGEN, as implemented in the 3D QSAR Pharmacophore Generation protocol within Discovery Studio, was employed to identify possible pharmacophoric binding modes assumed by HER2 inhibitors.[39,40,41,42,43,45] HYPOGEN implements an optimization algorithm that evaluates a large number of potential models for a particular target through fine perturbations to hypotheses that survived the constructive and subtractive phases of the modeling algorithm. [45]

3.3 Pharmacophore Hypothesis Generation

The extent of the evaluated pharmacophore space is reflected by the configuration (Config.) cost calculated for each modeling run. It is generally recommended that the Config. cost not exceed 17 (corresponding to 2^{17} hypotheses to be assessed by the software) to ensure thorough analysis of all models.[44,45] The size of the investigated pharmacophore space is a function of training compounds, selected chemical features and other software control parameters.[46] Restricting the extent of explored pharmacophore space should improve the efficiency of optimization by allowing effective evaluation of a limited number of pharmacophore models. Moreover, the fact that pharmacophore modeling requires a limited number of carefully selected training compounds (from 16 to 45 compounds only) that exhibit bioactivity variations attributable solely to the presence or absence of pharmacophore features (3D SAR), makes it impossible to explore the pharmacophore space of large training sets (such as all the 115 compounds) in one run. This is partly because HYPOGEN is not suited to handle large number of compounds and partly because pharmacophore modeling is generally confused by electronic and steric bioactivity modifying factors commonly encountered in SAR data. Nevertheless, the basic problem with this approach is to identify a particular training set capable of representing the whole list of collected compounds.

This problem can be very significant in cases of large SAR lists, as in our case. We found that the best way to deal with this problem is by exploring the pharmacophore space of several carefully selected training subsets from the whole list of collected compounds, followed by allowing the resulting pharmacophores to compete within the context of GFA-QSAR analysis such that the best pharmacophores that are capable of explaining bioactivity variations across the whole list of collected compounds are selected. However, since pharmacophore models fail in explaining the electronic bioactivity-modulating effects, the GFA-QSAR process should be allowed to select other 2D physicochemical descriptors to complement the selected pharmacophore(s). Therefore, we decided to explore the pharmacophore space of HER2 inhibitors under reasonably imposed "boundaries" through 24 HYPOGEN automatic runs and to employ three carefully selected training subsets (subsets I-III in Table A in supplementary material). The training compounds in these subsets were selected in such away as to guarantee maximal 3D structural diversity and continuous bioactivity spread over 3.5 logarithmic units.[44]

Guided by our rationally restricted pharmacophore exploration concept, we restricted the software to explore HYPOGEN pharmacophore models incorporating from one to three aromatic rings (RingArom), zero to two of hydrogen bond acceptors (HBA), hydrogen bond donor (HBD), hydrophobic (HBic), and positive ionizable (PosIon) features instead of the default range of zero to five, as shown in Table B (in supplementary material). Furthermore, the software was instructed to explore only 4- and 5-featured pharmacophores, i.e., ignore models with fewer features in order to further narrow the investigated pharmacophoric space and to represent the feature-rich nature of known HER2 ligands as shown in Table B in supplementary material. In each run, the resulting hypotheses were automatically ranked according to their corresponding "total cost" value, which is defined as the sum of error cost, weight cost and configuration cost.[44,60] Error cost provides the highest contribution to total cost and it is directly related to the capacity of the particular pharmacophore as a 3D-QSAR model, which correlates the molecular structures to the corresponding biological responses.[44,60] Moreover, HYPOGEN can also calculate 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 (residual cost, Table 2) the more likely that the hypothesis does not reflect a chance correlation.

An additional validation technique to assess the quality of the HYPOGEN pharmacophores is to cross-validate them based on Fischer's randomization test.[50] Eventually, 240 pharmacophore models emerged from 24 automatic HYPOGEN runs, all of these models illustrated scrambling confidence levels $\geq 95\%$ (Fisher scrambling criteria). These successful models were clustered and their best representatives (20 models, Table 2) were subsequently evaluated by Quantitative Structure Activity Relationship (QSAR) analysis, and by correlation between fit values and bioactivities (F-values) (Calculated through LINEST function in Excel), among collected compounds (1-115, Table 1), as displayed in Table 2. As revealed in the Table, representative models shared comparable features and acceptable statistical success criteria. In other words, emergence of several statistically comparable pharmacophore models suggests the ability of HER2 ligands to assume multiple binding modes within the binding pocket. Therefore, it is quite challenging to select any particular pharmacophore hypothesis as a sole representative of the binding process.

3.4. QSAR modeling

Although pharmacophore hypotheses provide excellent insights into ligand-macromolecule recognition and can be employed to mine for new biologically interesting scaffolds, their predictive value as 3D-QSAR models is limited by bioactivity-enhancing or reducing auxiliary groups[35]. This point, combined with the fact that pharmacophore modeling of HER2 inhibitors furnished numerous binding hypotheses of comparable success criteria (Table 2), prompted us to employ classical QSAR analysis to search for the best combination of pharmacophore(s) and 2D descriptors capable of explaining bioactivity variation across the whole list of collected inhibitors (1–115, Table 1 and Figure 1). Furthermore, QSAR modeling was implemented in the current case in order to select the best pharmacophore(s) that can explain bioactivity variation across the whole training list. However, since pharmacophore models fail in explaining electronic bioactivity-modulating effects, the GFA-QSAR process should be allowed to select other 2D physicochemical descriptors to complement the selected pharmacophore(s). Extensive QSAR modeling to

search for the best combination of pharmacophore(s) (i.e., among the best 20 models) and other 2D (physicochemical descriptors) capable of explaining bioactivity variation across the whole list of collected inhibitors was accomplished by employing genetic function approximation and multiple linear regression QSAR (GFA-MLR-QSAR) analysis to search for optimal QSAR model.[61]

The fit values obtained by mapping 20 representative hypotheses against collected inhibitors (1–115, Figure 1 and Table 1) were enrolled together with a selection of 2D descriptors as independent variables in GFA-MLR-QSAR analysis. Unfortunately, all our attempts to achieve self-consistent and predictive QSAR models were ineffective; this prompted us to evaluate an alternative modeling strategy, namely, to employ ligand efficiency [$\log(\text{IC}_{50})/\text{MW}$] [54] as the response variable instead of activity ($\log(\text{IC}_{50})$). We felt that normalizing the bioactivities, based on molecular weights, can achieve significant QSAR correlations because most reported (in particular potent) HER2 inhibitors are heavily functionalized and have large molecular weights, which seem to create structure-bioactivity noise that confuses attempts to correlate structural descriptors with bioactivity. The use of ligand efficiency values should neutralize molecular weight-related noise. Fortunately, this strategy proved successful in this case.

To assess the predictive power of the resulting QSAR models on an external set of inhibitors, we randomly selected 24 molecules (marked with double asterisks in Table 1) and employed them as an external testing set for validating QSAR models. Various statistical measures can be adopted to measure the fitness of a GFA model during the evolution process. All QSAR models were cross-validated automatically using the leave-one-out cross-validation implemented in the software.[60,62] Shown in Equation (3) and Figure 2 are the optimal QSAR model and the corresponding scatter plots of experimental versus estimated bioactivities for the training and testing inhibitors, respectively.

$$\frac{\log_{10}(1/\text{IC}_{50})}{\text{MW}} = -0.196 - 6.42 * 10^{-4}(\text{Kappa}_{1\text{AM}}) + 2.67 * 10^{-6}(\text{Wiener}) - 4.95 * 10^{-4}(\text{JurS}_{\text{DPSA}_1}) + 1.29 * 10^{-3}(\text{JurS}_{\text{PPSA}_1}) - 1.78 * 10^{-4}(\text{JurS}_{\text{WNSA}_1}) - 9.40 * 10^{-4}(\text{JurS}_{\text{WPSA}_1}) + 2.92 * 10^{-4}(\text{Hypo4/15}) + 2.16 * 10^{-4}(\text{Hypo 9/12}) + 4.10 * 10^{-7}(\text{JurS}_{\text{WPSA}_1})^2$$

$$n = 91, r^2_{91} = 0.79, r^2(\text{adj}) = 0.77, r^2(\text{LOO}) = 0.73, \text{Friedman L.O.F.} = 8.96 * 10^{-6}, r^2(\text{PRESS-24}) = 0.67 \quad \text{Eq. (3)}$$

Where n is the number of training compounds, r^2 is the correlation coefficient against 91 training compounds; $r^2(\text{adj})$ is r^2 adjusted for the number of terms in the model; $r^2(\text{LOO})$ is the leave-one-out cross-validation correlation coefficient; Friedman L.O.F. is the Friedman lack-of-fit score [52], and $r^2(\text{PRESS-24})$ is the predictive r^2 determined for 24 randomly selected test compounds. Hypo4/15 and Hypo9/12 represent the fit values of the training compounds against the 4th and 9th pharmacophore models generated in the 15th and 12th automatic pharmacophoric runs as shown in Table 2 (and Table C in supplementary material). Interestingly, the combination of Hypo4/15 and Hypo9/12 featuring in the highest-ranking QSAR equations suggests that they represent two complementary binding modes accessible to ligands within the binding pocket of HER2, i.e., one of them can optimally explain the bioactivities of some inhibitors, whereas other inhibitors are more appropriately explained by the second pharmacophore.

The Wiener index is the sum of the chemical bonds existing between all pairs of heavy atoms in the molecule. [63,64,65] κ^a_1 ($\text{Kappa}_{1\text{AM}}$) is the α -modified shape index of order one. This index is the refinement of the shape index κ_1 that takes into consideration the contribution from covalent radii and hybridization states to the shape of the molecule. κ_1 encodes the shapes of molecules in terms of the count of atoms and the presence of cycles relative to the minimal and maximal graphs. [66,67] JursPPSA1 is the partial positive surface area (sum of the solvent-accessible surface areas of all positively charged atoms), JursDPSA1 is the difference in charged partial surface areas (partial positive solvent accessible surface area minus partial negative solvent-accessible surface area), while JursWPSA1 and JursWNSA1 are both surface-weighted charged partial surface area descriptors. [60,62] Both the Wiener and κ^a_1 descriptors are topological 2D descriptors that help to differentiate the molecules according, mostly, to their size, degree of branching, flexibility, and overall shape. [68,69,70] The emergence of these two descriptors illustrates the role played by ligand topology in the binding process; also as both of these descriptors are related to the size and branching of the molecule, this can explain the importance of using the ligand efficiency instead of activity in our case to obtain successful QSAR models. However, despite the predictive significance of these descriptors, their information content is quite obscure. Jurs

descriptors are a set of spatial descriptors that combine shape and electronic information to characterize molecules based on partial charges mapped on the surface [71]. The descriptors are calculated by mapping atomic partial charges on solvent-accessible surface areas of individual atoms. On the other hand, the combined emergence of multiple hydrophilicity indicators, i.e. the Jurs partial charge surface area descriptors, JursPPSA1, JursDPSA1, JursWPSA1 and JursWNSA1 all in association with negative and positive regression coefficients suggests that our binding pocket is highly charge dependant, i.e. optimal ligand/HER2 affinity requires certain optimal hydrophilic/hydrophobic balance. JursPPSA1 and JursDPSA1 suggest that positively charged molecules favor binding to the HER2 binding site, as implied also from the emergence of Hypo4/15 and Hypo9/12 (both include a PosIon feature) in our best model shown in Eq. 3. However, obtaining the surface-weighted partial charge surface areas descriptors, JursWPSA1 and JursWNSA1 may be suggestive that at least one of the HER2 binding modes favors binding of more polar compounds.

The statistical significance of QSAR equation (3) is supported by r^2 , $r^2(\text{LOO})$, $r^2(\text{adj})$ and $r^2(\text{PRESS-24})$ against 24 randomly selected external compounds, where ($r^2_{91} = 0.79$, $r^2(\text{adj}) = 0.77$, $r^2(\text{LOO}) = 0.73$, $r^2(\text{PRESS-24}) = 0.67$).

3.5 In-Silico Screening and Subsequent Invitro Testing

Pharmacophore models are suitable for screening 3D molecular databases because they can capture compounds that exhibit optimally oriented binding features, complementary to a proposed binding site. [72,73,74,75]Hypo4/15 and Hypo9/12 were employed as 3D search queries against the NCIdatabase (2012 release of a multiconformer database includes 254,805 compounds stored at the National Cancer Institute) as well as DrugBank database [76] (bioinformatics and chemoinformatics resource that contains 6825 drugs) using the "Best Flexible Database Search" option implemented within Discovery Studio. Compounds that have their chemical groups, spatially overlapped (map) with corresponding features of the particular pharmacophoric model were captured as hits. Table 3 summarizes the numbers of captured hits by both pharmacophores. The NCI hits were then filtered based on Lipinski's and Veber's rules [57,58] of drug-likeness to help in finding hits more amenable for subsequent optimization into leads. However, DrugBank hits were left without subsequent filtration.

We believe that loosening Lipinski's and Veber's [57,58] prefilters (i.e., allowing hits of molecular weights > 500 D and rotatable bonds > 10 , respectively) allows reasonable structural diversity among the hits particularly under the combined restrictions of two searching pharmacophores and selection *via* a QSAR model. Furthermore, many anti-cancer drugs and especially the ones used for HER2 overexpressing cancers are of higher molecular weights. Therefore, we selected a few hits of molecular weights exceeding 500. The selected hits were with a high predicted activity but of molecular weights < 700 D and rotatable bonds < 13 . These selected hits were fitted against the two pharmacophores and their fit values were substituted in QSAR Equation (3) to determine their predicted bioactivities.

Surviving hits (Figure 5) were fitted against Hypo4/15 and Hypo9/12, and were evaluated against HER2-overexpressing SKOV3 Ovarian cancer cell line and MCF-7 human breast cells. MCF-7 which expresses low levels of HER2 was used to demonstrate the selective HER2 inhibitory activity of tested compounds.[27] The antitumor activities of compounds (**116-127**) were investigated by conducting cell viability assays using tetrazolium dye 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT). Initially, hits were screened at 20 μ M concentrations; subsequently compounds of inhibitory percentages $\geq 60\%$ at 20 μ M on SKOV-3 cells were further assessed to determine their IC_{50} values. Table 4 and Figure 5 show tested hits and their corresponding estimated and experimental inhibitory bioactivities against both SKOV-3 and MCF-7 cells. Results revealed that four NCI hits (**120**, **123**, **125**, and **126**) selectively inhibited growth of SKOV3 cells with IC_{50} values $< 5\mu$ M, while the rest were found to be less active or inactive. These four compounds gave low or no activity toward MCF-7 cells, which may indicate that these are good HER2 inhibitors. The best two hits 120 ($IC_{50} = 1.43 \mu$ M, Table 4) and 126 ($IC_{50} = 1.76 \mu$ M, Table 4) fitted bothHypo 4/15 and Hypo9/12 and how they map the features is shown in Figures 3 and 4, respectively.

Clearly, there are some correlation between the experimental biological activity findings and the predicted values. Such a correlation is exemplified in compound 120 which scored a relatively low IC_{50} values in both the predicted and experimental data. This compound also scored the second highest Fit value in the Hypo4/15 hypotheses. This correlation is also clear in compound 123, where both the predicted and reported IC_{50} values are relatively low. In addition, this compoundalso scored a high Hypo9/12 hypotheses

fit value. In the case of the other two selective compounds, 125 and 126, the correlations were much less pronounced as the predicted IC_{50} values for both of them were relatively high and both compounds exerted remarkable activities against the SKOV-3 Ovarian cell line scoring 3.92 (1.77) for compound 125 and 1.76 (0.19) for 126 compounds. This low correlation may be explained by looking at both the Hypo9/12 hypotheses Fit values as those compounds scored relatively high values. What is also important is that the four compounds (120, 123, 125, and 126) have distinguished structural features with high number of groups capable of forming hydrogen bonding, especially the oxygen and nitrogen containing groups. These groups perhaps are expected to make more hydrogen bonding with the targeted receptors.

4. Conclusions

Several structurally diverse compounds possessing growth inhibitory potency against HER2-overexpressing ovarian cancer cells were identified using pharmacophore modeling and QSAR analysis. The pharmacophore space of HER2 inhibitors was explored via three diverse sets of inhibitors using Discovery Studio HYPOGEN to identify high quality binding model(s). Two pharmacophore models suggest the existence of at least two distinct binding modes accessible to ligands within HER2 binding pocket. *In silico* screening revealed four compounds inhibited the growth of SKOV3 cells with IC_{50} values $< 5\mu M$; these compounds are potential leads for HER2 ovarian cancer. Further screening of hits compounds against breast cancer cells is in progress and will be reported in due course. Structural optimization to enhance their inhibitory activity is the next step.

5. Acknowledgements

We wish to thank the Deanship of Scientific Research at the University of Jordan and Abdul Hameed Shoman Foundation, Amman, Jordan, for financial support. We would like also to acknowledge the Developmental Therapeutics Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute (<http://dtp.cancer.gov>) for providing us with the NCI compounds.

FIGURES

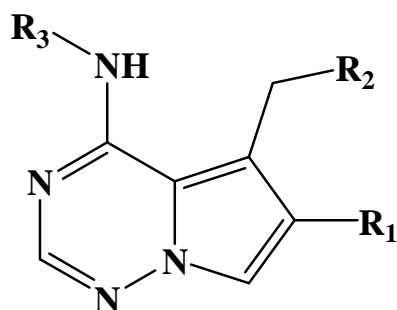
Figure 1. The chemical scaffolds of training compounds; the structures are as in Table 1.

Figure 2. Experimental versus (A) fitted (91 compounds, r^2 : 0.79), and (B) predicted (24 compounds, r^2 PRESS: 0.67) bioactivities calculated from the best QSAR model Eq. (3). 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 error margins.

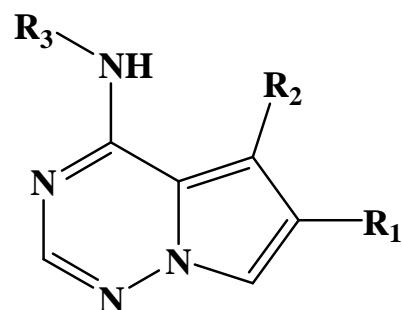
Figure 3. Hypo4/15. (A) Pharmacophoric features of the binding model **Hypo 4/15**: HBD as violet vectored spheres and Hbic as blue spheres, PosIon as red spheres, RingArom as orange vectored spheres, and exclusion volumes as grey spheres. (B) and (C) show **Hypo4/15** fitted against hits 120 ($IC_{50} = 1.43\mu M$, Table 4), 126 ($IC_{50} = 1.76\mu M$, Table 4) respectively. (D) and (E) chemical structure of hits 120 and 126 respectively.

Figure 4. Hypo9/12. (A) Pharmacophoric features of the binding model **Hypo 9/12**: Hbic as blue spheres, PosIon as red spheres, and RingArom as orange vectored spheres. (B) and (C) show **Hypo9/12** fitted against hits 120 ($IC_{50} = 1.43\mu M$, Table 4), 126 ($IC_{50} = 1.76\mu M$, Table 4) respectively. (D) and (E) chemical structure of hits 120 and 126 respectively.

Figure 5. Chemical structure of highest ranking tested hits.

Scaffold A

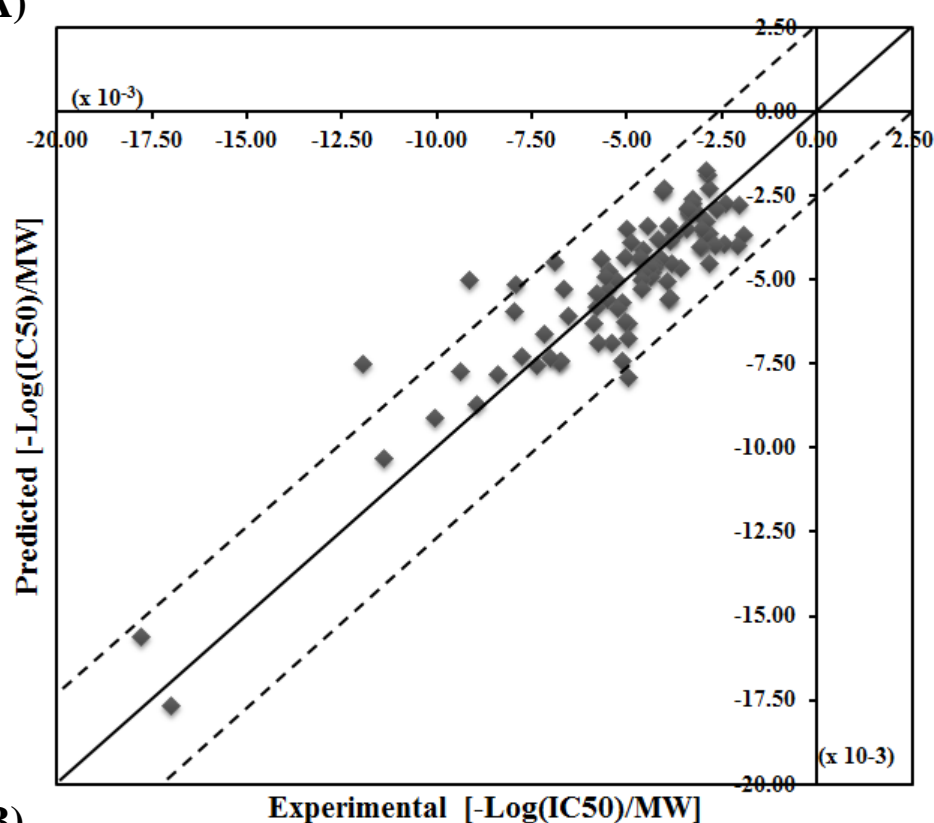
Scaffold A: compounds (1-62), (67-115)

Scaffold B

Scaffold B: compounds (63-66)

Figure 1

(A)



(B)

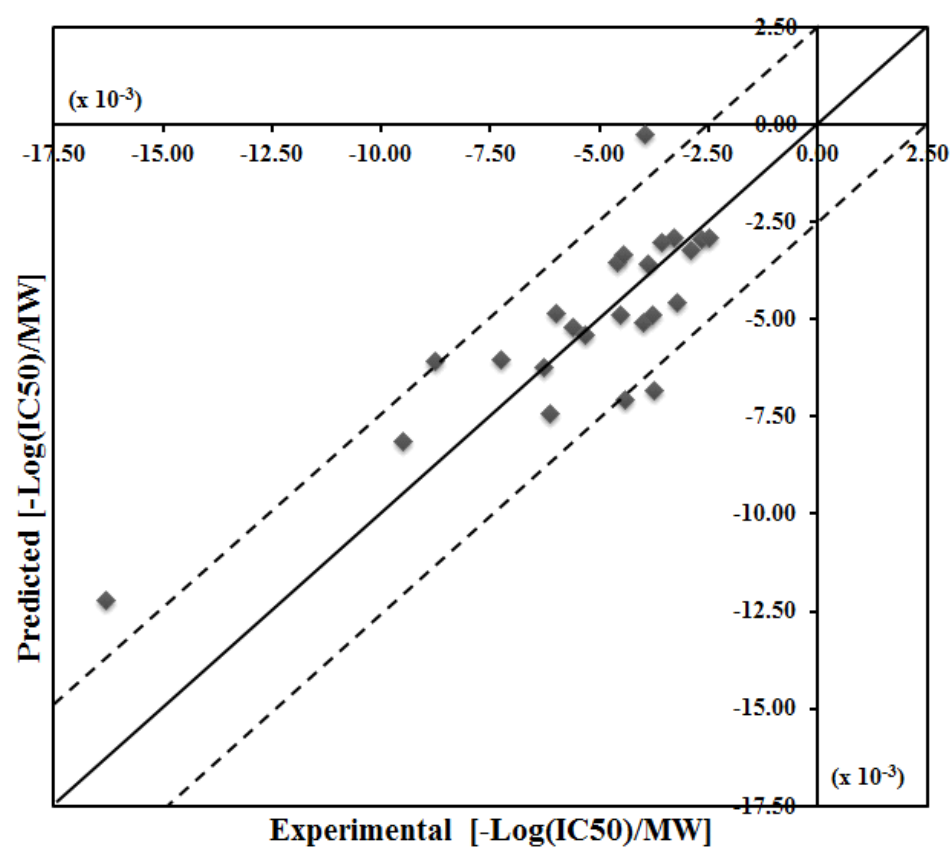


Figure 2

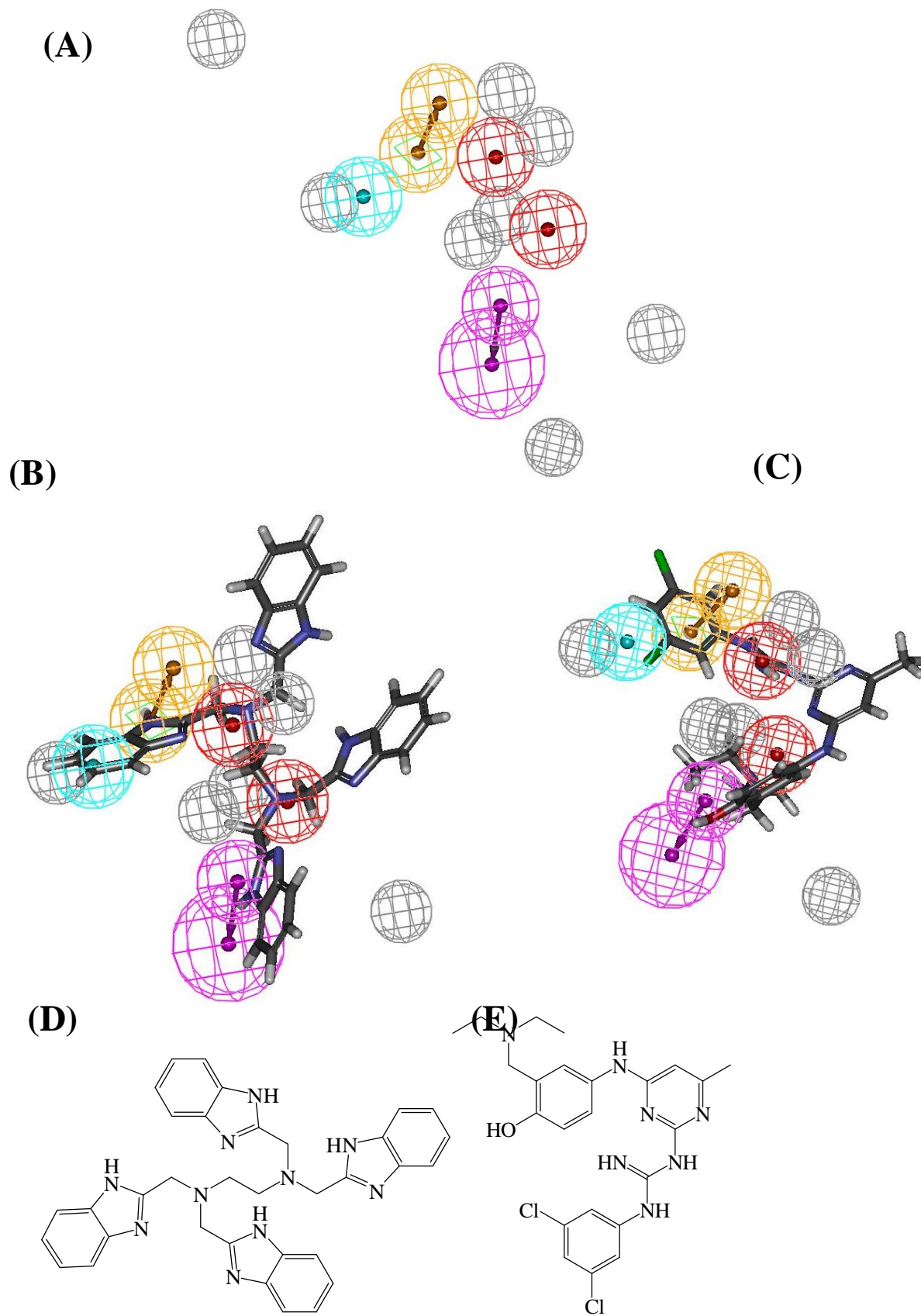


Figure 3

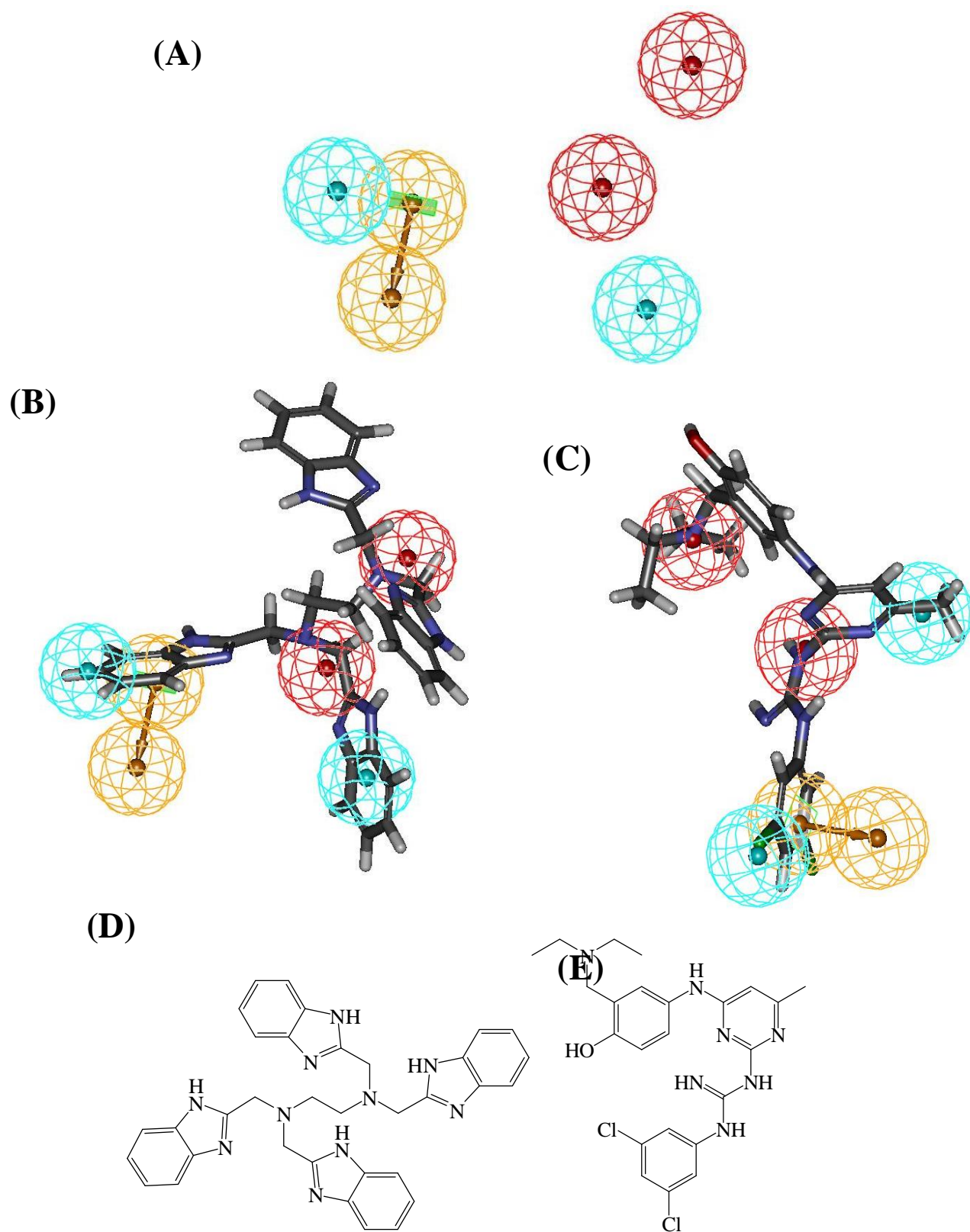
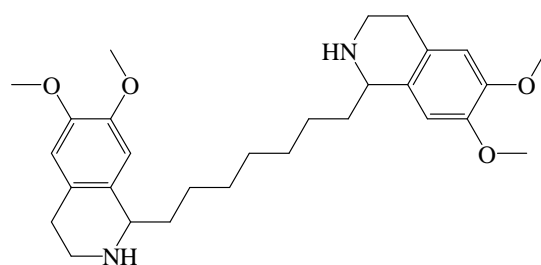
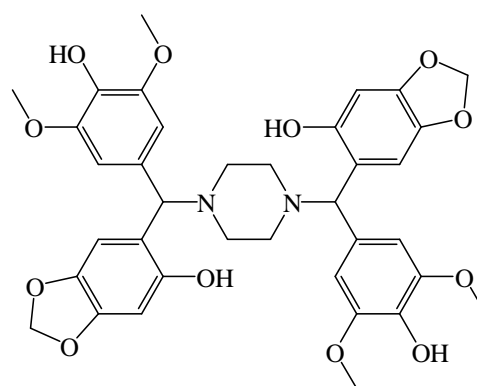
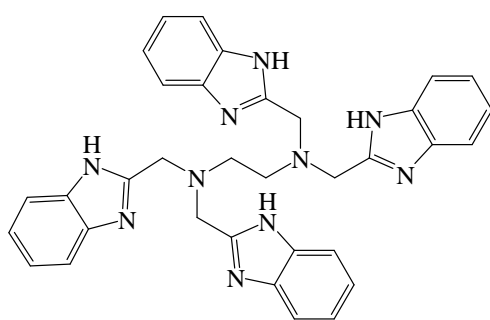
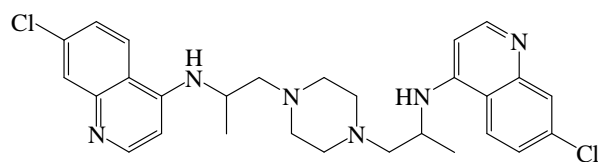
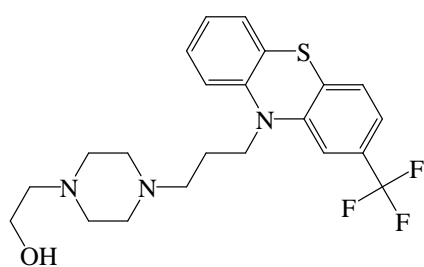
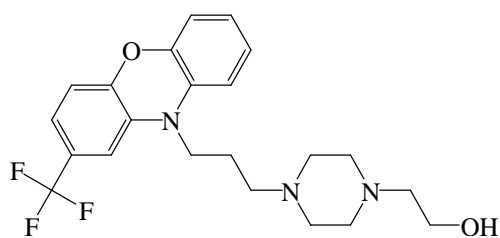
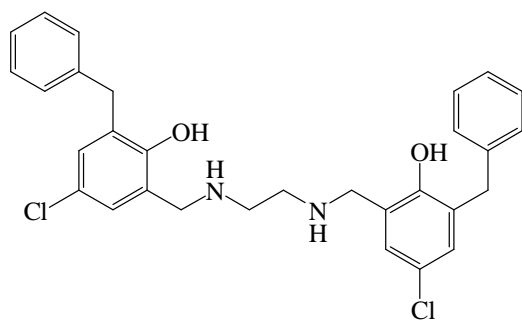
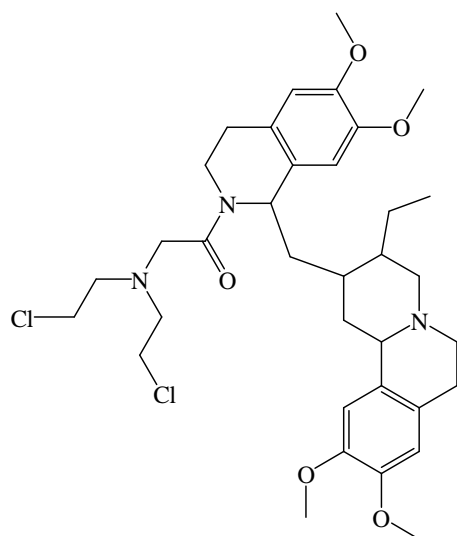
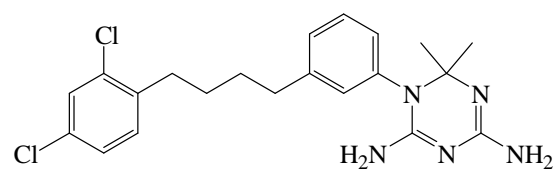
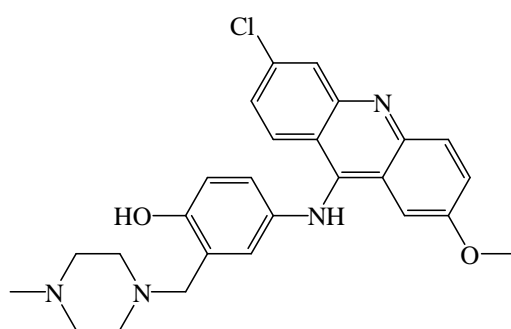
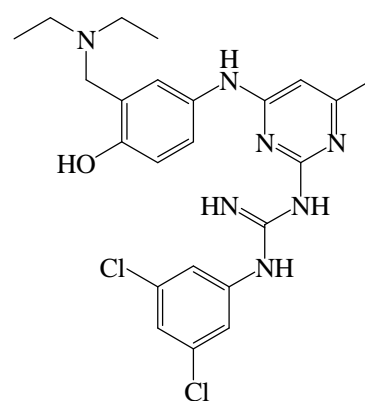
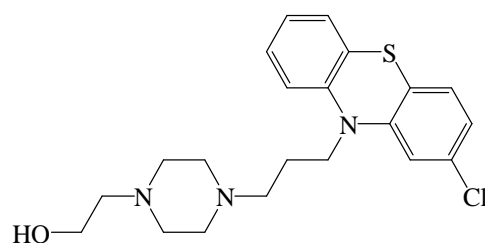


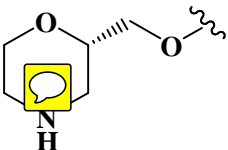
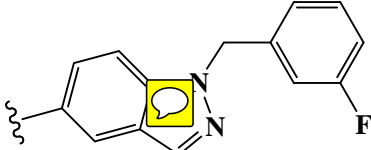

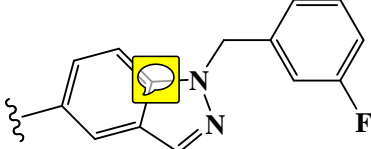
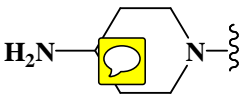
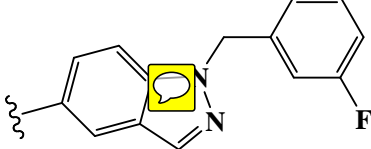
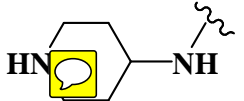
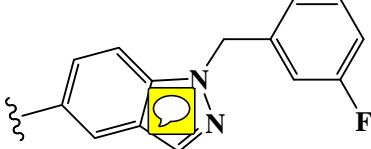

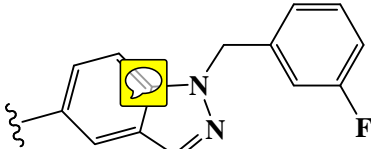

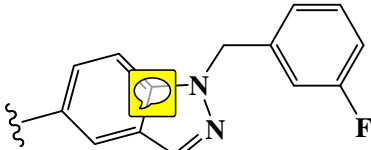
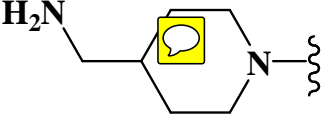
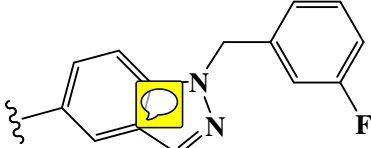
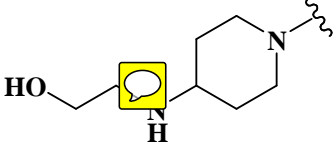
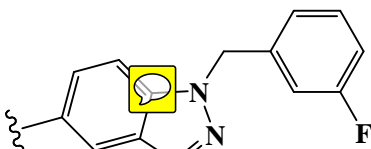
Figure 4

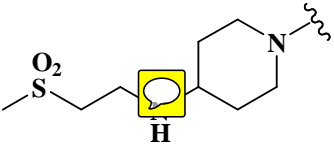
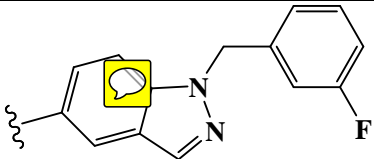

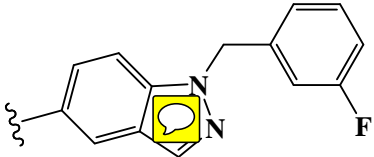
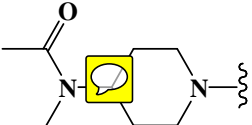
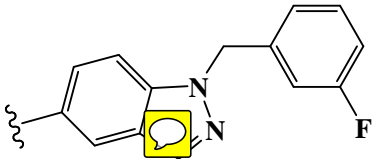
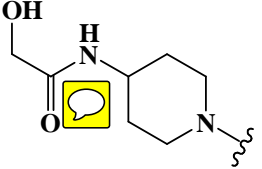
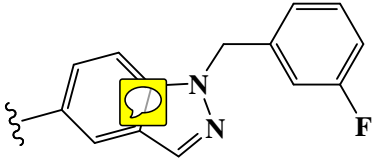
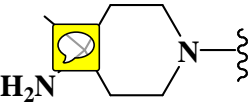
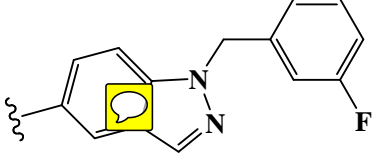
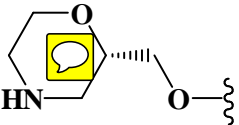
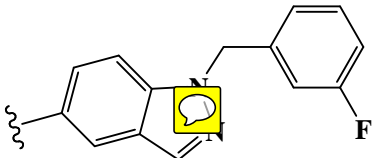
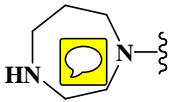
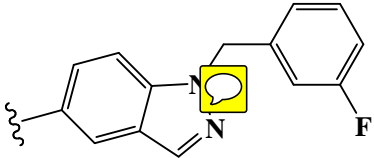
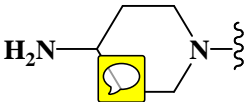
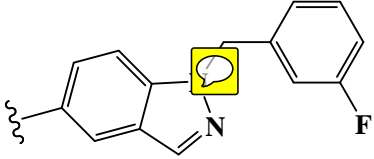
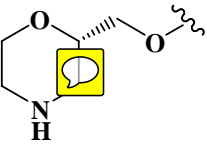
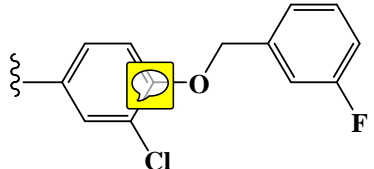



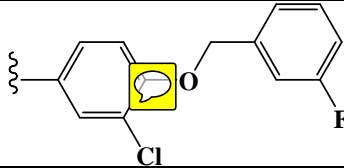

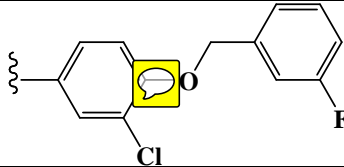
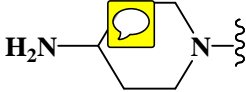
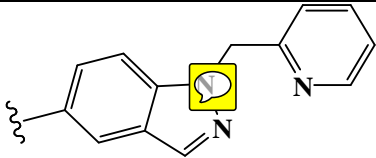

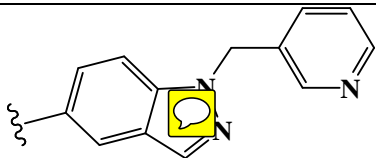

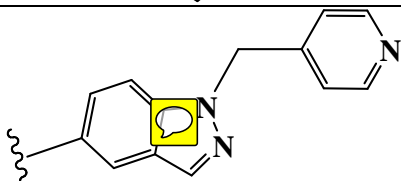
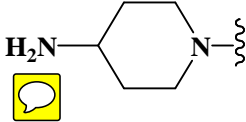
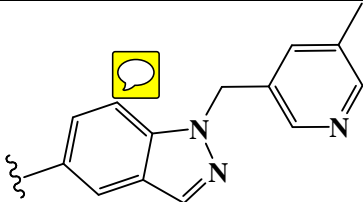
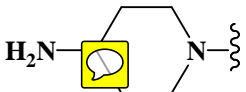
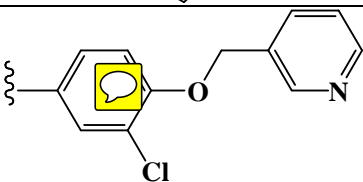

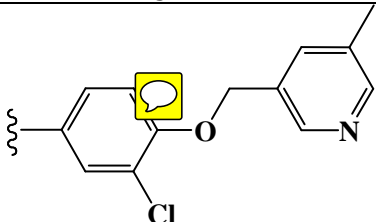
**123****124****125****126****127****Figure 5**


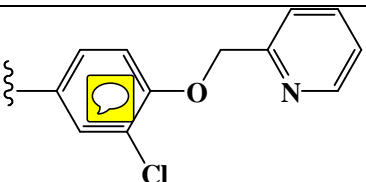
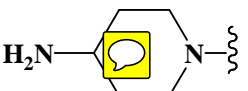
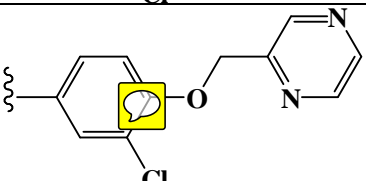

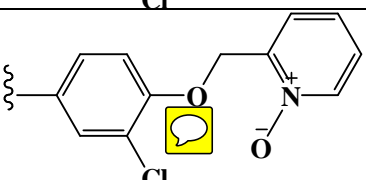
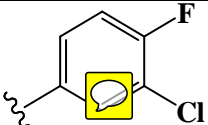
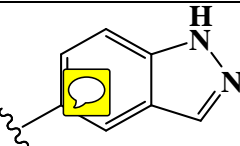
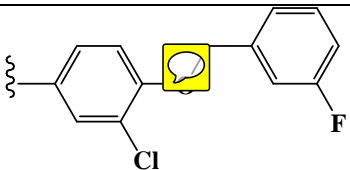
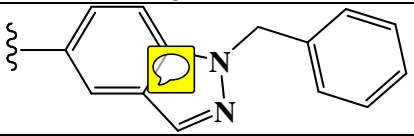
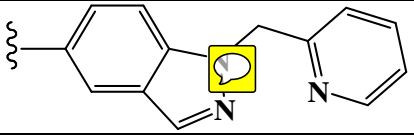
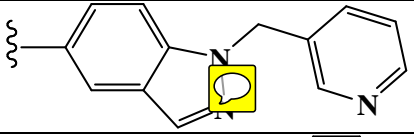
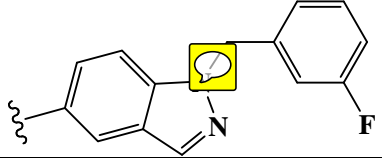
TABLES

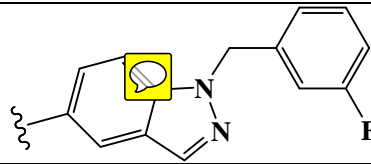
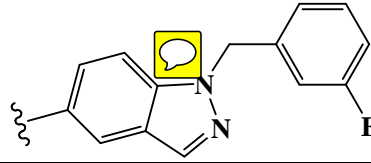
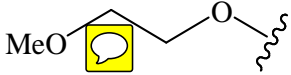
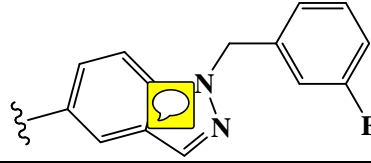
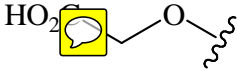
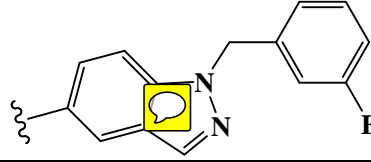
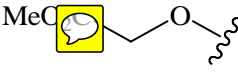
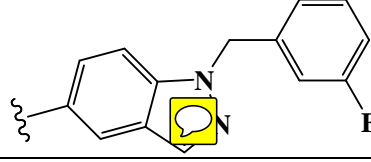
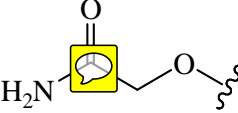
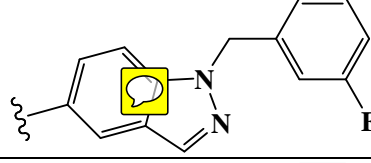

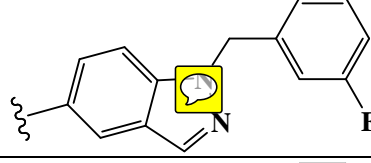
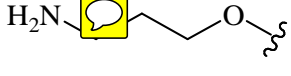
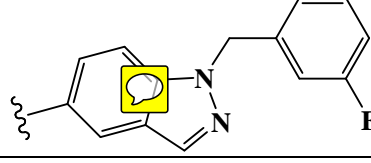
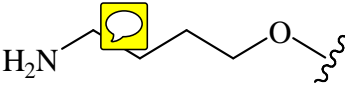
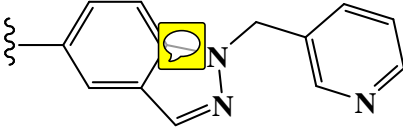
Table 1: HER2 Ligands used to build training set

Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
1	H			55
2	H			27
3	H			23
4	H			76
5	H			70
6**	H			54
7	H			38
8	H			31

Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
9**	H			190
10	H			11
11	H			18
12	H			12
13	H			43
14	OMe			250
15**	OMe			170
16	OMe			43
17	H			560

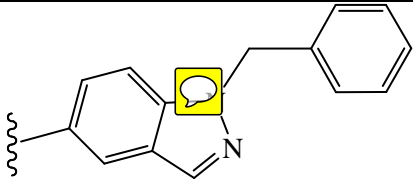
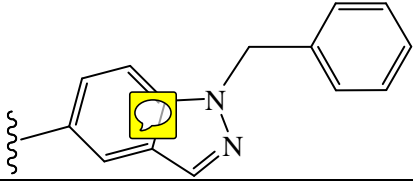
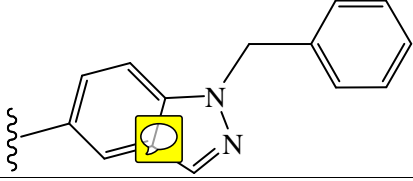
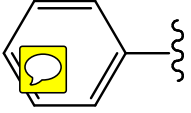
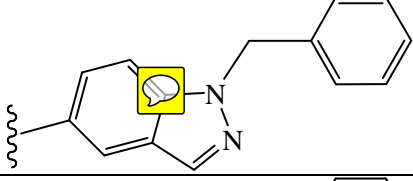
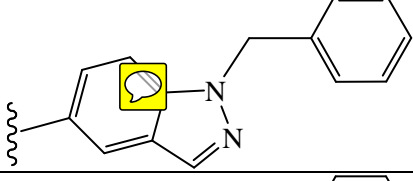
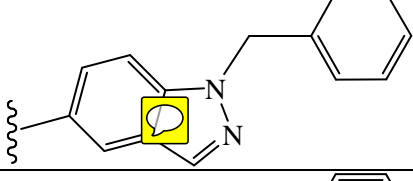
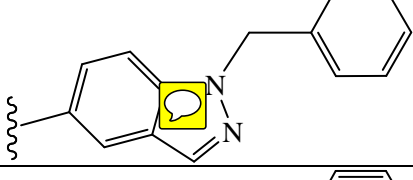
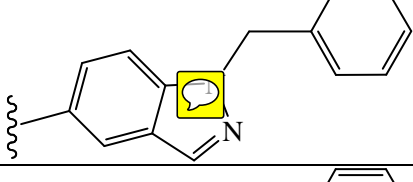
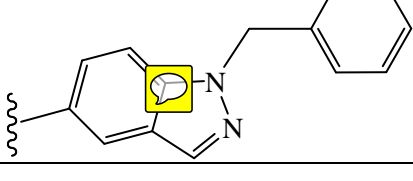
Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
18	H			41
19	H			32
20	H			35
21	H			34
22	H			190
23	H			13
24	H			36
25	H			60

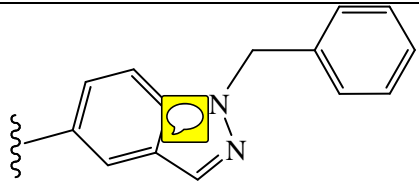
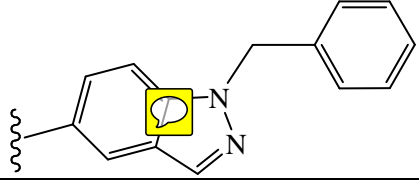
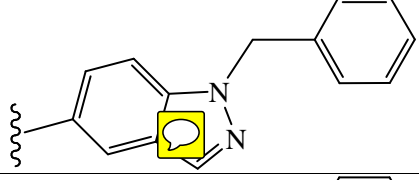
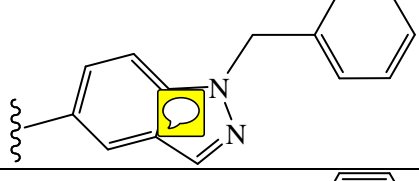
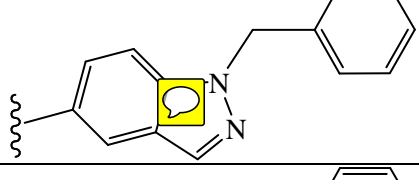
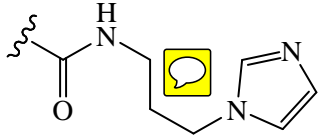
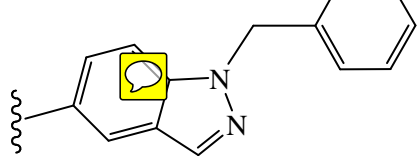
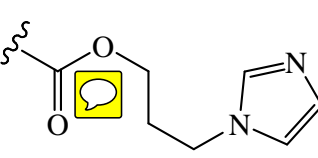
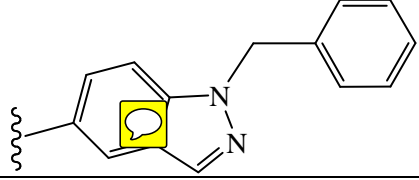
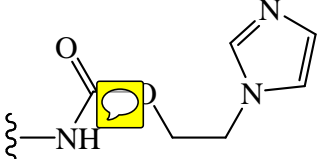
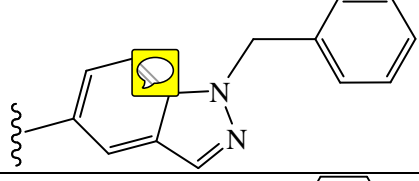
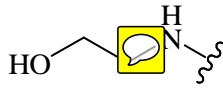
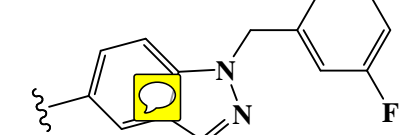
Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
26**	H			22
27	H			26
28	H			18
29	H	H		50000
30	H	H		50000
31	H	H		430
32	H	H		110
33	H	H		310
34	H	H		250
35	H	H		88

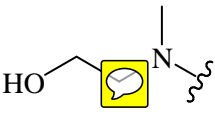
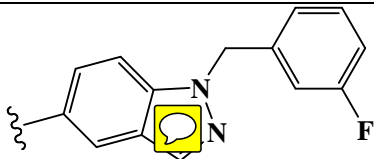
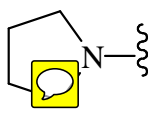
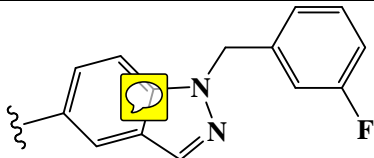
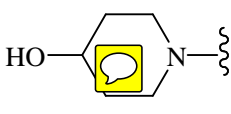
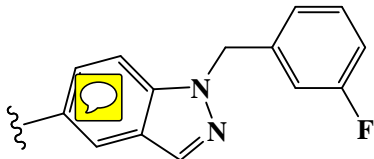
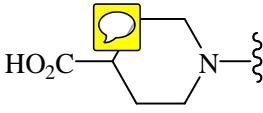
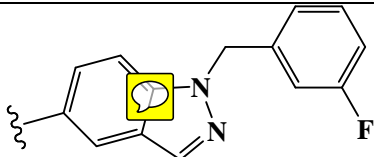
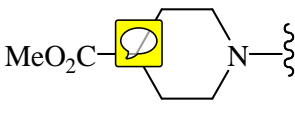
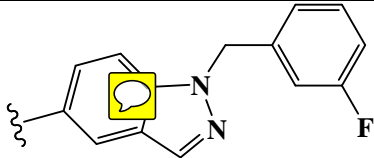
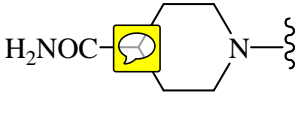
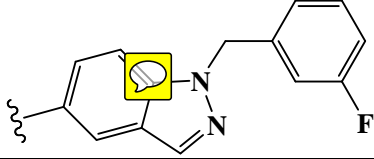
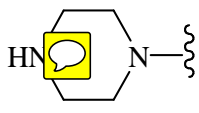
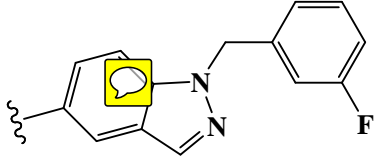

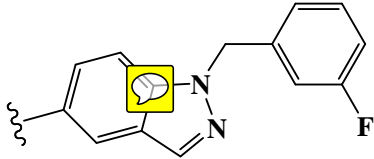
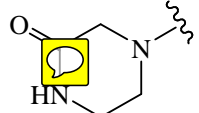
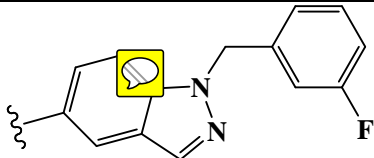
Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
36	H	HO		110
37	H	MeO		420
38	H			420
39	H			18
40**	H			280
41	H			110
42**	H			90
43**	H			60
44	H			130

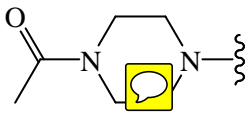
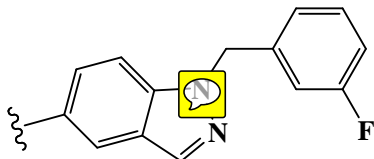

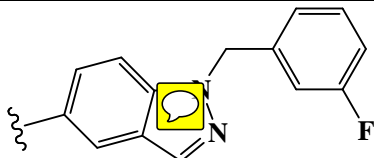
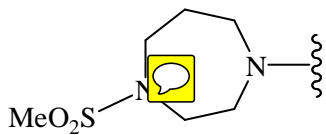
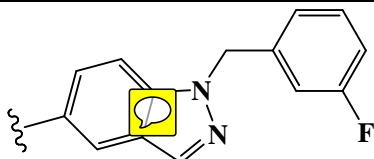

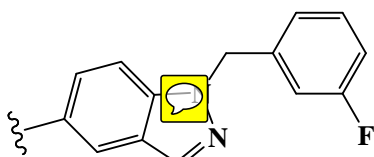
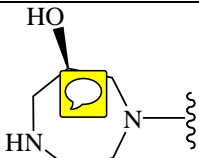
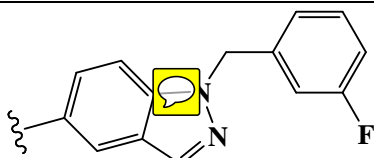
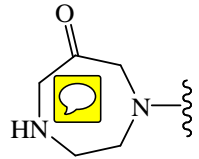
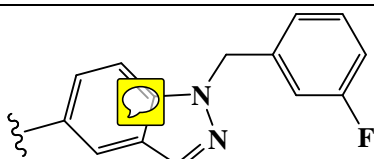
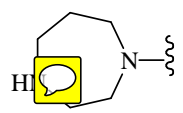
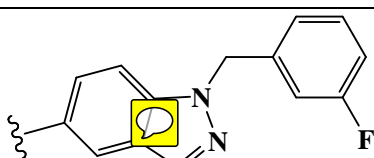
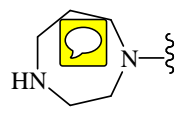
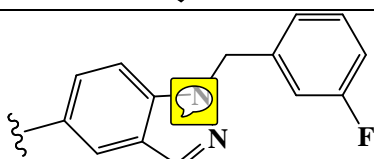
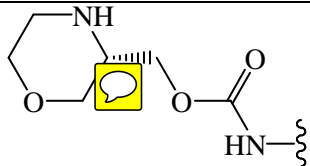
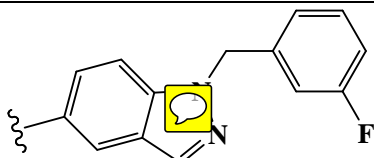
Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
45	H			560
46	H			70
47	H			120
48	H			170
49	H			590
50	H			140
51	H			560
52	H			380
53	H			410

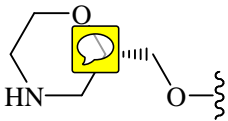
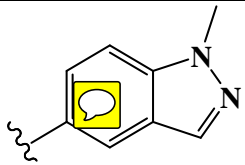
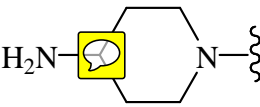
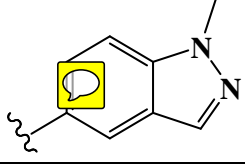
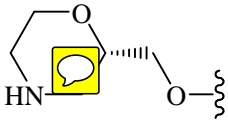
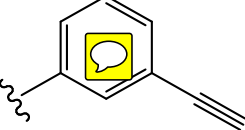
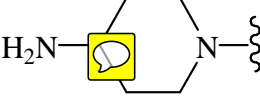
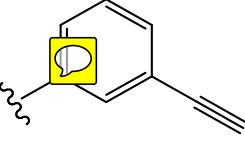
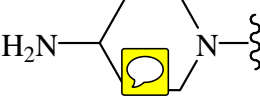

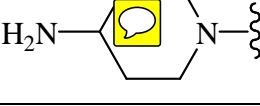
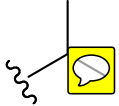
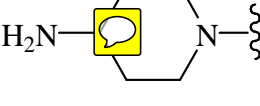

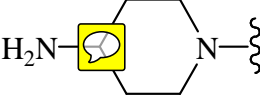

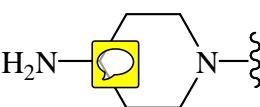

Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
54	H			130
55**	H			41
56	MeO			250
57**	EtO ₂ C	Me		2800
58	EtO ₂ C	Me		3000
59	EtO ₂ C	Me		3300
60	EtO ₂ C	Me		7200
61	EtO ₂ C	Me		180
62**	EtO ₂ C	Me		820
63**	EtO ₂ C	H		390

Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
64	EtO ₂ C	CH ₃		200
65	EtO ₂ C	iPr		210
66	EtO ₂ C	nPr		1200
67**	EtO ₂ C			≥ 25000
68	HO ₂ C	Me		40
69	C(O)NH ₂	Me		110
70	C(O)NHMe	Me		130
71**	C(O)NMe ₂	Me		500
72	CH ₂ OH	Me		730

Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
73	NHCO(O)Bn	Me		170
74	CH ₂ NHEt	Me		9700
75	CH ₂ OEt	Me		1400
76	CH ₂ C(O)NH ₂	Me		2400
77	CH ₂ CO ₂ H	Me		680
78		Me		30
79		Me		240
80		Me		40
81	H			47

Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
82**	H			110
83	H			360
84	H			63
85	H			110
86	H			50000
87	H			310
88	H			81
89	H			120
90	H			200

Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
91**	H			610
92	H			120
93**	H			120
94**	H			75
95**	H			16
96**	H			39
97	OMe			170
98	Me			74
99		H		23

Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
100	H			50000
101	H			950
102	H			670
103	H			10
104	H			40
105**	H			50000
106	H			660
107	H			20
108	H			50


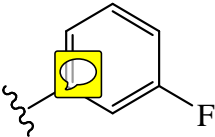

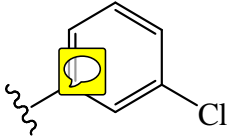
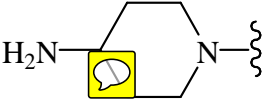
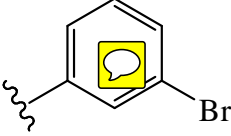
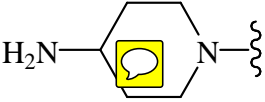
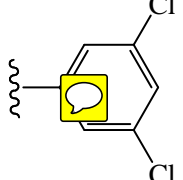
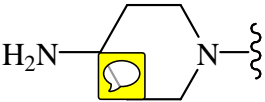
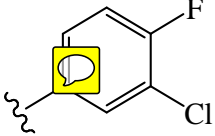

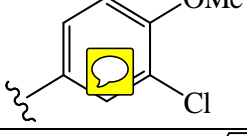
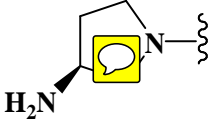
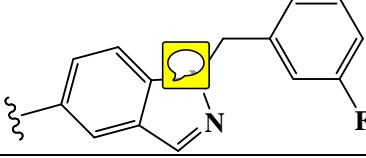
Cpd	R ₁	R ₂	R ₃	IC ₅₀ (nM)
109**	H			32
110	H			9
111	H			6
112	H			130
113**	H			10
114**	H			210
115	H			52

Table 2: Success criteria of representative pharmacophore hypotheses (cluster centers).

N	Run ^a	Hypotheses ^b	Features	Config. cost	Total cost	Null Hypothesis	Residual Cost ^c	R ^d	F-Stat ^e	Cat. Scramble
1	5	9 ^f	HBA, Hbic, PosIon, RingArom	16.10	122.24	160.25	38.01	0.94	4.30	99%
2	6	7	Hbic, PosIon, 2x RingArom, 1x ExVol	15.12	120.31	160.25	39.94	0.92	14.86	99%
3	3	10	HBA, Hbic, 2x PosIon, RingArom	14.70	122.23	160.25	38.02	0.91	13.32	99%
4	2	2	Hbic, 2x PosIon, RingArom	15.12	116.51	160.25	43.74	0.95	18.36	99%
5	2	8	2x Hbic, PosIon, RingArom	15.12	118.30	160.25	41.95	0.94	8.62	99%
6	7	7	Hbic, 2x PosIon, 2x RingArom	14.69	120.01	160.25	40.24	0.92	17.24	99%
7	14	1	HBA, Hbic, PosIon, RingArom, 6x ExVol	15.03	80.95	125.88	44.93	0.97	39.17	99%
8	14	3	HBA, Hbic, PosIon, RingArom, 4x ExVol	15.03	82.49	125.88	43.39	0.96	59.03	99%
9	10	8	Hbic, 2x PosIon, RingArom	14.73	83.14	125.88	42.74	0.95	34.91	99%
10	15	4^g	HBD, Hbic, 2x PosIon, RingArom, 7x ExVol	13.28	83.93	125.88	41.95	0.93	46.57	99%
11	16	5	HBD, 2x Hbic, PosIon, RingArom, 4x ExVol	13.06	93.93	125.88	31.95	0.91	60.00	99%
12	12	9^g	2x Hbic, 2x PosIon, RingArom	13.06	83.82	125.88	42.06	0.93	48.63	99%
13	21	6	HBD, Hbic, PosIon, RingArom	17.90	154.61	205.76	51.15	0.93	29.77	99%
14	21	9	HBA, HBD, Hbic, RingArom	17.90	156.02	205.76	49.74	0.92	24.84	99%
15	21	10	HBD, Hbic, 2x RingArom	17.90	156.48	205.76	49.28	0.92	16.26	99%
16	17	10	HBD, 2x Hbic, RingArom	17.90	155.69	205.76	50.17	0.92	22.61	99%
17	22	1	HBD, Hbic, 2x RingArom, 1x ExVol	17.11	152.92	205.76	52.84	0.94	17.58	99%
18	22	6	2x Hbic, PosIon, RingArom, 4x ExVol	17.11	155.17	205.76	50.59	0.92	11.51	99%
19	19	9	HBD, 2x Hbic, PosIon, RingArom	15.99	154.37	205.76	51.93	0.92	12.23	99%
20	24	7	2x Hbic, 2x PosIon, RingArom	14.73	152.31	205.76	53.45	0.92	12.99	99%

^aCorrespond to runs in Table 3, ^bHigh ranking representative hypotheses (in their corresponding clusters, see section 4.1.6). ^cDifference between total cost and the cost of the corresponding null hypotheses. ^dCorrelation coefficients between pharmacophore-based bioactivity estimates and bioactivities of corresponding training compound (subsets in Table 2). ^eFisher statistic calculated based on the linear regression between the fit values of all collected inhibitors(1-115, Table 1) against pharmacophore hypothesis (employing the "best fit" option) and their respective inhibitory bioactivities (log (1/IC₅₀) values). ^fRanking of hypotheses is as generated by software in each automatic run (ranging from 1-10). ^gBolded pharmacophores appeared in the best QSAR model Equation (3).

Table3.Numbers of captured hits by Hypo4/15 and Hypo9/12

3D Database ^a	Pharmacophore models		
	Post screening filtering ^b	Hypo4/15	Hypo9/12
NCI ^a	Before	126	996
	After	77	479
BrugBank ^c		11	34

^aNCI: national cancer institute list of available compounds (254,805 structures),

^bUsing Lipinski's and Veber's rules,

^c DrugBank: drug bank list of available compounds (6825 structures).

Table 4. Experimental bioactivities of high-ranking hit molecules.

Hits ^a		Best Fit values ^b		QSAR IC ₅₀ Predictions (μM)	Experimental Cytotoxicity MCF-7 Breast cell line	Experimental Cytotoxicity in HER2 Overexpressing SKOV-3 Ovarian cell line	
Compound	Name	Hypo4/15	Hypo9/12		%Inhibition at 20 μM	%Inhibition at 20 μM	IC ₅₀ (μM) ^{c,d}
116	NCI 11963	0	6.677	0.050	59.3± 0.010	68.5 ± 0.006	12.5 (2.96) ^e
117	NCI 53660	2.118	3.403	0.117	0.00	76.5± 0.004	12.97 (1.4) ^e
118	NCI 62323	1.471	4.007	0.164	59.4± 0.009	71.5± 0.003	50.2 (1.6) ^e
119	NCI 129790	2.502	6.721	0.117	38.1± 0.051	76.5± 0.004	6.71 (0.76) ^e
120	NCI 344032	5.745	3.507	0.0002	20.2±0.019	80.8 ± 0.004	1.43 (0.47)^e
121	NCI 369077	6.484	2.681	0.000003	41.4±0.071	74.8± 0.03	25.99 (0.06) ^e
122	NCI 44451	0	7.563	0.114	48.8±0.003	60.3± 0.002	81.6 (6.35) ^e
123	NCI 101142	0	7.323	0.006	7.27±0.039	82.2± 0.008	4.31 (0.48)^e
124	NCI 104124	0	7.389	0.144	1.16±0.006	66.2± 0.011	18.72 (0.81) ^e
125	NCI 130813	3.592	7.507	0.234	31.5±0.031	70.2± 0.002	3.92 (1.77)^e
126	NCI 157531	2.367	6.683	0.190	0.00	83.06± 0.003	1.76 (0.19)^e
127	Perphenazine	2.481	3.611	0.860	50.9±0.031	67.4± 0.006	41.8 (2.50) ^e
136	Trastuzumab (Herceptin) ^f	----	----	----		41.7 (6.56)	

^aChemical structures shown in figure 5. ^bFit values calculated against respective hypotheses, ^cExperimental IC₅₀ values determined at 11 inhibitor concentrations (100, 50, 25, 12.5, 6.25, 3.12, 1.56, 0.78, 0.39, 0.195 and 0.098 μM), ^dValues in brackets represent the correlation coefficients of the corresponding dose-response curve., ^eFitting was performed using Graph Pad Prism (version5.04) via fitting against sigmoidal dose-inhibition model (Please refer to figure A under Supplementary data for corresponding dose/response curves), ^fTrastuzumab (Herceptin) standard inhibitor (positive control). Bolded his are the highest ranking four.

References

- ¹ C. L. Arteaga, J.A. Engelman, ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell*. 25(2014), 282–303.
- ² J. Schlessinger, Cell signaling by receptor tyrosine kinases. *Cell*, 103 (2000), 211–225.
- ³ S. Scholl, P. Beuzeboc, P. Pouillart, Targeting HER2 in other tumor types. *Ann Oncol*. 12(2001), Suppl 1:S81-S87.
- ⁴ S. Menard, P. Casalini, M. Campiglio, S.M. Pupa, E. Tagliabue, Role of HER2/neu in tumor progression and therapy. *Cell Mol Life Sci*. 61(2004), 2965-2978.
- ⁵ J. Mendelsohn, Blockade of receptors for growth factors: an anticancer therapy. The Fourth Annual Joseph H. Burchenal American Association for Cancer Research Clinical Research Award Lecture, *Clin. Cancer Res.* 6 (2000), 747–753.
- ⁶ V. N. Kristensen, T. Sorlie, J. Geisler, A. Langerød, N. Yoshimura, R. Karesen, N. Harada, P.E. Lonning, A.L. Borresen-Dale, Gene expression profiling of breast cancer in relation to estrogen receptor status and estrogen-metabolizing enzymes: clinical implications. *Clin. Cancer Res.*, 11 (2005), 878s-883s.
- ⁷ C. M. Perou, T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P.O. Brown, D. Botstein, Molecular portraits of human breast tumours. *Nature* 406 (2000), 747-752.
- ⁸ S. M. Pupa, E. Tagliabue, S. Menard, A. J. Anichini, HER-2: a biomarker at the crossroads of breast cancer immunotherapy and molecular medicine. *Cell. Physiol.*, 205 (2005), 8-10.
- ⁹ D.J. Slamon, G.M. Clark, S.G. Wong, W.J. Levin, A. Ullrich, W.L. McGuire, Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene, *Science*, 235 (1987), 177–182.
- ¹⁰ J.S. Ross, J.A. Fletcher, The HER-2/neu oncogene in breast cancer: prognostic factor, predictive factor, and target for therapy, *Stem Cells*, 16 (1998), 413–428.
- ¹¹ J.S. Ross, J.A. Fletcher, G.P. Linette, J. Stec, E. Clark, M. Ayers, W.F. Symmans, L. Pusztai, K.J. Bloom. The HER-2/neu gene and protein in breast cancer: biomarker and target of therapy, *Oncologist*, 8 (2003), 307–325.
- ¹² D.J. Slamon, W. Godolphin, L.A. Jones, J.A. Holt, S.G. Wong, D.E. Keith, W.J. Levin, S.G. Stuart, J. Udove, A. Ullrich, Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer, *Science*, 244 (1989), 707–712.
- ¹³ D.B. Agus, P.A. Bunn Jr., W. Franklin, M. Garcia, R.F. Ozols, HER-2/neu as a therapeutic target in non-small cell lung cancer, prostate cancer, and ovarian cancer, *Semin. Oncol.*, 27 (2000), 53–63.

-
- ¹⁴G. Cox, M. Vyberg, B. Melgaard, J. Askaa, A. Oster, K.J. O'Byrne, Herceptest: HER2 expression and gene amplification in non-small cell lung cancer, *Int. J. Cancer*, 92 (2001), 480–483.
- ¹⁵F.R. Hirsch, M. Varella-Garcia, W.A. Franklin, R. Veve, L. Chen, B. Helfrich, C. Zeng, A. Baron, P.A. Jr. Bunn, Evaluation of HER-2/neu gene amplification and protein expression in non-small cell lung carcinomas, *Br. J. Cancer*, 86 (2002), 1449–1456.
- ¹⁶H.K.W. Koeppen, B.D. Wright, A.D. Burt, P. Quirke, A.M. McNicol, N.O. Dybdal, et al., Overexpression of HER2/neu in solid tumours: an immunohistochemical survey, *Histopathology* 38 (2001) 96–104
- ¹⁷B.F. Calvo, A.M. Levine, M. Marcos, Q.F. Collins, M.V. Iacocca, L.S. Caskey, C.W. Gregory, Y. Lin, Y.E. Whang, H.S. Earp, J.L. Mohler, Human epidermal receptor-2 expression in prostate cancer, *Clin. Cancer Res.*, 9 (2003), 1087–1097.
- ¹⁸I. Osman, H.I. Scher, M. Drobnjak, D. Verbel, M. Morris, D. Agus, J.S. Ross, C. Cordon-Cardo, HER-2/neu (p185neu) protein expression in the natural or treated history of prostate cancer, *Clin. Cancer Res.*, 7 (2001), 2643–2647.
- ¹⁹Santin, A.D, Bellone, S, Roman, JJ, McKenney, JK, Pecorelli S. Trastuzumab treatment in patients with advanced or recurrent endometrial carcinoma overexpressing HER2/neu. *Int. J. Gynaecol. Obstet.*, 102 (2008), 128–131.
- ²⁰D.M. Reese, E.J. Small, G. Magrane, F.M. Waldman, K. Chew, D. Sudilovsky, HER2 protein expression and gene amplification in androgen-independent prostate cancer, *Am. J. Clin. Pathol.*, 116 (2001), 234–239.
- ²¹N. E. Hynes, H. A. Lane, ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat. Rev. Cancer*, 5 (2005), 341–354.
- ²²P. Nagy, A. Jenei, A.K. Kirsch, J. Szollosi, S. Damjanovich, T.M. Jovin, Activation-dependent clustering of the erbB2 receptor tyrosine kinase detected by scanning near-field optical microscopy, *J. Cell Sci.*, 112 (1999). 1733-1741.
- ²³R. Kaufmann, P. Müller, G. Hildenbrand, M. Hausmann, C. Cremer, Analysis of Her2/neu membrane protein clusters in different types of breast cancer cells using localization microscopy, *J. Microsc.*, 242 (2010), 46-54.
- ²⁴R. Kaufmann, P. Müller, G. Hildenbrand, M. Hausmann, C. Cremer, Analysis of Her2/neu membrane protein clusters in different types of breast cancer cells using localization microscopy, *J. Microsc.*, 242 (2010), 46-54.
- ²⁵B. E.Hillner, T. J.Smith, Do the large benefits justify the large costs of adjuvant breast cancer trastuzumab?, *J. Clin. Oncol.* 25 (2007), 611-613.
- ²⁶B. Moy, P.Kirkpatrick, S. Kar, P. Goss, Lapatinib, *Nat. Rev. Drug Discovery*, 6 (2007), 431–432.

-
- ²⁷R. Jr. Roskoski, The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol. Res.*, 79 (2014) 34–74.
- ²⁸C. Larbouret, B. Robert, I. Navarro-Teulon, S. Thezenas, M.Z. Ladjemi, S. Morisseau, E. Campigna, F. Bibeau, J.P. Mach, A. Pelegrin, D. Azria, In vivo therapeutic synergism of anti-epidermal growth factor receptor and anti-HER2 monoclonal antibodies against pancreatic carcinomas. *Clin. Cancer Res.*, 13 (2007), 3356–3362.
- ²⁹A. Magnifico, L. Albano, S. Campaner, D. Delia, F. Castiglioni, P. Gasparini, G. Sozzi, E. Fontanella, S. Menard, E. Tagliabue, Tumor-initiating cells of HER2- positive carcinoma cell lines express the highest oncoprotein levels and are sensitive to trastuzumab. *Clin. Cancer Res.*, 15 (2009), 2010–2021.
- ³⁰L. C. Amler, Y. Wang and G. Hampton (2012). HER2 as a Therapeutic Target in Ovarian Cancer, *Ovarian Cancer - Clinical and Therapeutic Perspectives*, Dr. Samir Farghaly (Ed.), ISBN: 978-953-307-810-6.
- ³¹M.A. Olayioye, R.N. Neve, H.A. Lane, N.E. Hynes, The ErbB signaling network: receptor heterodimerization in development and cancer, *Eur. Mol. Biol. Org. J.*, 19 (2000), 3159–3167.
- ³²D.J. Riese II, D.F. Stern, Specificity within the EGF family/ErbB receptor family signaling network, *Bioessays* 20 (1998), 41–48.
- ³³J.T. Jones, R.W. Akita, M.X. Sliwkowski, Binding specificities and affinities of EGF domains for ErbB receptors, *Fed. Eur. Biochem. Soc. Lett.*, 447 (1999), 227–231.
- ³⁴E. Tzahar, H. Waterman, X. Chen, G. Levkowitz, D. Karunagaran, S. Lavi, et al., A hierarchical network of interreceptor interactions determines signal transduction by neu differentiation factor/neuregulin and epidermal growth factor, *Mol. Cell. Biol.*, 16 (1996), 5276–5287.
- ³⁵I.B. Bersuker, S. Bahçeci, J.E. Boggs, In *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, (2000) 457–473.
- ³⁶R. F. Tayyem, H. M. Zalloum, R. M. Elmaghrabi, AL-M. Yousef, and M. S. Mubarak. Ligand-based designing, *in silico* screening, and biological evaluation of new potent fructose-1,6-bisphosphatase inhibitors. *Eur. J. Med. Chem.*, 56 (2012), 70-95.
- ³⁷ B.A. Akhoon, S.K. Gupta, V. Verma, G. Dhaliwal, M. Srivastava, S. K. Gupta, R.F. Ahmad, In silico designing and optimization of anti-breast cancer antibody mimetic oligopeptide targeting HER-2 in women. *Mol Graph Model.* 28(2010), 664-669.
- ³⁸ M. Jamalan, M. Zeinali, A.E. Barzegari, Design of peptidomimetics for inhibition of HER2 receptor dimerization by a combination of virtual screening, MD simulations, and QSAR in silico methods. *Chem Biol Drug Des.* 81(2013), 455-462.

-
- ³⁹H. Mastalerz, M. Chang, P. Chen, B.E. Fink, A. Gavai, W.C. Han, W. Johnson, D. Langley, F.Y. Lee, K. Leavitt, P. Marathe, D. Norris, S. Oppenheimer, B. Slecicka, J. Tarrant, J.S. Tokarski, G.D. Vite, D.M. Vyas, H. Wong, T.W. Wong, H. Zhang, G. Zhang. 5-((4-Aminopiperidin-1-yl)methyl)pyrrolotriazine dual inhibitors of EGFR and HER2 protein tyrosine kinases. *Bioorg Med Chem Lett.*, 17(2007), 4947-54.
- ⁴⁰H. Mastalerz, M. Chang, P. Chen, P. Dextraze, B.E. Fink, A. Gavai, B. Goyal, W.C. Han, W. Johnson, D. Langley, F.Y. Lee, P. Marathe, A. Mathur, S. Oppenheimer, E. Ruediger, J. Tarrant, J.S. Tokarski, G.D. Vite, D.M. Vyas, H. Wong, T.W. Wong, H. Zhang, G. Zhang. New C-5 substituted pyrrolotriazine dual inhibitors of EGFR and HER2 protein tyrosine kinases. *Bioorg Med Chem Lett.*, 17 (2007), 2036-2042.
- ⁴¹H. Mastalerz, M. Chang, A. Gavai, W. Johnson, D. Langley, F.Y. Lee, P. Marathe, A. Mathur, S. Oppenheimer, J. Tarrant, J.S. Tokarski, G.D. Vite, D.M. Vyas, H. Wong, T.W. Wong, H. Zhang, G. Zhang. Novel C-5 aminomethyl pyrrolotriazine dual inhibitors of EGFR and HER2 protein tyrosine kinases. *Bioorg. Med. Chem. Lett.*, 17 (2007), 2828-2833.
- ⁴²B.E. Fink, D. Norris, H. Mastalerz, P. Chen, B. Goyal, Y. Zhao, S.H. Kim, G.D. Vite, F.Y. Lee, H. Zhang, S. Oppenheimer, J.S. Tokarski, T.W. Wong, A.V. Gavai. Novel pyrrolo[2,1-f][1,2,4]triazin-4-amines: Dual inhibitors of EGFR and HER2 protein tyrosine kinases. *Bioorg. Med. Chem. Lett.* 21(2011), 781-785.
- ⁴³B.E. Fink, G.D. Vite, H. Mastalerz, J.F. Kadow, S.H. Kim, K.J. Leavitt, K. Du, D. Crews, D. T. Mitt, T.W. Wong, J.T. Hunt, D.M. Vyas, J.S. Tokarski. New dual inhibitors of EGFR and HER2 protein tyrosin kinases. *Bioorg. Med. Chem. Lett.*, 15 (2005), 4774-4779.
- ⁴⁴Catalyst, version 4.11 Users' Manual, Accelrys Software Inc., San Diego, 2005.
- ⁴⁵H. Li, J. Sutter, R. Hoffmann, In *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, (2000), 173–189.
- ⁴⁶J. Sutter, O. Güner, R. Hoffman, H. Li, M. Waldman, in: O.F. Güner (Ed.), *Effect of Variable Weights and Tolerances on Predictive model Generation. Pharmacophore Perception, Development and Use in Drug Design*, International University Line, California, (2000), 501–511.
- ⁴⁷Y. Kurogi, O. F. Guner, *Pharmacophore modeling and three-dimensional database searching for drug design using catalyst*, *Curr Med Chem.*, 8 (2001), 1035-1055.
- ⁴⁸I.B. Bersuker, S. Bahçeci, J.E. Boggs, In *Pharmacophore Perception, Development, and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, (2000) 457–473.

-
- ⁴⁹K. Poptodorov, T. Luu R. Hoffmann In: T. Langer and R.D. Hoffmann, Editors, *Methods and Principles in Medicinal Chemistry, Pharmacophores and Pharmacophores Searches 2*, Wiley-VCH, Weinheim Germany (2006), 17-47.
- ⁵⁰R. Fischer. *The Principle of Experimentation Illustrated by a Psycho-Physical Experiment (Chapter II)*, Hafner Publishing Co., (8th ed.) New York, (1966).
- ⁵¹E.M. Krovat, T. Langer, Non-peptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification. *J. Med. Chem.*, 46 (2003), 716–726.
- ⁵²J. H. Friedman, *Multivariate Adaptive Regression Splines (with discussions) (MARS)*, *Anal. Stat.*, 19 (1991), 1-67.
- ⁵³E. Perola, An analysis of the binding efficiencies of drugs and their leads in successful drug discovery programs, *J. Med. Chem.*, 53 (2010), 2986–2997.
- ⁵⁴L.H. Andrew, R.G. Colin, A. Alexander, Ligand efficiency: a useful metric for leadselection, *Drug Discov. Today*, 9 (2004), 430–431.
- ⁵⁵Reynolds, C.H.; Tounge, B.A.; Bembenek, S.D. Ligand Binding Efficiency: Trends, Physical Basis, and Implications *J. Med. Chem.*, 51 (2008), 2432-2438.
- ⁵⁶M. Congreve, G. Chessari, D. Tisi, A.J. Woodhead, Recent Developments in Fragment-Based Drug Discovery. *J. Med. Chem.*, 51(2008), 3661-3680.
- ⁵⁷C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv DrugDel Rev.*, 46 (2001), 3–26.
- ⁵⁸D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, *J. Med. Chem.*, 45 (2002), 2615–2623.
- ⁵⁹R.P. Sheridan, S.K. Kearsley, Why do we need so many chemical similarity search methods, *Drug Discov. Today*, 7 (2002), 903–910
- ⁶⁰Discovery Studio version 2.5.5 (DS 2.5) User Manual (2009) Accelrys Inc, San Diego.
- ⁶¹L.F. Ramsey, W.D. Schafer, *The Statistical Sleuth*, 1st ed.; Wadsworth Publishing Company: Belmont, CA, (1997).
- ⁶²Accelrys Software Inc., CERIOUS2 (2005) QSAR Users' Manual, Version 4.10; Accelrys Inc., San Diego.
- ⁶³H. Wiener, Structural determination of paraffin boiling points. *J. Am. Chem. Soc.*, 1 (1947), 17–20.
- ⁶⁴D. H. Rouvray, The rich legacy of half a century of the Wiener index, in Rouvray, Dennis H.; King, Robert Bruce, *Topology in Chemistry: Discrete Mathematics of Molecules*, Horwood Publishing, (2002), 16–37.

-
- ⁶⁵W. R. Muller, K. Szymanski, J. von Knop, N. Trinajstić, An Algorithm for construction of the molecular distance matrix, *J. Comput. Chem.*, 8 (1987), 170-173.
- ⁶⁶L.B. Kier, Indexes of molecular shape from chemical graphs, *Computational Chemical Graph Theory*, Rouvray, D. H., Ed., Nova Science: New York (1990).
- ⁶⁷L.H. Hall, L.B. Kier, The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling, *Rev. Comput. Chem. II.*, K.B. Lipkowitz and D.B. Boyd, Eds, (1991), 367-422.
- ⁶⁸L.B. Kier, L.H. Hall, Molecular Connectivity Indices in Chemistry and Drug Research.. *Medicinal Chemistry*, ed. G. deStevens, New York, Academic Press, 14 (1976).
- ⁶⁹L.B. Kier, L.H. Hall, Molecular Connectivity in Structure-Activity Analysis.. *Chemometrics Series*, ed. D.D. Bawden, New York: Research Studies Press Ltd. 9 (1985).
- ⁷⁰A.R. Katritzky, E.V. Gordeeva, Traditional Topological Indices vs. Electronic, Geometrical, and Combined Molecular Descriptors in QSAR/QSPR Research., *J. Chem. Inf. Comput. Sci.*, 33 (1993), 835-857.
- ⁷¹D.T. Stanton, P.C. Jurs, Development and Use of Charge Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies., *Anal. Chem.*, 62 (1990), 2323-2329.
- ⁷²O. Clement, A. Mehl, In: O. F. Güner (Ed), *Pharmacophore Perception, Development, and Use in Drug Design-IUL Biotechnology Series*, International University Line, La Jolla, California, (2000), 71-84.
- ⁷³P.W. Sprague, R. Hoffmann, In *Computer Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry*; Van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; VHCA: Basel, Switzerland, (1997), 230-240.
- ⁷⁴D. Barnum, J. Greene, A. Smellie, P. Sprague, Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.*, 36 (1996), 563-571.
- ⁷⁵J. Singh, C.E. Chuaqui, P.A. Boriack-Sjodin, W.C. Lee, T. Pontz, M.J. Corbley, H.K. Cheung, R.M. Arduini, J.N. Mead, M.N. Newman, J.L. Papadatos, S. Bowes, S. Josiah, L.E. Ling, 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.*, 13 (2003), 4355-4359.
- ⁷⁶DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS. *Nucleic Acids Res.* (2011) Jan; 39 (Database issue):D1035-1041.