Ligand-based pharmacophore exploration and QSAR analysis of transition state analogues followed by in silico screening guide the discovery of new sub-micromolar β-secreatase inhibitors

Afaf Al-Nadaf & Mutasem O. Taha

Medicinal Chemistry Research

ISSN 1054-2523
Volume 22
Number 4

Med Chem Res (2013) 22:1979-1997
DOI 10.1007/s00044-012-0204-x

Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Springer

Author's personal copy

Med Chem Res (2013) 22:1979–1997
DOI 10.1007/s00044-012-0204-x

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

Ligand-based pharmacophore exploration and QSAR analysis of transition state analogues followed by in silico screening guide the discovery of new sub-micromolar β-secreatase inhibitors

Afaf Al-Nadaf · Mutasem O. Taha

Received: 28 January 2012 / Accepted: 10 August 2012 / Published online: 24 August 2012
© Springer Science+Business Media, LLC 2012

Abstract Transition state analogues (TSA)-based inhibitors have particularly potent inhibitory profiles. The fact that β-secretase (BACE) inhibitors have potential as anti-Alzheimer's treatments prompted us to explore the pharmacophoric space of 68 known TSA BACE inhibitors. Subsequently, quantitative structure–activity relationship (QSAR) analysis was employed to select optimal combination of binding model(s) and 2D physicochemical descriptors capable of explaining bioactivity variation. One pharmacophoric model emerged in the successful QSAR equation. However, to closely mimic the sterically demanding transition state of BACE we were obliged to complement the successful pharmacophore with strict shape-based query decorated with carefully positioned hydroxyl fragment that simulates the statin hydroxyl of known TSA BACE inhibitors. Both models, i.e., the hybrid shape–OH and optimal pharmacophore, were validated via receiver-operating characteristic curve analysis and were found to exhibit excellent abilities in discerning active compounds from decoys. Subsequent in silico screening against the National Cancer Institute structural database using the two models yielded two novel inhibitors of nanomolar and low micromolar IC_{50} values.

Keywords β-Secretase inhibitors · Transition state · Alzheimer's · Pharmacophore modeling · Quantitative structure–activity relationship

Electronic supplementary material The online version of this article (doi:10.1007/s00044-012-0204-x) contains supplementary material, which is available to authorized users.

A. Al-Nadaf
Department of Medicinal Chemistry and Pharmacognosy, Faculty of Pharmacy, Applied Science University, Amman, Jordan

M. O. Taha (✉)
Drug Discovery Unit, Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan
e-mail: mutasem@ju.edu.jo

Introduction

Alzheimer's disease (AD) is the most common cause of dementia in older people. The progression of AD involves the destruction of cells that control memory by the formation of neuritic plaques of aberrantly folded proteins in the brain (Maria *et al*., 2009). Neuritic plaques consist mainly of amyloid β-protein (Aβ): a 40–42-residue protein generated by cleavage of the extra-cellular domain of the transmembrane amyloid precursor protein (APP) is catalyzed by the proteases γ- and β-secretase (Shawn, 2009). This prompted significant recent interest in designing potent γ- and β-secretase inhibitors as potential therapeutic agents against AD (Cooper, 2002; Varghese, 2006).

Several classes of β-secreatase (BACE) inhibitors were recently reported (Wen-Hai *et al*., 2009; Cumming *et al*., 2008; Yoshio *et al*., 2008; Lorna *et al*., 2008; Yoshiari *et al*., 2008; Heuisul *et al*., 2008; Derek *et al*., 2008; Huang *et al*., 2006). More recently, several nanomolar transition state analogues (TSA) BACE inhibitors were reported (Heuisul *et al*., 2008; Clarke *et al*. 2008a, b; Beswick *et al*., 2008).

Clearly, the main focus of recent efforts toward the development of new BACE inhibitors concentrate on structure-based ligand design (Huang *et al*., 2006; Arun *et al*., 2001).

To date, several human BACE X-ray complexes are documented in the Protein Data Bank (Godemann *et al*., 2009; Charrier *et al*., 2009a, b, c; Clarke *et al*., 2008a, b; Beswick *et al*., 2008). However, although considered most reliable structural information for drug design, crystallographic structures are limited by inadequate resolution (Beeley and Sage,

Author's personal copy

2003) and crystallization-related artifacts of the ligand–protein complex (Klebe, 2006; Steuber *et al*., 2006; Stubbs *et al*., 2002). Moreover, crystallographic structures generally ignore structural heterogeneity related to protein anisotropic motion and discrete conformational substrates (DePristo *et al*., 2004)

The continued interest in designing new drug-like TSA BACE inhibitors, combined with the drawbacks of structure-based design and the significant-induced fit flexibility observed for BACE (Georgia *et al*., 2007) prompted us to explore the possibility of developing ligand-based TSA three-dimensional (3D) pharmacophore(s) integrated within self-consistent quantitative structure–activity relationship (QSAR) model. This approach avoids the pitfalls of structure-based techniques; furthermore, the pharmacophore model(s) can be used as 3D template(s) to search for new inhibitors. Although QSAR analysis is normally conducted to identify appropriate combination(s) of physicochemical or structural descriptors capable of explaining bioactivity variation across a group of bioactive compounds (Taha *et al*., 2008a), we implemented QSAR modeling herein as a competition arena to allow the selection of optimal pharmacophore(s) capable, upon combination with other 2D descriptors, of explaining bioactivity variation across a relatively long list of BACE inhibitors. We previously reported using this approach toward the discovery of new inhibitory leads against glycogen synthase kinase 3β (Taha *et al*., 2008b), hormone-sensitive lipase (Taha *et al*., 2008c), bacterial MurF (Taha *et al*., 2008a), protein tyrosine phosphatase 1B (Taha *et al*., 2007), influenza neuraminidase (Abu Hammad and Taha, 2009), cholesteryl ester transfer protein (Abu Khalaf *et al*., 2010), β-secretase (Al-Nadaf *et al*., 2010) renin (Al-Nadaf and Taha, 2011), cyclin-dependent kinase (Al-Sha'er and Taha, 2010), glycogen phosphorylase (Taha *et al*., 2011), β-glucosidase (Abu Khalaf *et al*., 2011), and β-galactosidase (Abdula *et al*., 2011).

However, TSAs resemble the substrate at its postulated transition to products; and therefore, TSA inhibitors require stringent steric and 3D provisions to dock into the enzymatic binding site during its sterically demanding high-energy transition state (TS). Therefore, TSA–enzyme complexes show pronounced sensitivity to slight misalignments among their complementary attractive groups (Schramm, 2003, 2005; Amyes and Richard, 2007). This conduct is expected to result in steep structure–activity surface, which limits the ability of the pharmacophore theory to explain activity/inactivity variations among training compounds. Pharmacophore modeling requires continuous bioactivity variation attributable to the presence or the absence of certain chemical features, i.e., smooth SAR surface. Hence, unveiling novel slow binding TSA inhibitors for BACE should complicate pharmacophore modeling.

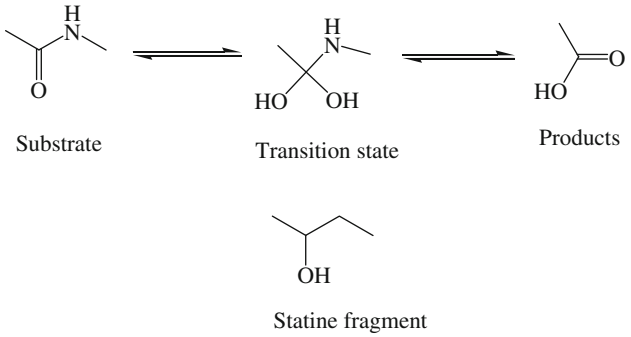


Fig. 1 Similarity between amide–hydrolysis transition state and the statin fragment of statin-based transition state analogues developed for inhibition of aspartic proteinases

The pronounced sensitivity of TSA to slight structural modifications should also complicate the subsequent use of pharmacophore models as 3D search queries to mine for new hits. Pharmacophore models would be too lax and, therefore, promiscuous in capturing TSAs as in silico hits, i.e., they may identify many inactive hits (false positives) (Catalyst User Guide, 2005). In fact, all previous pharmacophore modeling efforts to discover BACE inhibitors were reported for substrate analogues (Al-Nadaf *et al*., 2010; Huang *et al*., 2008). Accordingly, we were prompted to complement our QSAR-based pharmacophore models with tight shape constrains and to use the combination, i.e., optimal pharmacophore and shape mode, as consequent 3D search queries to mine for new TSA inhibitors. Furthermore, we introduced a central hydroxyl fragment into our shape query to mimic the catalytic intermediate derived by nucleophilic attack of a water molecule on the scissile peptide bond carbonyl, i.e., the statin hydroxyl seen in many BACE TSA inhibitions, shown in Fig. 1 (Rizzi *et al*., 2009; Clarke *et al*., 2008a, b; Beswick *et al*., 2008).

Materials and methods

Molecular modeling

Software and hardware

The following software packages were utilized in the present research.

- CATALYST (Version 4.11), Accelrys Inc. (www.accelrys.com), USA.
- CERIUS2 (Version 4.10), Accelrys Inc. (www.accelrys.com), USA.
- CS ChemDraw Ultra 6.0, Cambridge Soft Corp. (http://www.cambridgesoft.Com), USA.
- Pharmacophore and QSAR modeling studies were performed using CATALYST (HYPOGEN module)

Author's personal copy

and CERIUS2 software suites from Accelrys Inc. (San Diego, California, www.accelrys.com) installed on a Silicon Graphics Octane2 desktop workstation equipped with a dual 600 MHz MIPS R14000 processor (1.0 GB RAM) running the Irix 6.5 operating system. Structure drawing was performed employing ChemDraw Ultra 6.0 which was installed on a Pentium 4 PC.

Data set

The structures of 68 BACE-1 inhibitors (Fig. 2; Table A of Supplementary material) were collected from published literature (Clarke *et al*., 2008a, b; Beswick *et al*., 2008). The in vitro bioactivities of the collected inhibitors were expressed as the concentration of the test compound that inhibited the activity of BACE by 50 % (IC_{50}). Figure 2 and Table A of Supplementary material show the structures and IC_{50} values of the considered inhibitors. The logarithm of measured IC_{50} (μM) values were used in pharmacophore modeling and QSAR analysis, thus correlating the data linear to the free energy change.

In cases where IC_{50} is expressed as being higher than 500 μM (e.g., **2**, **3**, and **4**), it was assumed it equals 500 μM. This assumption 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 sketched using ChemDraw Ultra, installed on a PC, and saved in MDL-mol file format. Subsequently, they were imported into CATALYST, converted into corresponding standard 3D structures and energy minimized to the closest local minimum using the molecular mechanics CHARMm force field implemented in CATALYST. The resulting 3D structures were utilized as starting conformers for conformational analysis.

Conformational analysis

The molecular flexibilities of the collected compounds were taken into account by considering each compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a given energy range. Accordingly, the conformational space of each inhibitor (**1**–**68**, Fig. 2; Table A of Supplementary material) was explored adopting the "best conformer generation" option within CATALYST. Default parameters were employed in the conformation generation procedure, i.e., a conformational ensemble was generated with an energy threshold of 20 kcal/mol from the local minimized structure which has the lowest energy level and a maximum limit of 250 conformers per molecule (Catalyst User Guide, 2005).

Pharmacophoric hypotheses generation

All 68 molecules with their associated conformational models were regrouped into a spreadsheet. The biological data of the inhibitors were reported with an "Uncertainty" value of 3, which means that the actual bioactivity of a particular inhibitor is assumed to be situated somewhere in an interval ranging from one-third to three times the reported bioactivity value of that inhibitor (Kurogi and Güner, 2001; Li *et al*., 2000; Poptodorov *et al*., 2006). Typically, CATALYST requires informative training sets that include at least 16 compounds of evenly spread bioactivities over at least 3.5 logarithmic cycles. Lesser training lists could lead to chance correlation and thus faulty models (Kurogi and Güner, 2001; Li *et al*., 2000; Poptodorov *et al*., 2006). Three structurally diverse training subsets (Table B of Supplementary material) were carefully selected from the collected compounds for pharmacophore modeling.

Each training subset was utilized to conduct four modeling runs to explore the pharmacophoric space of BACE inhibitors. Different hypotheses were generated by altering the inter-feature spacing and the number of allowed features in the resulting pharmacophores (Table C of Supplementary material). See SM-1 of Supplementary material for details about CATALYST Pharmacophore Generation Algorithm (Kurogi and Güner, 2001; Li *et al*., 2000; Poptodorov *et al*., 2006).

Assessment of the generated hypotheses

When generating hypotheses, CATALYST attempts to minimize a cost function consisting of three terms: Weight cost, error cost, and configuration cost (see SM-2 of Supplementary material) (Catalyst User Guide, 2005; Sutter *et al*., 2000; Kurogi and Güner, 2001; Li *et al*., 2000; Poptodorov *et al*., 2006; Bersuker *et al*., 2000).

CATALYST-HYPOGEN implements additional approach to assess the quality of generated pharmacophores via the Cat-Scramble program implemented in CATALYST. This validation procedure is based on Fisher's randomization test (Poptodorov *et al*., 2006). In this validation test, a 95 % confidence level was selected, which instruct CATALYST to generate 19 random spreadsheets by the Cat-Scramble command. Subsequently, CATALYST-HYPOGEN is challenged to use these random spreadsheets to generate hypotheses using exactly the same features and parameters used in generating the initial unscrambled hypotheses. Success in generating pharmacophores of comparable cost criteria to those produced by the original unscrambled data reduces the confidence in the training compounds and the unscrambled original pharmacophore models. Table 1 shows the success criteria of best representative models generated during the pharmacophore exploration process.

Author's personal copy

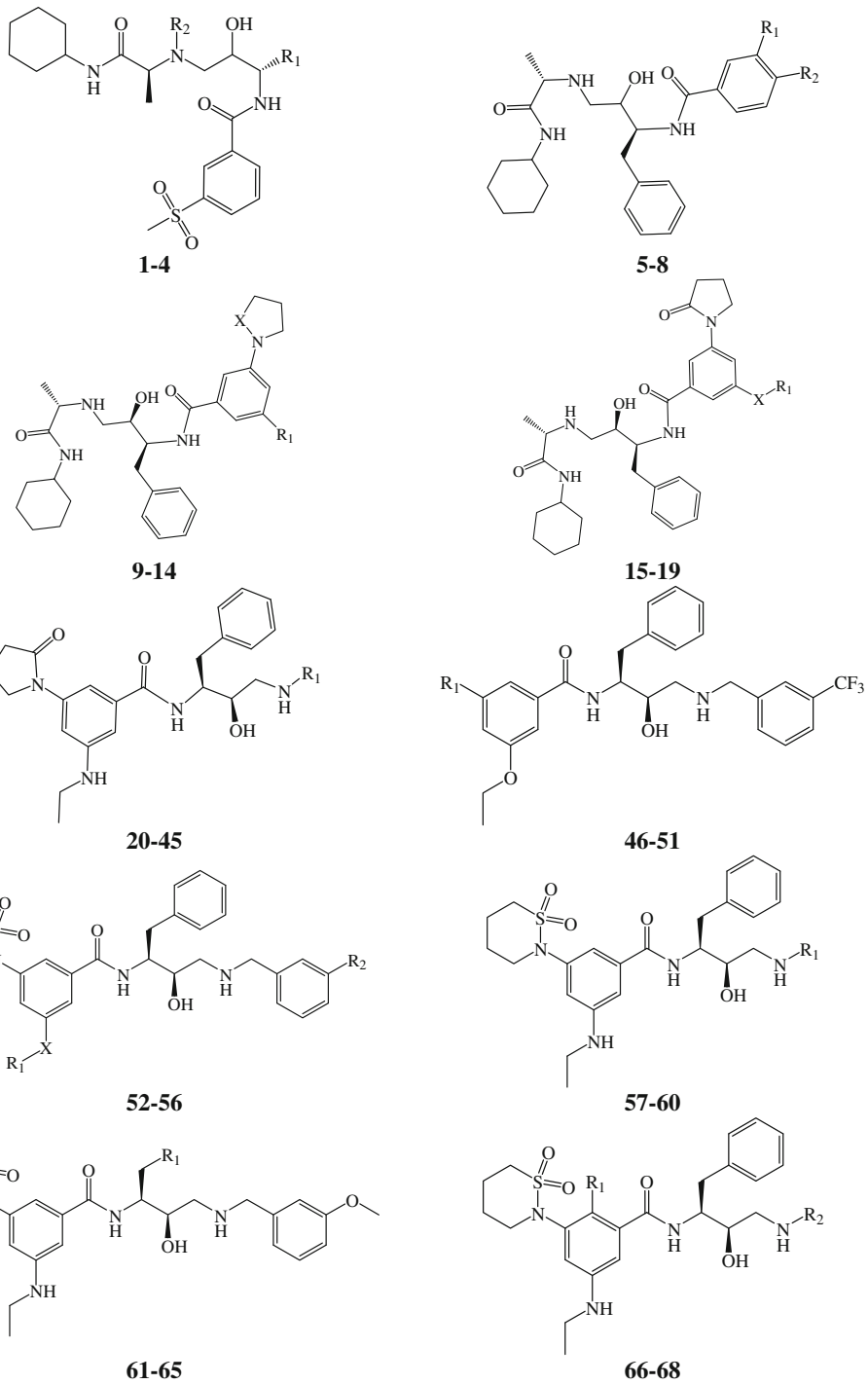


Fig. 2 The chemical scaffolds of training compounds, the corresponding structures and bioactivities are in Table A of supplementary material

Clustering of the generated pharmacophore hypotheses

The successful models (103 pharmacophores, see section "Exploration of BACE pharmacophoric space") were clustered into 21 groups utilizing the hierarchical average linkage method available in CATALYST. Therefore, closely related pharmacophores were grouped in 5-membered clusters. Subsequently, the highest-ranking representatives, as judged based on their fit-to-bioactivity correlation *F* values (calculated against collected compounds **1–68**, Fig. 2; Table A of Supplementary material), were selected to represent their corresponding clusters in subsequent QSAR modeling (see Table 1).

QSAR modeling

A subset of 55 compounds from the total list of inhibitors (**1–68**, Table A of Supplementary material) was utilized as

Author's personal copy

Table 1 The performance of the best representatives of clustered pharmacophore hypotheses generated for BACE TSA inhibitors

Training set	Run[a]	Hypotheses[b]	Pharmacophoric features in generated hypotheses	Total cost	Cost of null hypothesis	Residual cost[c]	R^2[d]	F-statistic[e]	Cat-Scramble (%)
A	1	2[f]	HBA, Hbic, 2× RingArom	85.3	118.2	32.9	0.93	37.4	95
		6	HBA, Hbic, PosIon, RingArom	87.6	118.2	30.6	0.92	47.1	95
		7	HBA, HBD, Hbic, RingArom	87.8	118.2	30.4	0.93	44.4	95
	2	3	HBD, Hbic, 2× RingArom	90.1	118.2	28.1	0.92	46.1	85
		10	2× Hbic, PosIon, RingArom	92.1	118.2	26.1	0.90	48.9	80
	3	2	2× HBA, 3× Hbic	83.4	118.2	34.8	0.93	46.0	95
		3	HBD, 2× Hbic, PosIon, RingArom	85.1	118.2	33.1	0.90	47.0	95
		4	HBA, 2× Hbic, PosIon, RingArom	85.4	118.2	32.8	0.93	47.0	95
B	5	4	HBD, Hbic, PosIon, RingArom	99.0	120.5	21.5	0.86	22.5	90
		6	2× Hbic, PosIon, RingArom	101.2	120.5	19.3	0.84	68.5	80
		9	HBD, Hbic, 2× RingArom	102.2	120.5	18.2	0.83	47.5	80
	7	1	HBD, 2× Hbic, PosIon, RingArom	95.0	120.5	25.5	0.93	35.9	95
		4	HBA, 2× Hbic, PosIon, RingArom	98.7	120.5	21.8	0.84	39.7	95
		6	2× HBA, 2× Hbic, PosIon	100.2	120.5	20.3	0.84	47.9	90
		7	HBA, 2× Hbic, PosIon, RingArom	100.3	120.5	20.1	0.85	55.4	90
	8	9	HBA, HBD, 2× Hbic, PosIon	94.5	120.5	26.0	0.89	9.4	90
C	9	10	3× HBA, Hbic	96.4	123.0	26.6	0.92	16.6	95
	10	3	Hbic, PosIon, 2× RingArom	90.7	123.0	32.3	0.94	46.5	95
		9	2× Hbic, PosIon, RingArom	95.3	123.0	27.7	0.89	46.4	95
		10	**HBA, HBD, Hbic, RingArom**	**95.3**	**123.0**	**27.7**	**0.91**	**54.5**	**95**
	11	3	HBA, HBD, 3× Hbic	90.5	123.0	32.5	0.91	35.0	85

Training subsets shown in Table B of Supplementary material; bolded pharmacophore appeared in the best QSAR equation

[a] Correspond to runs in Table C of Supplementary material

[b] High ranking representative BACE hypotheses (i.e., among their corresponding clusters of pharmacophores)

[c] Difference between total cost and the cost of the corresponding null hypotheses

[d] Correlation coefficients between bioactivity estimates and bioactivities of corresponding training compounds (subsets in Table B of Supplementary material)

[e] Fisher statistic calculated based on the linear regression between the fit values of collected inhibitors (**1**–**68**, Table A of Supplementary material; Fig. 2) against pharmacophore hypothesis (employing the "best fit" option and Eq. D of Supplementary material) and their respective anti-BACE bioactivities ($\log(1/IC_{50})$) values

[f] CATALYST ranking of each hypothesis in respective runs

a training set for QSAR modeling. However, as it is essential to access the predictive power of the resulting QSAR models on an external set of inhibitors, the remaining 13 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**–**68**, Fig. 2; Table A of Supplementary material) were ranked according to their IC_{50} values, then every fifth compound was selected for the test set starting from the high-potency end (Table A of Supplementary material). This selection considers the fact that the test molecules must represent a range of biological activities similar to that of the training set.

The logarithm of measured $1/IC_{50}$ (μM) values was used in QSAR, thus correlating the data linear to the free energy change. The chemical structures of the inhibitors were imported into CERIUS2 as standard 3D single conformer representations generated via the rule-based methods implemented in CERIUS2 (in SD format). Subsequently, different descriptor groups were calculated for each compound employing the C2.DESCRIPTOR module of CERIUS2. The calculated descriptors included various simple and valence connectivity indices, electro-topological state indices and other molecular descriptors (e.g., logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, energies of the lowest and highest occupied molecular orbitals, etc.) (CERIUS2 QSAR Users' Manual, 2005). Furthermore, the training compounds were fitted (using the Best-fit option in CATALYST) against the representative pharmacophores (21 models, Table 1), and their fit values were added as additional descriptors. The fit

Author's personal copy

value for any compound is obtained automatically via Eq. (D) of section SM-2 in supplementary material (Catalyst User Guide, 2005).

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, i.e., multiple linear regression modeling (MLR). The fitness function employed herein is based on Friedman's "lack-of-fit" (LOF) (CERIUS2 QSAR Users' Manual, 2005).

Our preliminary diagnostic trials suggested the following optimal GFA parameters: explore linear, quadratic, and spline equations at mating and mutation probabilities of 50 %; population size = 500; number of genetic iterations = 30,000, and LOF smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (QSAR descriptors), it was decided to scan and evaluate all possible QSAR models resulting from 3 to 11 explanatory terms.

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

$$r^2_{PRESS} = SD - PRESS/SD, \quad (1)$$

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. QSAR modeling selected Hypo10/10 as optimal pharmacophore (see section "QSAR modeling").

within BACE (PDB code: 2vj6, resolution 1.8 Å) (Clarke *et al*., 2008a) as template. However, in an attempt to generate the tightest possible shape query, several shape-tolerance values were evaluated. The tolerance range of 90–110 % was used for virtual screening as it allowed capture of reasonable number of hits (see section "Creation of shape query combined with OH fragment"). Furthermore, to further enhance the success rate of the overall virtual screen, we complemented our shape query with appropriately positioned statin hydroxyl fragment. This fragment corresponds to the position and direction of central hydroxyl of **15** bound within BACE complex, as in Fig. 4.

Receiver-operating characteristic (ROC) curve analysis

Both our QSAR-selected pharmacophore (Hypo10/10) and hydroxyl-complemented shape model were validated by assessing their abilities to selectively capture diverse BACE active compounds from a large testing list of actives and decoys.

The testing list was prepared as described by Verdonk *et al*. (2004; Kirchmair *et al*., 2008). Briefly, decoy compounds were selected based on three basic one-dimensional (1D) properties that allow the assessment of distance (D) between two molecules (e.g., i and j): (1) the number of hydrogen-bond donors (NumHBD); (2) number of hydrogen-bond acceptors (NumHBA); and (3) count of nonpolar atoms (NP, defined as the summation of Cl, F, Br, I, S, and C atoms in a particular molecule). For each active compound in the test set, the distance to the nearest other active compound is assessed by their Euclidean distance (Eq. 2)

$$D(i,j) = \sqrt{(\mathrm{NumHBD}_i - \mathrm{NumHBD}_j)^2 + (\mathrm{NumHBA}_i - \mathrm{NumHBA}_j)^2 + (\mathrm{NP}_i - \mathrm{NP}_j)^2} \quad (2)$$

Generation of shape and hydroxyl constrains

TSs are strained proteins states, and, therefore, are highly demanding vis-à-vis conformational and shape requirements of the corresponding binding ligands. Accordingly, we were prompted to develop shape-based constrains derived from the co-crystallized structure of potent TSA inhibitor **15** (Fig. 1; Table A of Supplementary material, IC_{50} 13 nM). These constrains were used as prefilter prior to virtual screening by Hyp10/10 to enrich the success rate.

The shape query was built using the Cat.Shape module within Catalyst (Catalyst User Guide, 2005) by implementing the crystallographic structure of **15** complexed

The minimum distances are then averaged over all active compounds (D_{min}). Subsequently, for each active compound in the test set, around 7 decoys were randomly chosen from the ZINC database (Irwin and Shoichet, 2005). These were selected in such a way that they occur within D_{min} from their corresponding active analogue.

To diversify active members in the list, we excluded any active compound having zero distance ($D(i, j)$) from other active compound(s) in the test set. Active testing compounds were defined as those possessing BACE affinities ranging from 3.0 to 72.0 nM. The test set included 10 active compounds and 71 ZINC decoys.

Author's personal copy

The test set (81 compounds) was screened by each particular pharmacophore employing the "Best flexible search" option implemented in CATALYST, while the conformational spaces of the compounds were generated employing the "Fast conformation generation option" implemented in CATALYST. Compounds missing one or more features were discarded from the hit list. In silico hits were scored employing their fit values as calculated by Eq. (D) of Supplementary material. Subsequently, hit lists were used to construct ROC curves for corresponding pharmacophores (see section SM-3 of Supplementary material for details on ROC analysis) (Verdonk *et al*., 2004; Kirchmair *et al*., 2008; Irwin and Shoichet, 2005; Triballeau *et al*., 2005; Jacobsson *et al*., 2003; Gao *et al*., 1999).

In silico screening for new BACE-1 inhibitors

The hydroxyl-complemented shape query was employed as preliminary 3D search query against the National Cancer Institute (NCI) database of compounds (238,819 compounds) using the "Best Flexible Database Search" option implemented within CATALYST to capture 27 hits. However, 18 of them were captured by subsequent screening against Hypo10/10. Only four hits were available from the NCI for subsequent in vitro testing.

In vitro experimental studies

Materials

Materials BACE-1 assay kit and dimethylsulfoxide (DMSO) were purchased from Sigma-Aldrich CS0010 and BDH Laboratory Supplies (England), respectively. NCI hits were kindly donated by the National Caner Institute.

Preparation of hit compounds for in vitro assay

The tested compounds were provided as dry powders. They were initially dissolved in DMSO to provide 0.02 mM stock solutions and subsequently diluted to the required concentrations using 50 mM sodium acetate, pH 4.5. The inhibition of BACE activity by the hit compounds was measured using the fluorometric assay. The final concentration of DMSO was adjusted to be <0.1 %.

Quantification of BACE-1 activity in a fluorometric assay

The BACE fluorescence resonance energy transfer (FRET) assay was performed as described by the manufacturer (Sigma, CS0010) (Sigma-Aldrich). Principle of the assay: the substrate is linked to a fluorescent dye on one end and to a quenching group on its other. The fluorescence of the substrate is significantly reduced due to intramolecular resonance energy transfer to the quenching group. Upon substrate cleavage by the enzyme, there is a disturbance of the energy transfer resulting in the enhancement of the fluorescent signal. The assay procedure can be described briefly as follows: the BACE substrate is prepared in the buffer to a concentration of 50 µM. BACE enzyme is prepared in the same buffer to a concentration of approximately 0.3 U/µl. Stock solutions of test samples are prepared in DMSO, and then serially diluted in the buffer to give the desired working concentrations. Triton X-100 was added to each well to a final concentration of 160 µM. BACE enzyme, substrate, standard, test samples, and buffer are then added to the wells for a total volume of 100 µl, with the BACE-1 enzyme being added last, just prior to reading. Baseline fluorescence is recorded immediately after the addition of the BACE enzyme on a fluorometer set at excitation 320 nm, emission 405 nm. The reaction rate was monitored for 2 h at 37 °C using FLX800TBI Microplate Fluorimeter (BioTek Instruments, Winooski, USA) and the linear time-relative fluorescence units (RFU) sections were taken for rate calculation (Al-Nadaf *et al*., 2010).

BACE inhibition by hit compounds

The inhibition of BACE activity by the hit compounds was measured using the fluorometric assay described above. The percentage of residual activity of BACE was determined for each compound by comparing the activity of BACE in the presence and absence of the tested compound. Blank and standard inhibitor (Lys-Thr-Glu-Glu-Ile-Ser-Glu-Val-Asn-Sta-Val-Ala-Glu-Phe) (Sigma-Aldrich, product (A1847)) was used as negative and positive controls, respectively. Measurements were conducted at least in duplicates.

Results and discussion

In this study, we employed CATALYST-HYPOGEN (Catalyst User Guide, 2005) to explore the pharmacophoric space of BACE TSA inhibitors utilizing 68 reported inhibitors. Three carefully selected training subsets were used to construct 103 possible binding hypotheses. Subsequently, genetic function algorithm (GFA) and MLR analyses were employed to search for optimal QSAR that combines high-quality binding pharmacophore with other molecular descriptors capable of explaining bioactivity variation across the whole collection BACE inhibitors. Subsequently, we employed the crystallographic structure of a co-crystallized potent TSA inhibitor (IC_{50} = 13 nM) complexed within BACE to generate tight shape constrains,

Author's personal copy

which we combined with a central hydroxyl fragment to simulate the statin hydroxyl in TSA inhibitors of BACE (Fig. 1). The resulting models, i.e., QSAR-selected pharmacophore and hydroxyl-decorated shape model, were used sequentially as 3D search queries to screen the NCI compounds database for new potential BACE inhibitors (Hahn, 1997; Moffat *et al*., 2008). The captured hits were subsequently experimentally validated against BACE.

CATALYST models drug–receptor interaction using information derived only from the drug structure (Catalyst User Guide, 2005; Sprague and Hoffmann, 1997; Smellie *et al*., 1995). HYPOGEN identifies a 3D array of a maximum of five chemical features common to active training molecules, which provides a relative alignment for each input molecule consistent with their binding to a proposed common receptor site. The chemical features considered can be hydrogen bond donors and acceptors (HBDs and HBAs), aliphatic and aromatic hydrophobes (Hbic), positive and negative charges, positive and negative ionizable groups (NegIon and PosIon), and aromatic planes (RingArom). The conformational flexibility of training ligands is modeled by creating multiple conformers, judiciously prepared to emphasize representative coverage over a specified energy range (Smellie *et al*., 1995; Sutter *et al*., 2000).

The SHAPE module in CATALYST is a shape-based similarity searching method. The van der Waals surface of a molecule (in certain conformation) is calculated and represented as a set of points of uniform average density on a grid. The surface points enclose a volume on the grid. The geometric center of the set of points is computed along with the three principal component vectors passing through the center. The maximum extents along each principal axis and the total volume are calculated. These provide shape indices that can be compared with the query and used in an initial screening step to eliminate poor matches from further consideration (Abu Hammad and Taha, 2009; Moffat *et al*., 2008; Smellie *et al*., 1995; Sutter *et al*., 2000; Singh *et al*., 2003).

In CATALYST-SHAPE, conformational flexibility is handled by pre-computing an ensemble of conformers for each library compound and comparing each conformer with the query shape indices (Moffat *et al*., 2008). CATALYST pharmacophores, with or without shape constraints, have been used as 3D queries for database searching and in 3D-QSAR studies (Moffat *et al*., 2008; Smellie *et al*., 1995; Sutter *et al*., 2000; Singh *et al*., 2003).

Data mining and conformational coverage

The literature was extensively surveyed to collect large group of diverse TSA BACE inhibitors (**1**–**68**, Fig. 2; Table A of Supplementary material) (Clarke *et al*. 2008a, b; Beswick *et al*., 2008). The 2D structures of the training inhibitors were imported into CATALYST and converted automatically into plausible 3D single conformer representations via the rule-based methods implemented within the package. The resulting single conformer 3D structures were later used as starting points for conformational analysis and in the determination of various molecular descriptors for QSAR modeling.

The conformational space of each inhibitor was extensively sampled utilizing the poling algorithm employed within CATALYST. Poling promotes conformational variation via employing molecular mechanical force field algorithm that penalizes similar conformers (Sutter *et al*., 2000). Conformational coverage was performed employing the "Best" module to ensure extensive sampling of conformational space to guarantee minimal conformation-related noise during pharmacophore generation and validation stages. Inadequate conformational sampling within the training compounds is known to affect pharmacophore modeling and pharmacophore-based in silico search procedures (Sheridan and Kearsley, 2002).

Exploration of BACE pharmacophoric space

CATALYST-HYPOGEN enables automatic pharmacophore construction using a collection of molecules with bioactivities spanning over several orders of magnitude (Catalyst User Guide, 2005; Sutter *et al*., 2000; Kurogi and Güner, 2001; Li *et al*., 2000; Poptodorov *et al*., 2006; Bersuker *et al*., 2000). Accordingly, as we collected a list of 68 diverse TSA BACE inhibitors of evenly spread bioactivities over more than 4 log cycles, we were prompted to employ HYPOGEN to identify possible pharmacophoric-binding modes assumed by different TSA inhibitors within BACE.

However, HYPOGEN implements an optimization algorithm that evaluates large number of potential models for a particular target through fine perturbations to hypotheses that survived the subtractive and constructive phases (see sections "Pharmacophoric hypotheses generation" and SM-1 of Supplementary material) (Li *et al*., 2000). The extent of the evaluated space is reflected by the configuration (Config.) cost calculated for each modeling run. It is generally recommended that the Config.cost of any HYPOGEN run not to exceed 17 (corresponding to 2^{17} hypotheses to be assessed by CATALYST), to guarantee thorough analysis of all models (Sutter *et al*., 2000).

The size of the investigated pharmacophoric space is a function of training compounds, selected input chemical features and other CATALYST control parameters. Restricting the extent of explored pharmacophoric space should improve the efficiency of optimization via allowing

Author's personal copy

effective evaluation of limited number of pharmacophoric models. On the other hand, extensive restrictions imposed on the pharmacophoric space might reduce the possibility of discovering optimal pharmacophoric hypothesis, as they might occur outside the "boundaries" of the pharmacophoric space. Accordingly, a balanced exploration approach is necessary to unveil optimal pharmacophores.

Moreover, the fact that pharmacophore modeling requires limited number of carefully selected training compounds (from **16** to **45** compounds only) (Catalyst User Guide, 2005) that exhibit bioactivity variations attributable solely to the presence or the absence of pharmacophoric features, i.e., not due to steric or electronic factors, makes it impossible to explore the pharmacophore space of large training sets in one shot (e.g., 68 compounds), partly because CATALYST-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 large SAR data. This dilemma prompted us to break our collected compounds (68) into smaller training subsets compatible with pharmacophore modeling, i.e., of bioactivity variations attributable solely to the presence or the absence of pharmacophoric features (3D SAR). Nevertheless, the basic problem in 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 relatively large SAR lists, as in our case. We found that the best way to solve this problem is by exploring the pharmacophoric space of several carefully selected training subsets, i.e., selected from the whole list of collected compounds, followed by allowing the resulting pharmacophores (originating from all training subsets) to compete within the context of GFA-QSAR analysis such that the best pharmacophore(s) that is(are) capable of explaining bioactivity variations across the whole list of collected compounds is(are) selected (Taha *et al*., 2007, 2008a, b, c, 2011; Abu Hammad and Taha, 2009; Abu Khalaf *et al*., 2010, 2011; Al-Nadaf *et al*., 2010; Al-Nadaf and Taha, 2011; Al-Sha'er and Taha, 2010; Abdula *et al*., 2011).

Accordingly, we explored the pharmacophoric space of collected BACE inhibitors under reasonably imposed "boundaries" through 12 HYPOGEN automatic runs employing three carefully selected training subsets (i.e., from the collected compounds): subsets **A**, **B**, and **C** (Table B of Supplementary material). The training compounds in these subsets were selected in such away to guarantee maximal 3D diversity and continuous bioactivity spread over more than 3.5 logarithmic cycles. Moreover, training subsets were selected in such a way that their member compounds apparently share certain 3D SAR rules (by visual evaluation) and their differences in anti-BACE bioactivities are primarily attributable to the presence or the absence of pharmacophoric features rather than steric shielding and/or bioactivity-enhancing or bioactivity-reducing auxiliary groups (e.g., electron donating or withdrawing groups).

Guided by our reasonably restricted pharmacophoric exploration concept, the software was restricted to explore pharmacophoric models, of each training subset, having from zero to three features of any particular selected feature type (i.e., HBA, HBD, hydrophobic, and Ring Aromatic), i.e., instead of the default range of zero to five. Furthermore, to further limit the investigated pharmacophoric space, only four- and five-featured pharmacophores were explored (Table C of Supplementary material). Three- and two-featured pharmacophores are rather promiscuous as 3D search queries and probably not adequate descriptions of ligand-BACE binding as judged from the structural diversity of the training compounds.

In each run, the resulting binding hypotheses were automatically ranked according to their corresponding "total cost" value: defined as the sum of error cost, weight cost, and configuration cost (see sections "Assessment of the generated hypotheses" under Experimental and SM-2 of Supplementary material). Error cost provides the highest contribution to total cost and it is directly related to the capacity of the particular pharmacophore as 3D-QSAR model, i.e., in correlating the molecular structures to the corresponding biological responses. HYPOGEN also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost (residual cost, Table 1), the more likely that the hypothesis does not reflect a chance correlation (Catalyst User Guide, 2005; Sutter *et al*., 2000; Kurogi and Güner, 2001; Li *et al*., 2000; Poptodorov *et al*., 2006; Bersuker *et al*., 2000).

CATALYST includes an additional validation technique known as Cat.Scramble (Catalyst User Guide, 2005). This procedure is based on Fisher's randomization test (Fisher, 1966). In this test, the biological data and the corresponding structures are scrambled several times, and the software is challenged to generate pharmacophoric models from the randomized data. The confidence in the parent hypothesis (i.e., generated from unscrambled data) is lowered proportional to the number of times the software succeeds in generating binding hypothesis from scrambled data of apparently better cost criteria than the parent hypothesis (see section **"**Assessment of the generated hypotheses" under Experimental).

Eventually, 120 pharmacophore models emerged from 12 automatic HYPOGEN runs performed on three training subsets (Table C of Supplementary material), out of which

Author's personal copy

103 models illustrated Cat.Scramble confidence levels of ≥85 %. These were clustered (see section "Clustering of the Generated Pharmacophore Hypotheses") and the best representatives (21 models) were used in subsequent QSAR modeling (as in Table 1).

QSAR modeling

The fact that pharmacophore modeling of BACE inhibitors furnished several binding hypothesis of comparable statistical criteria (Table 1) prompted us to employ classical QSAR analysis to search for the best pharmacophore(s) capable of explaining bioactivity variation across the whole list of collected inhibitors (**1**–**68**, Table A of supplementary material). GFA and MLR QSAR (GFA-MLR-QSAR) analysis were employed to search for an optimal QSAR equation(s).

GFA-MLR-QSAR selects optimal descriptor combinations based on the Darwinian concept of genetic evolution whereby the statistical criteria of regression models from different descriptor combinations (chromosomes) are employed as fitness criteria (Irwin and Shoichet, 2005). GFA-MLR-QSAR analysis was employed to explore various combinations of pharmacophores and other structural descriptors and to evaluate their statistical properties as predictive QSAR models.

The fit values obtained by mapping the 21 representative hypotheses against collected BACE-1 inhibitors were enrolled, together with around 250 physicochemical descriptors, as independent variables (genes) in a cycle of GFA-MLR-QSAR analysis over 30,000 iterations employing Friedman's LOF fitness criterion (see section "QSAR Modeling" under Experimental) (CERIUS2 QSAR Users' Manual, 2005; Ramsey and Schafer, 1997). The complementary physicochemical descriptors were generally independent of the 3D conformation/alignment of modeled compounds, and therefore, we implemented standard rule-based generated 3D conformers for QSAR modeling without any particular alignment rule. However, since it is essential to access the predictive power of the resulting QSAR models on an external set of inhibitors, we randomly selected 13 molecules and employed them as external test molecules for validating the QSAR models (marked with asterisk in Table A of Supplementary material, see section "QSAR Modeling" under Experimental). Moreover, all QSAR models were cross-validated automatically using the leave-one-out cross-validation in CERIUS2 (CERIUS2 QSAR Users' Manual, 2005; Ramsey and Schafer, 1997).

Equation (3) shows the details of the optimal QSAR model. Figure 3 shows the corresponding scatter plots of experimental versus estimated bioactivities for the training and testing inhibitors

$$\begin{aligned}\mathrm{Log}(1/(\mathrm{IC}_{50})) &= -40.243 + 2.69(\mathrm{Hypo10/10}) \\ &\quad + 4.71 \times 10^{-4}(\mathrm{Apol}) - 0.25(\mathrm{SaasC}) \\ r^2_{55} &= 0.77, \quad F = 56.22, \quad r^2_{\mathrm{BS}} = 0.77, \quad r^2_{\mathrm{LOO(13)}} = 0.74, \\ r^2_{\mathrm{PRESS(13)}} &= 0.55, \end{aligned} \tag{3}$$

where r^2_{55} is the correlation coefficient of the QSAR equation against 55 training compounds, F is Fisher statistical parameter, r^2_{BS} is the boot strapping regression coefficient, r^2_{LOO} is the leave-one-out correlation coefficient, and $r^2_{\mathrm{PRESS(13)}}$ is the predictive r^2 determined for 13 randomly selected test compounds (CERIUS2 QSAR Users' Manual, 2005; Ramsey and Schafer, 1997).

Apol is an electronic descriptor corresponding to the sum of atomic polarizabilities within a particular molecule. SaasC is the elctrotopological state index descriptor for aromatic carbon atoms. Hypo10/10 represent the fit values of the training compounds against the 10th pharmacophore

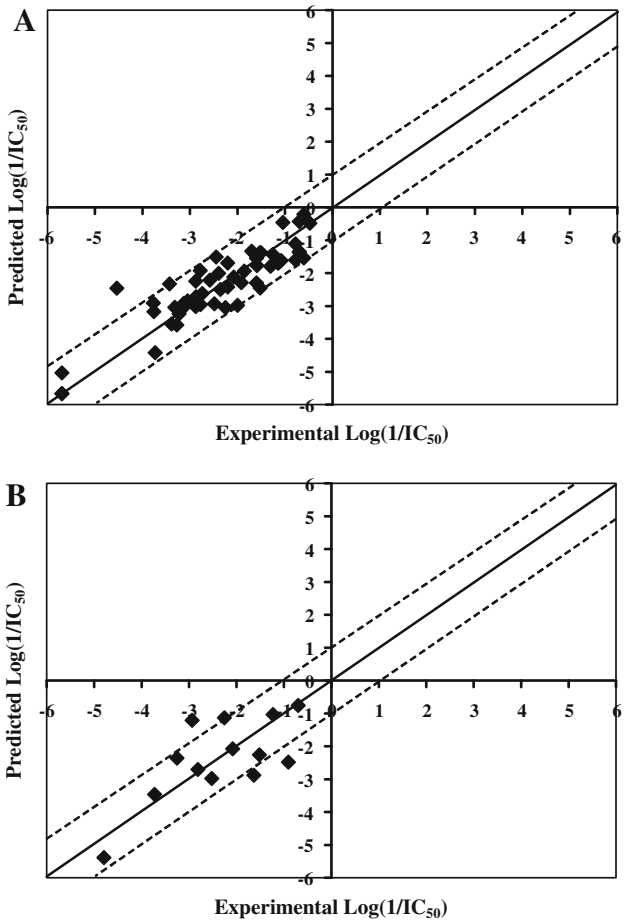


Fig. 3 Experimental versus fitted (**a** 55 compounds, $r^2_{\mathrm{LOO}} = 0.74$) and predicted (**b** 13 compounds, $r^2_{\mathrm{PRESS}} = 0.55$) bioactivities calculated from the best QSAR model (Eq. 3). The *solid lines* are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the *dotted lines* indicate the 1.0 log point error margins

Author's personal copy

model generated in the 10th modeling run, as in Table 1 (and Table C of Supplementary material) (CERIUS2 QSAR Users' Manual, 2005). Figures 4 and 5 show Hypo10/10 pharmacophore and how it maps two of the most potent TSA BACE inhibitors: **15** (IC_{50} = 13 nM) and **33** (IC_{50} = 17 nM), respectively, while Table 2 shows the X, Y, and Z coordinates of the best pharmacophore.

Judging from the regression coefficients of different descriptors in Eq (3) and the ranges of their values across all training compounds (see Table D of Supplementary material), one can clearly see that Hypo10/10 is the most prominent contributor to bioactivity with a maximum enhancement of 29.03 log cycles to bioactivity (average bioactivity enhancement of 28.50 log cycles). On the other hand, the analysis shows that SaasC is the least influential with a maximum reduction to bioactivity of 1.4 log cycles (average bioactivity reduction of 0.73 log cycles). On the other hand, Apol has moderate contributions to bioactivity as reflected in its maximal bioactivity enhancement of 11.29 log cycles (average bioactivity enhancement of 10.07). Accordingly, one can safely conclude from this analysis that pharmacophoric recognition and complementarities among ligand–receptor-binding features are the dominant contributors to bioactivity, while molecular polarizability comes next. On the other hand, the presence of aromatic rings (i.e., SaasC) seems to have marginal detrimental effect on bioactivity. Apparently, the presence of polarizable atoms, e.g., sulfur, halogens, and aromatic rings, promote ligand–receptor affinity via promoting van der Waals' interactions with prolarizable groups within the binding pocket, e.g., the aromatic side chains of TRP176, PHE169, TYR132, and TYR259 in the binding

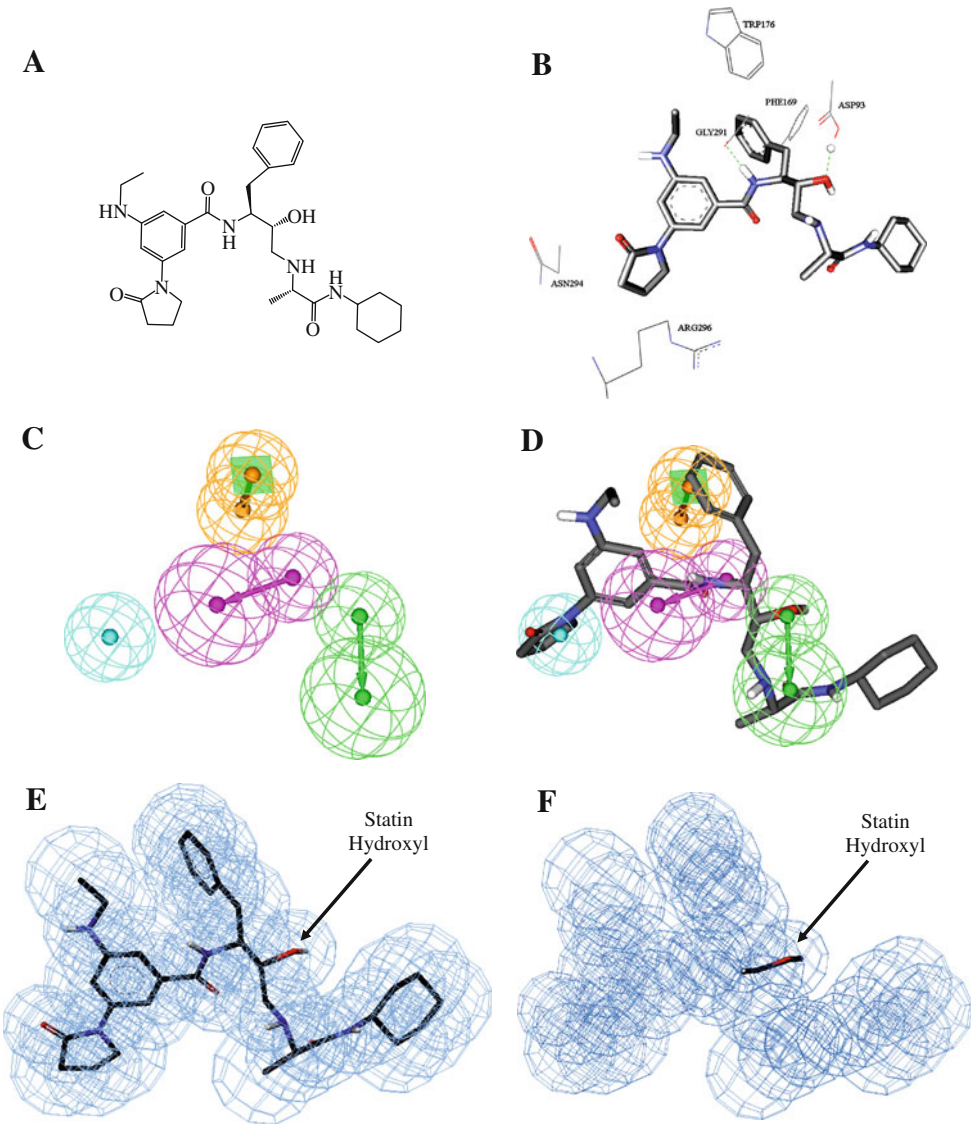


Fig. 4 **a** Chemical structure of **15**, **b** co-crystal structure of **15** (IC_{50} = 13 nM, Table A of supplementary material) complexed within BACE (PDB code: 2VJ6, resolution 1.8 Å) (Clarke *et al*., 2008a). **c** Pharmacophoric features of Hypo10/10: *HBD* hydrogen bond donor as *pink* vectored spheres, *HBA* hydrogen bond acceptor as *green* vectored spheres, *Hbic* hydrophobic features as *blue* spheres, *RingArom* ring aromatic as vectored *orange* spheres, **d** Hypo10/10 fitted against **15** employing rigid mapping (i.e., without allowing conformational flexibility), **e** Shape constrains patterned against the co-crystallized complexed structure of **15** within BACE, **f** Shape constrains merged with the statin hydroxyl fragment of **15**. This model was used as prefilter prior to virtual screening using Hypo10/10 (Color figure online)

Author's personal copy

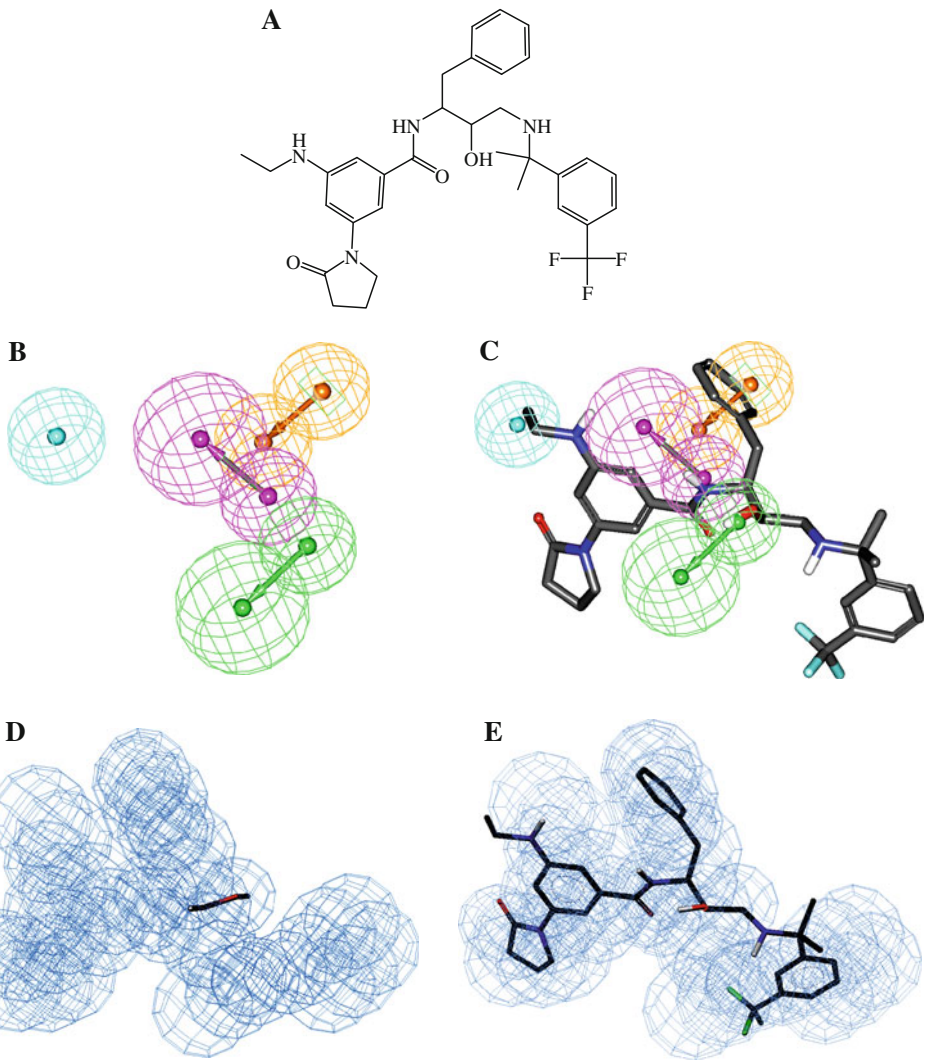


Fig. 5 **a** Chemical structure of training compound **33** (IC_{50} = 17 nM), **b** Hypo10/10, **c** Hypo10/10 mapped against training compound **33**, **d** hybrid shape–OH model, and **e** merged shape–OH model mapped against **33** (Color figure online)

Table 2 Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of Hypo10/10

Definitions	Chemical features						
	HBA		HBD		Hbic	RingArom	
Weights	2.726		2.726		2.726	2.726	
Tolerances	1.60	2.20	1.60	2.20	1.60	1.60	1.60
Coordinates							
X	2.26	3.78	−0.82	−2.05	−5.44	−3.72	−4.02
Y	−2.27	0.34	−1.45	−0.35	4.26	−4.24	−1.65
Z	−1.54	−1.78	−1.01	−3.56	−2.70	−1.37	0.11

pocket (Fig. 4 and PDB entry 2VJ6 (Clarke *et al*. 2008a). However, it seems that aromatic rings marginally undermine bioactivity due to their steric sizes that clash with the tight steric requirements of the binding pocket during TS.

Comparison of Hypo10/10 with the active site of BACE

To emphasize the validity of our pharmacophore/QSAR modeling approach, we compared Hypo10/10 with the crystallographic structure of BACE-binding pocket complexed to **15** (IC_{50} = 13.0 nM, PDB code: 2VJ6 (Clarke *et al*. 2008a). Figure 4 shows the chemical structure of the ligand and compares its BACE complex (Fig. 4b) with the way it maps Hypo10/10 (Fig. 4d) employing rigid mapping, i.e., fitting the ligand's bound state against the pharmacophore without conformational adjustments.

A marked similarity was observed between the features proposed by the pharmacophore model and the ligand-binding features in crystallographic structure, as can be clearly seen in Fig. 4.

Mapping the benzyl aromatic ring of **15** against RingArom feature in Hypo10/10 (Fig. 4c) corresponds to stacking against the indole and phenyl side chain of

Author's personal copy

TRP176 and PHE169, respectively (Fig. 5b), in the complex. While mapping the central amidic NH in **15** against HBD in Hypo10/10 (Fig. 4c) corresponds to hydrogen bond interactions tying this NH with the peptidic carbonyl of GLY291 in the crystallographic complex (Fig. 4b). Furthermore, fitting the statin hydroxyl of **15** against HBA in Hypo10/10 (Fig. 4c) matches hydrogen-bonding interactions tying the oxygen of the statin hydroxyl in **15** with the carboxylic OH of ASP93 (Fig. 4b). Finally, mapping the pyrrolidinone of **15** against an Hbic feature in Hypo10/10 (Fig. 4c) seems to agree with squeezing this fragment within a hydrophobic pouch comprised of the aliphatic linkers of the side chains of ASN294 ARG296 (Fig. 4b).

Creation of shape query combined with OH fragment

To enrich the probability of active hits among captured compounds, it was necessary to guarantee the tightest 3D analogy between hit compounds and sterically demanding TS inhibitors. Therefore, it was decided to use shape-based constraints derived from a potent training inhibitor as virtual prefilter prior to virtual screening via Hypo10/10. Shape constraints encode for the extent of 3D spatial similarity between screened compounds and the template molecule used to build shape limitations (Hahn, 1997).

To generate the shape query, we selected the co-crystallized conformation of the TSA inhibitor **15** complexed within BACE (PDB entry 2vj6) (Clarke *et al*., 2008a) as template. We converted the template molecule into corresponding shape constrains employing Cat.SHAPE module implemented within CATALYST suite (see section "Generation of shape and hydroxyl constrains"). However, to create the tightest possible shape query, we evaluated several shape tolerance values (Catalyst User Guide, 2005). Tolerance values that allowed the capture of reasonable number of hits were used for virtual screening. A range of 300–400 hits was considered reasonable as it is manageable for possible subsequent in vitro screening and large enough to allow satisfactory diversity within the hit list. A tolerance range of 90–110 % was found to capture optimal number of hits (i.e., 371 compounds).

However, to further limit our prefiltration search hits, we complemented our shape prefilter model with hydroxyl fragment corresponding to the statin central hydroxyl group of **15** (see Fig. 1). This should restrict captured compounds to those possessing central hydroxyl moieties necessary for TSA recognition and binding within BACE. Incorporation of the statin hydroxyl group within the shape model reduced the number of captured hits to 27. Figures 4 and 5 show the hydroxyl complemented shape model and how it fits compounds **15** and **33**, respectively.

ROC curve analysis

To validate the resulting models (QSAR, pharmacophore and hydroxyl-complemented shape query), we subjected our merged shape–hydroxyl query and Hypo10/10 to ROC analysis. In ROC analysis, the ability of a particular 3D search query (e.g., pharmacophore or shape model) to correctly classify a list of compounds as actives or inactives is indicated by the area under the curve (AUC) of the corresponding ROC as well as other parameters: overall accuracy, overall specificity, overall true positive rate (TPR), and overall false negative rate (see sections "ROC curve analysis" under Experimental and SM-3 ROC analysis of Supplementary material for more details) (Verdonk *et al*., 2004; Kirchmair *et al*., 2008; Irwin and Shoichet, 2005; Triballeau *et al*., 2005; Jacobsson *et al*. 2003; Gao *et al*., 1999).

Table 3 shows the ROC results of Hypo10/10 and the shape–OH hybrid model. Both the models illustrated excellent overall performances with ROC-AUC values of 100 %. However, Hypo10/10 outperformed the shape–OH model by capturing all seeded active compounds in the testing list (i.e., with decoys) and, therefore, achieving a TPR of 100 %, while the hybrid shape–OH query scored a TPR of 60 % because it excluded 4 out of 10 actives. On the other hand, the shape–OH query excelled by excluding all inactive decoys compared to Hypo10/10, which captured 57 out of 71 decoy compounds in the testing list. This doubled the false negative rate score of Hypo10/10 compared to the shape–OH model (14.1 and 8.5 %, respectively).

Overall, ROC statistics of both the models suggest that they complement each other as sequential search queries, i.e., the shape–OH prefilter should exclude most inactive compounds, while Hypo10/10 preferentially capture active compounds.

In silico screening of databases

In silico screening was conducted first employing the shape–OH prefilter against the NCI list of compounds (includes 238,819 compounds). It captured 27 hits (Catalyst User Guide, 2005). Subsequently, Hypo10/10 was used and short-listed them into 18 hits only. Hits are defined as

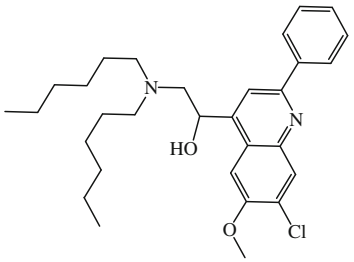
Table 3 ROC curve analysis criteria for Hypo10/10 and the merged shape–OH query

Pharmacophore model	ROC-AUC (%)	ACC (%)	SPC (%)	TPR (%)	FNR (%)
Hypo10/10	100	87.6	85.9	100	14.1
Shape–OH hybrid	100	87.6	91.6	60	8.5

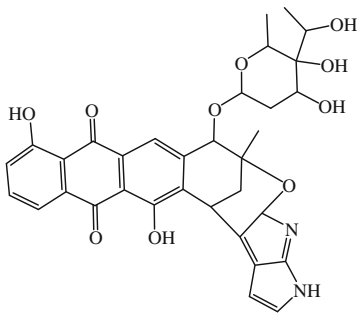
ROC receiver operating characteristic curve, *AUC* area under the curve, *ACC* overall accuracy, *SPC* overall specificity, *TPR* overall true positive rate, *FNR* overall false negative rate

Author's personal copy

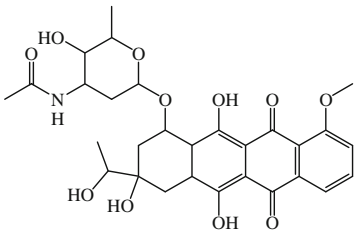
Fig. 6 The chemical structures of the captured hits (fit Hypo10/10 with complemented shape–OH fragment)



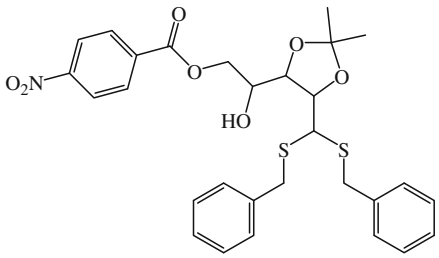
69 (NSC-13306)



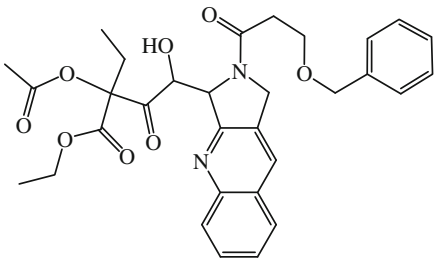
70 (NSC-194617)



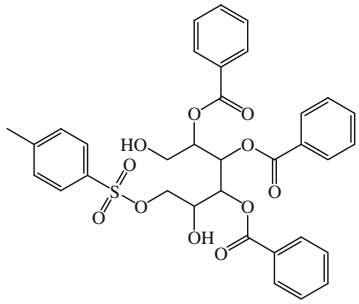
71 (NSC-277812)



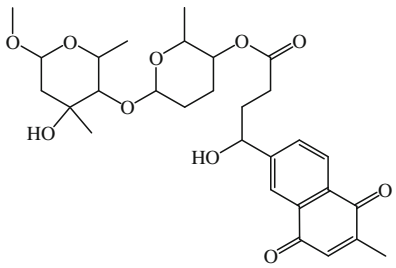
72 (NSC-109168)



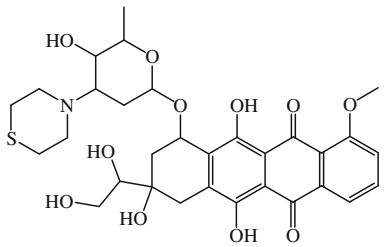
73 (NSC-141558)



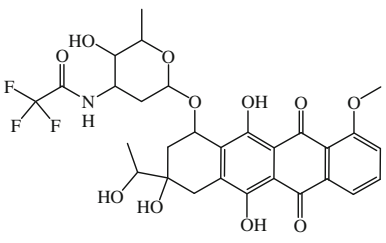
74 (NSC-231895)



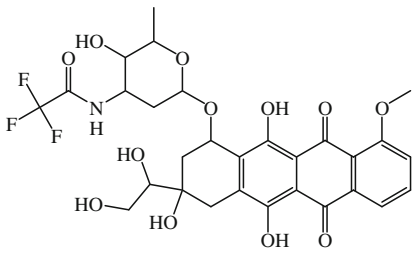
75 (NSC-220970)



76 (NSC-376755)



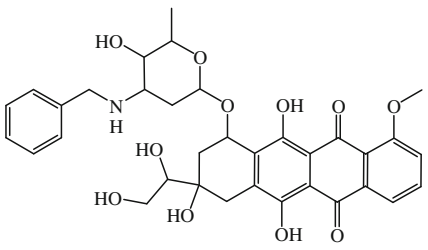
77 (NSC-263466)



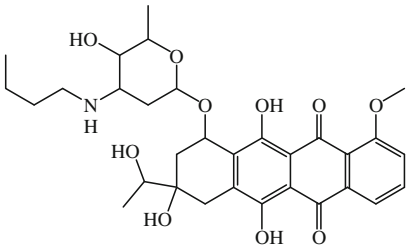
78 (NSC-268711)

Author's personal copy

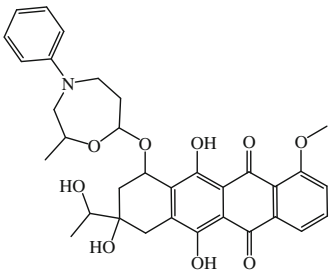
Fig. 6 continued



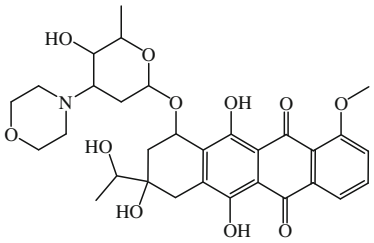
79 (NSC-269434)



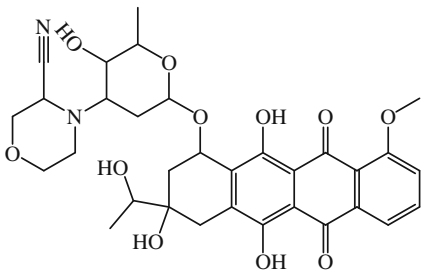
80 (NSC-269742)



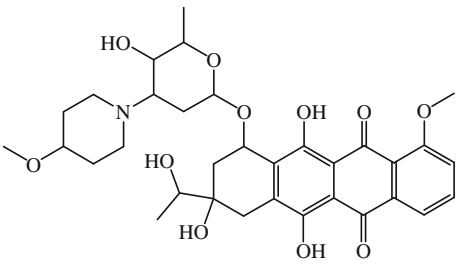
81 (NSC-304945)



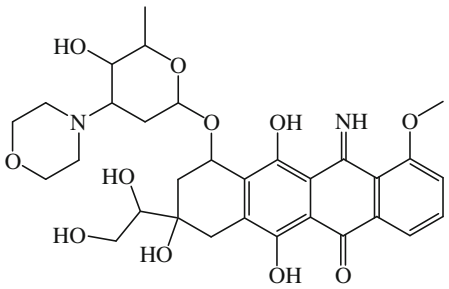
82 (NSC-327450)



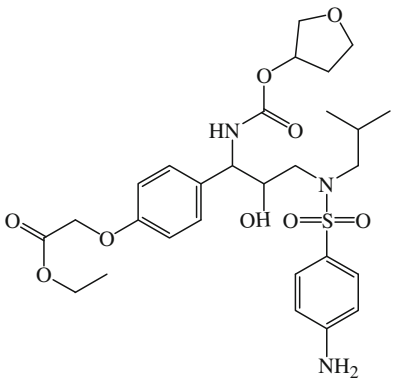
83 (NSC-332305)



84 (NSC-334354)



85 (NSC-355465)



86 (NSC-699058)

those compounds that have their chemical moieties spatially overlap (map) with corresponding features in the pharmacophoric query and fit the shape–OH prefilter. Figure 6 shows the chemical structures of the final list of captured hits. Unfortunately, only four hits were readily available from the NCI and, therefore, were tested in vitro against BACE, namely, **69**, **70**, **71**, and **72** (Fig. 6).

However, only two hits: **69** (NSC13306) and **70** (NSC194617) inhibited BACE by >50 % at 50 μM concentrations prompting us to evaluate their anti-BACE profiles at other concentrations. The two hits illustrated IC_{50} values of 209 nM and 5.62 μM, respectively. Figure 7 shows the inhibition curves of **69** and **70** against BACE. However, judging from the inhibitory profiles values of the two

Author's personal copy

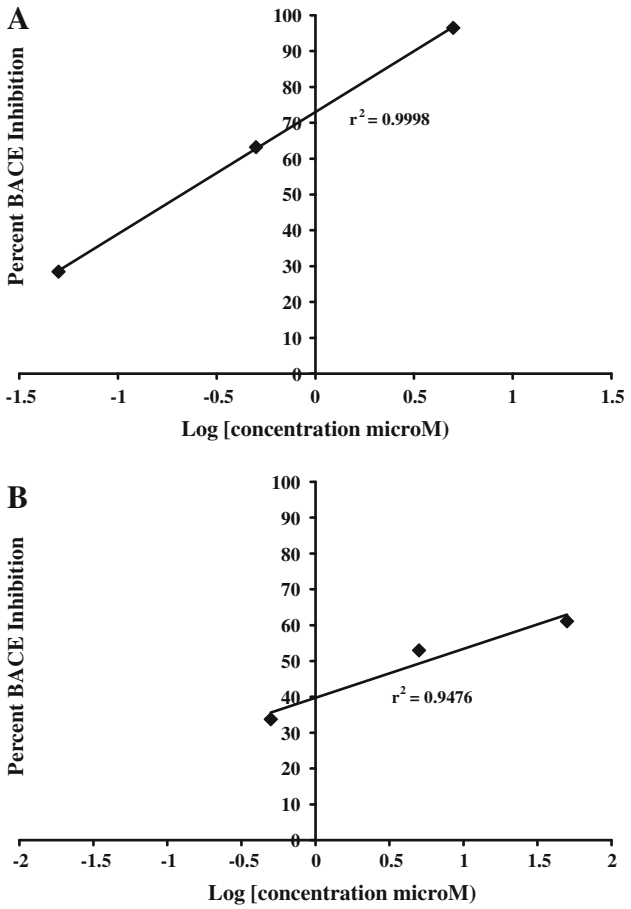


Fig. 7 Dose-inhibition curves of hit compounds **a 69** (IC_{50} = 209 nM) and **b 70** (IC_{50} = 5.62 μM)

compounds, we believe only **69** is actually a TSA inhibitor of BACE as it illustrated nanomolar IC_{50}. We believe the strict steric requirements of BACE TS are responsible for the inactivity of hits **71** and **72** and for the relatively weak activity of **70** (i.e., micromolar). Figure 8 shows how **69** fits the shape–OH hybrid model and Hypo10/0.

Anyway, both the active hits have completely novel scaffolds and can, therefore, be used as good leads for subsequent optimization.

Similarity analysis between training compounds and active hits

We employed three library comparison methods implemented in Discovery Studio 2.5 to assess the structural similarity/diversity between the modeled compounds (**1**–**68**, Fig. 2, library A; Table A of Supplementary material) (Clarke *et al*. 2008a, b; Beswick *et al*., 2008) compared to captured hits (**69**–**86** as in Fig. 6, library B), namely, Murcko Assemblies, Bayesian Model and Global Fingerprints. The comparison clearly establishes the uniqueness of captured hits.

In Murcko Assemblies, the algorithm breaks the ligands of each library into unique occurrences of molecular rings, ring assemblies, bridge assemblies, chains, Murcko assemblies, or any combination of these. Murcko assemblies are contiguous ring systems plus chains that link two or more rings (Bemis and Murcko, 1996). The two libraries are compared using a Tanimoto similarity of the assemblies based on the fragments that are common and unique to each library (Discovery Studio 2.5.5 User Guide, 2010).

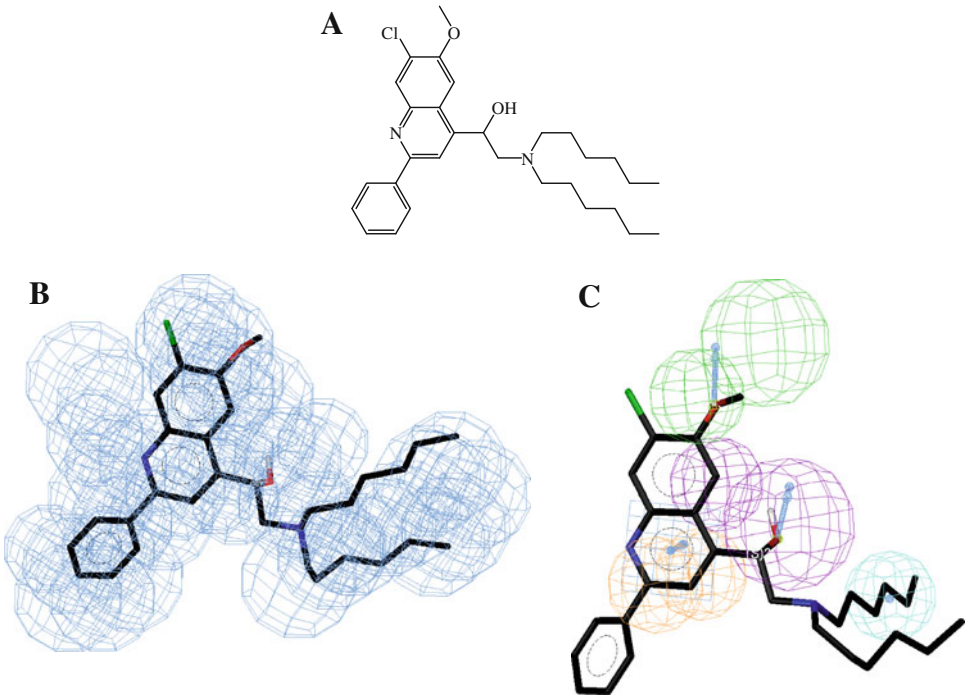


Fig. 8 **a** Chemical structure of hit compound **69** (NCI code: 13306, IC_{50} = 209 nM), **B 69** fitted against the shape–OH hybrid model, and **c 69** fitted against Hypo10/10 (Color figure online)

Author's personal copy

Table 4 Results of similarity analysis between training compounds and NCI with OH fragment hits

Murcko assemblies[a]		Bayesian model[b]		Global fingerprints[b,c]	
Number of total assemblies	41	Average LibA score of library A ligands	12.49	Number of total global fingerprint bits	888
Number of common assemblies	1	Average LibB score of library A ligands	−112.82	Number of common global fingerprint bits	80
Number of assemblies only in library A	28	Average LibA score of library B ligands	−90.27	Number of global fingerprint bits only in library A	421
Number of assemblies only in library B	12	Average LibB score of library B ligands	57.05	Number of global fingerprint bits only in library B	387
Similarity score between the two libraries	0.0244	Bayesian distance between the two libraries	272.63	Similarity score between the two libraries	0.0901

[a] See section "Similarity analysis between training compounds and active hits" and reference (Bemis and Murcko, 1996)

[b] See section "Similarity analysis between training compounds and active hits" reference (Discovery Studio 2.5.5 User Guide, 2010)

[c] Done by implementing the finger print descriptor FCFC_6 which correspond to functional-class extended-connectivity fingerprint count up to diameter 6 (Discovery Studio 2.5.5 User Guide, 2010)

Library A: the list includes all training and testing compounds employed in pharmacophore and QSAR modeling (**1–68**, Fig. 2; Table A of Supplementary material) (SD file of this list is provided as Supplementary material)

Library B: captured hits (**69–86**, Fig. 6)

On the other hand, in Bayesian Model approach, two Bayesian models were built, one to learn library A and one to learn library B. Finally, it scores all ligands using both the models. A distance is computed as Eq. (4))

$$\text{Distance} = \text{ScoreAA} + \text{ScoreBB} - \text{ScoreAB} - \text{ScoreBA}, \quad (4)$$

where ScoreAA is the average score of library A molecules scored by the Bayesian model that learned library A molecules, while ScoreBB is the average score of library B molecules scored by the Bayesian model that learned library B. ScoreAB and ScoreBA are the average scores of libraries A and B molecules scored by the Bayesian models that learned libraries B and A, respectively. The higher the distance, the more dissimilar the libraries are (Discovery Studio 2.5.5 User Guide, 2010).

Finally, the Global Fingerprint comparison algorithm generates a global fingerprint for all ligands in the training list and all ligands in the hits list and then computes a Tanimoto similarity coefficient between the two libraries (Discovery Studio 2.5.5 User Guide, 2010).

Table 4 shows the results of the three similarity/diversity assessment procedures. Clearly, the three methods suggest minimal structural similarity between known BACE inhibitors and our captured hits providing further support to the significance of these hits.

Conclusions

In the current project, we explored the pharmacophoric space of 68 known TSA BACE inhibitors. We subsequently employed QSAR analysis to select optimal combination of pharmacophoric models and 2D physicochemical descriptors capable of explaining bioactivity variation within 68 training compounds. One optimal pharmacophoric model emerged in the successful QSAR equation. However, to closely mimic the TS, we were obliged to complement the successful pharmacophore model with strict shape constraints and carefully positioned hydroxyl fragment to mimic steric requirements of BACE TSA. The resulting shape–OH hybrid model and optimal pharmacophore were validated via ROC curve analysis and were found to exhibit excellent ability to separate active compounds from decoys. Screening against the NCI structural database by employing the two models, i.e., the shape–OH query followed the optimal pharmacophore, culminated in the discovery of two novel BACE inhibitors of nanomolar and low micromolar IC_{50} values.

Acknowledgments The authors thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds.

Author's personal copy

References

Abdula AM, Khalaf RA, Mubarak MS, Taha MO (2011) Discovery of new β-D-galactosidase inhibitors via pharmacophore modeling and QSAR analysis followed by in silico screening. J Comput Chem 32:463–482

Abu Hammad AM, Taha MO (2009) Pharmacophore modelling, quantitative structure–activity relationship analysis and shape-complemented in silico screening allow access to novel influenza neuraminidase inhibitors. J Chem Inf Model 49:978

Abu Khalaf R, Abu Sheikha G, Bustanji Y, Taha MO (2010) Discovery of new cholesteryl ester transfer protein inhibitors via ligand-based pharmacophore modeling and QSAR analysis followed by synthetic exploration. Eur J Med Chem 45:1598–1617

Abu Khalaf R, Abu Sheikha G, Bustanji Y, Taha MO (2011) Discovery of new β-D-glucosidase inhibitors via pharmacophore modeling and QSAR analysis followed by in silico screening. J Mol Model 17:443–464

Accelrys, Inc. (2005) CERIUS2, version 4.10 QSAR Users' Manual. Accelrys, Inc, San Diego, pp 221–235

Al-Nadaf A, Taha MO (2011) Discovery of new rennin inhibitory leads via sequential pharmacophore modeling, QSAR analysis, in silico screening and in vitro evaluation. J Mol Graph Model 29:843–864

Al-Nadaf A, Abu Sheikha G, Taha MO (2010) Elaborate ligand-based pharmacophore exploration and QSAR analysis guide the synthesis of novel pyridinium-based potent beta-secretase inhibitory leads Bioorg. Med Chem 18:3088–3115

Al-Sha'er M, Taha MO (2010) Elaborate ligand-based modeling reveals new nanomolar heat shock protein 90α inhibitors. Eur J Med Chem 45:4316–4330

Amyes T, Richard J (2007) Rational design of transition-state analogues as potent enzyme inhibitors with therapeutic applications. ACS Chem Biol 2:711–714

Arun KG, Geoffrey B, Cynthia H, Reiko K, Dongwoo S, Khaja A, Lin H, Jeffrey AL, Chan N, Gerald K, Jacques E, Jordan T (2001) Structure-based design: potent Inhibitors of human brain memapsin 2 (β-secretase). J Med Chem 44:2865

Beeley NRA, Sage C (2003) GPCRs: an update on structural approaches to drug discovery. Targets 2:19–25

Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frame-works. J Med Chem 39:2887–2893

Bersuker IB, Bahçeci S, Boggs JE (2000) Pharmacophore perception, development and use in drug design. International University Line, San Diego, pp 457–473

Beswick P, Charrier N, Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Gleave R, Hawkins J, Hussain I, Johnson CN, Macpherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Skidmore J, Soleil V, Smith KJ, Stanway S, Stemp G, Stuart A, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 3: identification of hydroxy ethylamines (HEAs) with nanomolar potency in cells. Bioorg Med Chem Lett 18:1022

Catalyst User Guide, Version 4.11 (2005) Accelrys Software Inc., San Diego

Charrier N, Clarke B, Cutler L, Demont E, Dingwall C, Dunsdon R, Hawkins J, Howes C, Hubbard J, Hussain I, Maile G, Matico R, Mosley J, Naylor A, O'Brien A, Redshaw S, Rowland P, Soleil V, Smith KJ, Sweitzer S, Theobald P, Vesey D, Walter DS, Wayne G (2009a) Second generation of BACE-1 inhibitors. Part 1 The need for improved pharmacokinetics. Bioorg Med Chem Lett 19:3664

Charrier N, Clarke B, Demont E, Dingwall C, Dunsdon R, Hawkins J, Hubbard J, Hussain I, Maile G, Matico R, Mosley J, Naylor A, O'Brien A, Redshaw S, Rowland P, Soleil V, Smith KJ, Sweitzer S, Theobald P, Vesey D, Walter DS, Wayne G (2009b) Second generation of BACE-1 inhibitors part 2: optimisation of the non-prime side substituent. Bioorg Med Chem Lett 19:3669

Charrier N, Clarke B, Cutler L, Demont E, Dingwall C, Dunsdon R, Hawkins J, Howes C, Hubbard J, Hussain I, Maile G, Matico R, Mosley J, Naylor A, O'Brien A, Redshaw S, Rowland P, Soleil V, Smith KJ, Sweitzer S, Theobald P, Vesey D, Walter DS, Wayne G (2009c) Second generation of BACE-1 inhibitors part 3: towards non hydroxyethylamine transition state mimetics. Bioorg Med Chem Lett 19:3674

Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Hawkins J, Hussain I, Macpherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Soleil V, Smith K, Stanway S, Stemp G, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008a) BACE-1 inhibitors part 2: identification of hydroxy ethylamines (HEAs) with reduced peptidic character. Bioorg Med Chem Lett 18:1017

Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Hawkins J, Hussain I, Macpherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Soleil V, Smith K, Stanway S, Stemp G, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008b) BACE-1 inhibitors part 1: identification of novel hydroxy ethylamines (HEAs). Bioorg Med Chem Lett 18:1011

Cooper JB (2002) Aspartic proteinases in disease: a structural perspective. Curr Drug Targets 3:155–173

Cumming JN, Le TX, Babu S, Carroll X, Chen L, Favreau P, Gaspari T, Guo DW, Hobbs Y, Huang U, Iserloh M (2008) Rational design of novel, potent piperazinone and imidazolidinone BACE1 inhibitors. Bioorg Med Chem Lett 18:3236

DePristo MA, de Bakker PIW, Blundell TL (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. Structure 12:831–838

Derek CC, Joseph RS, Rajiv C, Rebecca C, John WE, Kristi YF, Boyd LH, Yun H, Steve J, Lee DJ, Guixian J, Peter AL, Michael SM, Eric SM, William JM, Mary-Margaret O, Andrea MO, Albert JR, Kristine S, JunJun W, Eric W, Jonathan B (2008) Acylguanidine inhibitors of β-secretase: optimization of the pyrrole ring substituents extending into the S_1 and S_3 substrate binding pockets. Bioorg Med Chem Lett 18:1063

Discovery Studio 2.5.5 User Guide (2010)

Fisher R (1966) The principle of experimentation illustrated by a psychophysical experiment, 8th edn. Hafner Publishing Co, New York

Gao H, Williams C, Labute P, Bajorath J (1999) Binary quantitative structure–activity relationships (QSAR) analysis of estrogen receptor ligands. J Chem Inf Comput Sci 39:164–168

Georgia BM, Dennis C, Samuel LG, Ming-Tain L, Sanjeev KM, Philippe GN, Beth P, Hemaka AR, Harold GS, Shaun RS, Katharine MH (2007) β-Secretase (BACE-1) inhibitors: accounting for 10 s loop flexibility using rigid active sites. Bioorg Med Chem Lett 17:1117

Godemann R, Madden J, Kramer J, Smith MA, Fritz U, Hesterkamp T, Barker J, Hoeppner S, Hallett D, Cesura A, Ebneth A, Kemp J (2009) Fragment-based discovery of BACE1 inhibitors using functional assays. Biochemistry 48:10743

Hahn M (1997) Three-dimensional shape-based searching of conformationally flexible compounds. J Chem Inf Comput Sci 37:80–86

Heuisul P, Kyeongsik M, Hyo-Shin K, Ki DK, Dongchul L, Sang-Won S, Jae-Ung C, Bettina P, Deog-Young C (2008) Synthesis, SAR, and X-ray structure of human BACE-1 inhibitors with cyclic urea derivatives. Bioorg Med Chem Lett 18:2900

Author's personal copy

Huang D, Urs L, Peter K, Marco C, Alcide B, Amedeo C (2006) In silico discovery of β-secretase inhibitors. J Am Chem Soc 128:5436

Huang W, Yu H, Sheng R, Li J, Hu Y (2008) Identification of pharmacophore model, synthesis and biological evaluation of N-phenyl-1-arylamide and N-phenylbenzenesulfonamide derivatives as BACE 1 inhibitors. Bioorg Med Chem 16:10190

Irwin JJ, Shoichet BK (2005) ZINC—a free database of commercially available compounds for virtual screening. J Chem Inf Comput Sci 45:177–182

Jacobsson M, Liden P, Stjernschantz E, Bostroem H, Norinder U (2003) Improving structure-based virtual screening by multivariate analysis of scoring data. J Med Chem 46:5781–5789

Kirchmair J, Markt P, Distinto S, Wolber G, Langer T (2008) Evaluation of the performance of 3D virtual screening protocols: rMSD comparisons, enrichment assessments, and decoy selection—what can we learn from earlier mistakes? J Comput Aided Mol 22:213–228

Klebe G (2006) Virtual ligand screening: strategies, perspectives and limitations. Drug Discovery Today 11:580–594

Kurogi Y, Güner O (2001) Pharmacophore modeling and three-dimensional database searching for drug design using catalyst Curr. Med Chem 8:1035

Li H, Sutter J, Hoffmann R (2000) Pharmacophore perception, development, and use in drug design. International University Line, San Diego, pp 173–189

Lorna P, Andrea C, Francesco C, Federica B, Manuela B, Francesca M, Maurizio R, Vincenza A, Angela R (2008) Multi-target-directed coumarin derivatives: hAChE and BACE1 inhibitors as potential anti-Alzheimer compounds. Bioorg Med Chem Lett 18:423

Maria LB, Riccardo M, Anna M, Michela R, Carlo M (2009) Alzheimer's disease: new approaches to drug discovery. Curr Opin Chem Biol 13:303

Moffat K, Gillet VJ, Whittle M, Bravi G, Leach ARA (2008) A comparison of field-based similarity searching methods: cat-Shape, FBSS, and ROCS. J Chem Inf Model 48:719–729

Poptodorov K, Luu T, Langer T, Hoffmann R (2006) Methods and principles in medicinal chemistry, pharmacophores and pharmacophores searches, vol 2. Wiley-VCH, Weinheim, pp 17–47

Ramsey LF, Schafer WD (1997) The statistical sleuth, 1st edn. Wadsworth Publishing Company, Belmont

Rizzi L, Vaiana N, Sagui F, Genesio E, Pilli E, Porcari V, Romeo S (2009) Design, synthesis and docking studies of hydroxyethylamine and hydroxyethylsulfide BACE-1 inhibitors. Protein Pept Lett 16(5):86–90

Schramm V (2003) Enzymatic transition state poise and transition state analogues. Acc Chem Res 36:588–596

Schramm V (2005) Enzymatic transition states and transition state analogues. Curr Opin Struct Biol 15:604–613

Shawn JS (2009) Progress toward the development of a viable BACE-1 inhibitor. Drug Dev Res 70:101

Sheridan RP, Kearsley SK (2002) Why do we need so many chemical similarity search methods. Drug Discovery Today 7:903–911

Singh J, Chuaqui CE, Boriack-Sjodin PA, Lee W, Pontz T, Corbley MJ, Cheung H, Arduini RM, Mead JN, Newman MN, Papadatos JL, Bowes S, Josiah S, Ling LE (2003) Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGFβ receptor kinase (TβRI). Bioorg Med Chem Lett 13: 4355–4359

Smellie A, Teig S, Towbin P (1995) Poling: promoting conformational variation. J Comput Chem 16:171

Sprague PW, Hoffmann R (1997) CATALYST pharmacophore models and their utility as queries for searching 3D databases. In: Van de Waterbeemd H, Testa B, Folkers G (eds) Computer assisted lead finding and optimization. Verlag Helvetica Chimica Acta, Zürich, pp 223–240

Steuber H, Zentgraf M, Gerlach C, Sotriffer CA, Heine A, Klebe G (2006) Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions. J Mol Biol 363:174–187

Stubbs MT, Reyda S, Dullweber F, Moller M, Klebe G, Dorsch D, Mederski W, Wurziger H (2002) pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. ChemBioChem 3:246–249

Sutter J, Güner O, Hoffmann R, Li H, Waldman M (2000) Effect of variable weights and tolerances on predictive model generation. In: Güner OF (ed) Pharmacophore perception, development, and use in drug design. International University Line, San Diego, pp 501–511

Taha MO, Bustanji Y, Al-Bakri AG, Yousef M, Zalloum WA, Al-Masri IM, Atallah N (2007) Discovery of new potent human protein tyrosine phosphatase inhibitors via pharmacophore and QSAR analysis followed by in silico screening. J Mol Graph Modell 25:870–884

Taha MO, Atallah N, Al-Bakri AG, Paradis-Bleau C, Zalloum H, Younis K, Levesque RC (2008a) Discovery of new murf inhibitors via pharmacophore modeling and QSAR analysis followed by in silico screening. Bioorg Med Chem 16: 1218–1235

Taha MO, Bustanji Y, Al-Ghussein M, Mohammad M, Zalloum H, Al-Masri IM, Atallah N (2008b) Pharmacophore modeling, quantitative structure-activity relationship analysis, and in silico screening reveal potent glycogen synthase kinase-3beta inhibitory activities for cimetidine, hydroxychloroquine, and gemifloxacin. J Med Chem 51:2062–2077

Taha MO, Dahabiyeh LA, Bustanji Y, Zalloum H, Saleh S (2008c) Combining ligand-based pharmacophore modeling, QSAR analysis and in silico screening for the discovery of new potent hormone sensitive lipase inhibitors. J Med Chem 51:6478–6494

Taha MO, Habash M, Al-Hadidi Z, Al-Bakri A, Younis K, Sisan S (2011) Docking-based comparative intermolecular contacts analysis as new 3-D QSAR concept for validating docking studies and in silico screening: nMT and GP inhibitors as case studies. J Chem Inf Model 51:647–669

Triballeau N, Acher F, Brabet I, Pin J-P, Bertrand H-O (2005) Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. J Med Chem 48:2534–2547

Varghese J (2006) Human beta-secretase (BACE) and BACE inhibitors: progress report. Curr Top Med Chem 6:569

Verdonk ML, Marcel L, Berdini V, Hartshorn MJ, Mooij WTM, Murray CW, Taylor RD, Watson P (2004) Virtual screening using protein-ligand docking: avoiding artificial enrichment. J Chem Inf Comput Sci 44:793–806

Wen-Hai H, Rong S, Yong-Zhou H (2009) Progress in the development of nonpeptidomimetic BACE 1 inhibitors for Alzheimers disease. Curr Med Chem 16:1806

Yoshiari S, Takeshi K, Akinori A, Tetsuhiro N, Hachiro S (2008) Flavonols and flavones as BACE-1 inhibitors: structure–activity relationship in cell-free, cell-based and in silico studies reveal novel pharmacophore features. Biochim Biophys Acta 1780:819

Yoshio H, Hiroko O, Naoko M, Ryoji Y, Abdellah Y, Koushi H, Tooru K, Kazuki S, Yoshio H, Shoichi I, Yoshiaki K (2008) Novel non-peptidic and small-sized BACE1 inhibitors. Bioorg Med Chem Lett 18:1654