J Mol Model
DOI 10.1007/s00894-010-0737-1

ORIGINAL PAPER

# Discovery of new β-D-glucosidase inhibitors via pharmacophore modeling and QSAR analysis followed by in silico screening

**Reema Abu Khalaf · Ahmed Mutanabbi Abdula · Mohammad S. Mubarak · Mutasem O. Taha**

Received: 26 January 2010 / Accepted: 28 April 2010


**Abstract** Glycosidases, including β-D-glucosidase, are involved in a variety of metabolic disorders such as diabetes, viral or bacterial infections and cancer. Accordingly, we were prompted to find new β-D-glucosidase inhibitors. Towards this end we scanned the pharmacophoric space of this enzyme using a set of 41 known inhibitors. Genetic algorithm and multiple linear regression analyses were employed to select an optimal combination of pharmacophoric models and physicochemical descriptors to yield self-consistent and predictive quantitative structure-activity relationship (QSAR). Three pharmacophores emerged in the QSAR equations, suggesting the existence of more than one binding mode accessible to ligands within the β-D-glucosidase pocket. The successful pharmacophores were complemented with strict shape constraints in an attempt to optimize their receiver-operating characteristic (ROC) curve profiles. The validity of the QSAR equations and the associated pharmacophoric models were established experimentally by the identification of several β-D-glucosidase inhibitors retrieved via in silico search of two structural databases, namely the National Cancer Institute (NCI) list of compounds, and our in-house structural database of established drugs and agrochemicals (DAC).

**Keywords** β-D-glucosidase inhibitor · Quantitative structure-activity relationship · In silico screening · Pharmacophore modeling · Shape constraints · Receiver-operating characteristic curve

## Introduction

Glycosidases are enzymes that are widespread in living organisms [1, 2]. They catalyze the hydrolysis of glycosidic bonds in carbohydrates and glycoconjugates, resulting in low-molecular weight monosaccharides and oligosaccharides. Glycosidases, including β-D-glucosidase, are involved in a variety of metabolic disorders and diseases such as diabetes, viral or bacterial infections and cancer. Therefore, the inhibition of glycosidases, including β-D-glucosidase, has many potential applications, e.g., antidiabetic, antiviral (HIV, influenza) and anticancer [3–8].

The intense interest during the last decade in the chemistry, biochemistry and pharmacology of glycosidase inhibitors has led to the discovery of many types of natural and synthetic glycosidase inhibitors [9–14]. All previous studies have dealt with glycosidase inhibitors including sugar-mimics (azasugars and carbasugars) and their analogues; however, no attempts have been made to discover new scaffolds as glycosidase inhibitors that have better chemical stabilities, pharmacokinetic profiles and higher potencies [9–14].

The current interest in the development of new β-D-glucosidase inhibitors, combined with the lack of adequate

**Electronic supplementary material** The online version of this article (doi:10.1007/s00894-010-0737-1) contains supplementary material, which is available to authorized users.

R. A. Khalaf
Department of Pharmaceutical Sciences, Faculty of Pharmacy, Al Zaytoonah Private University of Jordan, Amman, Jordan

A. M. Abdula · M. S. Mubarak
Department of Chemistry, Faculty of Science, University of Jordan, 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

Published online: 21 May 2010

computer-aided drug discovery efforts in this area, prompted us to explore the possibility of developing ligand-based three-dimensional (3D) pharmacophore(s) integrated within a self-consistent quantitative structure-activity relationship (QSAR) model for β-D-glucosidase inhibitors. The pharmacophore model(s) can be used as 3D search query(ies) to mine 3D libraries for new β-D-glucosidase inhibitors, while the QSAR model helps to predict the biological activities of the captured compounds and therefore prioritize them for in vitro evaluation.

The fact that all reported β-D-glucosidase inhibitors are slow binding/transition state analogue (TSA) sugar-mimics [9–14] complicates pharmacophore modeling and subsequent in-silico search. TSAs resemble the substrate at its postulated transition to products, implying that such inhibitors require stringent steric and 3D provision in order to dock into the enzymatic binding site during its sterically demanding high-energy transition state. TSAs are known to be much more tightly bound to the targeted enzyme than their ground state counterparts (i.e., substrate analogues), which further supports the notion of pronounced sensitivity of TSA-enzyme complexes to slight misalignments among their complementary attractive groups [15–17]. This conduct is expected of a rugged structure-activity surface, which limits the ability of the pharmacophore theory to explain activity/inactivity variations among training compounds. In fact, pharmacophore modeling requires continuous bioactivity variation attributable to the presence or absence of certain chemical features, i.e., it requires a smooth SAR surface.

The pronounced sensitivity of TSAs to slight structural modifications is also expected to 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 [18].

In fact, no previous pharmacophore modeling efforts have been reported for β-D-glucosidase inhibitors, probably as a consequence of their rugged SAR. Accordingly, we were prompted to hybridize our QSAR-based pharmacophore models with tight ligand shapes and to use the resulting combination as 3D search queries.

We previously reported the use of this innovative approach towards the discovery of new inhibitory leads against glycogen synthase kinase 3β [19], dipeptidyl peptidase [20], hormone sensitive lipase [21], bacterial MurF [22], protein tyrosine phosphatase 1B [23], influenza neuraminidase [24] and cholesteryl ester transfer protein [25].

We employed the HYPOGEN module from the CATALYST software package to construct numerous plausible binding hypotheses for β-D-glucosidase inhibitors [26]. Subsequently, genetic function algorithm (GFA) and multiple linear regression (MLR) analyses were employed to search for an optimal QSAR that combines high-quality binding pharmacophores with other molecular descriptors, and is capable of explaining bioactivity variation across a collection of diverse β-D-glucosidase inhibitors. Thereafter, the optimal pharmacophores were complemented with tight shape constraints and used as 3D search queries to screen several available virtual molecular databases for new β-D-glucosidase inhibitors.

The optimal pharmacophores were further validated by evaluating their ability to successfully classify a list of compounds as active or inactive by assessing their receiver-operating characteristic (ROC) curves. Subsequently, the optimal pharmacophores were complemented with tight shape constraints to enhance their ROC profiles. Thereafter, the resulting shape-complemented pharmacophores were used as 3D search queries to screen several available virtual molecular databases for new β-D-glucosidase inhibitors.

CATALYST models drug-receptor interactions using information derived from the ligand structures [26–34]. HYPOGEN identifies a 3D array of a maximum of five chemical features common to active training ligands that provide relative alignment for each input molecule consistent with binding to a proposed common receptor site. The conformational flexibility of training ligands is modeled by creating multiple conformers that cover a specified energy range for each input molecule [23, 29–31, 35–39].

The SHAPE module in CATALYST is a shape-based similarity searching method. The Van der Waals surface of a molecule (in a 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 [40]. CATALYST pharmacophores, with or without shape constraints, have been used as 3D queries for database searching and in 3D-QSAR studies [29, 31, 35, 40].

## Materials and methods

### Molecular modeling

#### *Software and hardware*

The following software packages were utilized in the present research.

- ♦ CATALYST (Version 4.11), Accelrys (www.accelrys.com).
- ♦ CERIUS2 (Version 4.10), Accelrys (www.accelrys.com).

♦ CS ChemDraw Ultra 6.0, Cambridge Soft Corporation (http://www.cambridgesoft.com).

Pharmacophore and QSAR modeling studies were accomplished using CATALYST (HYPOGEN module) and CERIUS2 software suites from Accelrys (San Diego, CA, 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 accomplished employing ChemDraw Ultra 6.0 installed on a Pentium 4 PC.

## Data set of β-D-glucosidase inhibitors

The structures of 41 β-D-glucosidase inhibitors **1**–**41** (Fig. 1, Table 1) were collected from recently published literature [41–45]. The in vitro bioactivities of the collected inhibitors were determined by employing identical bioassay conditions and were expressed as $K_i$ values (inhibition constants against almond β-D-glucosidase, μM), which allowed us to pool them for pharmacophore and QSAR analysis. The logarithms of measured $1/K_i$ values were used in 3D-QSAR, thus correlating the data linear to the free energy change. Collected compounds that were reported to be devoid of activity (e.g., **19**, **21**, **24**, **25**, **26**, **27**, **28**, **29**, **31**, **32**, **33**, **35**, **36**, **38**, **39**, **40** and **41**, Fig. 1 and Table 1) were assumed to have $K_i$ values of 2,000 μM, which is four logarithmic cycles from the most potent compound (**10**, $K_i$= 0.04 μM). These assumptions are necessary to allow statistical correlation and QSAR analysis. The logarithmic transformation of $K_i$ values should minimize any potential errors resulting from such assumptions [21, 22]. The two-dimensional (2D) chemical structures of the inhibitors were sketched using ChemDraw Ultra and saved in MDL-molfile 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 CATALYST conformational analysis.

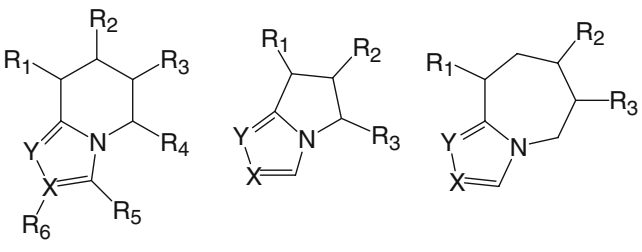



**Fig. 1** The chemical scaffolds of β-D-glucosidase training compounds. Detailed structures are given in Table 1

## Conformational analysis

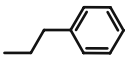
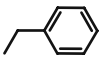
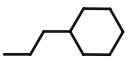
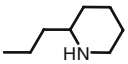
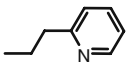
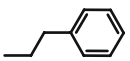
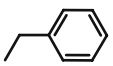
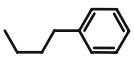
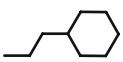
The molecular flexibilities of the collected compounds were taken into account by considering each compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a given energy range. Accordingly, the conformational space of each inhibitor (**1**–**41**; Fig. 1, Table 1) was explored adopting the "best conformer generation" option within CATALYST based on the generalized CHARMm force field implemented in the program. Default parameters were employed in the conformation generation procedure, i.e., a conformational ensemble was generated with an energy threshold of 20 kcal $mol^{-1}$ from the local minimized structure, which has the lowest energy level and a maximum limit of 250 conformers per molecule. This search procedure will probably identify the best 3D arrangement of chemical functionalities explaining the activity variations among the training set [26].

## Pharmacophoric hypotheses generation

All 41 molecules with their associated conformational models were regrouped into a spreadsheet. The biological data of the inhibitors were reported with uncertainty values of 2 or 3, which means that the actual bioactivity of a particular inhibitor is assumed to be situated somewhere in an interval ranging from 1/2 to 2 or 1/3 to 3 times the reported bioactivity value of that inhibitor, respectively [30, 32, 34]. The uncertainty value is of great impact on the qualities of the resulting pharmacophores, as it controls the number of training compounds within the "most potent category".

Subsequently, a structurally diverse training set was selected for pharmacophore modeling of β-D-glucosidase: **1**, **9**, **10**, **13**, **14**, **15**, **21**, **22**, **23**, **24**, **25**, **27**, **34**, **37**, **38** and **39** (Fig. 1, Table 1). Typically, CATALYST requires informative training sets that include at least 16 compounds of evenly spread bioactivities over at least three and a half logarithmic cycles. Lesser training lists could lead to chance correlation and thus faulty models [30, 32, 34]. The selected training set was utilized to conduct eight modeling runs (Table A, Electronic Supplementary Material) to explore the pharmacophoric space of β-D-glucosidase inhibitors. The exploration process included an altering interfeature spacing parameter (100 and 300 pm) 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, third, fifth, and seventh runs and from 4 to 4 for the second, fourth, sixth and eighth runs (Table A, Electronic Supplementary Material). Pharmacophore modeling employing CATALYST proceeds through three successive phases: constructive phase, subtractive phase and optimization phase [30, 32, 34]. During the constructive phase, CATALYST generates common conformational alignments among the most-active

**Table 1** The structures of β-D-glucosidase inhibitors utilized in modeling. The corresponding scaffolds are displayed in Fig. 1

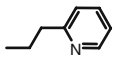
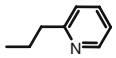
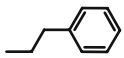
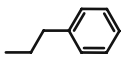
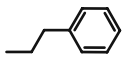
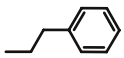
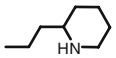
| No | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | X | Y | Ki values (μM) against β-D-glucosidase |
|---|---|---|---|---|---|---|---|---|---|
| **1*** | (*S*)OH | (*S*)OH | (S)OH | H | H | H | C | N | 1.00 |
| **2** | (*S*)OH | (*S*)OH | (*S*)OH | H |  | H | C | N | 15 |
| **3** | (*S*)OH | (*S*)OH | (*S*)OH | H |  | H | C | N | 2.80 |
| **4** | (*S*)OH | (*S*)OH | (*S*)OH | H |  | H | C | N | 1.60 |
| **5** | (*S*)OH | (*S*)OH | (*S*)OH | H |  | H | C | N | 170 |
| **6** | (*S*)OH | (*S*)OH | (*S*)OH | H |  | H | C | N | 2.80 |
| **7*** | (*S*)OH | (*S*)OH | (*S*)OH | H | $CH_3$ | H | C | N | 5.20 |
| **8** | (*S*)OH | (*S*)OH | (*S*)OH | H | $CH_2OH$ | H | C | N | 20.0 |
| **9** | (*S*)OH | (*S*)OH | (*S*)OH | H | H |  | C | N | 0.08 |
| **10** | (*S*)OH | (*S*)OH | (*S*)OH | H | H |  | C | N | 0.04 |
| **11*** | (*S*)OH | (*S*)OH | (*S*)OH | H | H |  | C | N | 0.09 |
| **12** | (*S*)OH | (*S*)OH | (*S*)OH | H | H |  | C | N | 0.06 |

training compounds. Only molecular alignments based on a maximum of five chemical features are considered. The program identifies a particular compound as being within the most active category if it satisfies Eq. 1 [30, 32, 34].

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

Where "MAct" is the activity of the most active compound in the training set, "Unc" is the uncertainty of the compounds and "Act" is the activity of the training compounds under question. However, if there are more than eight most-active inhibitors, only the top eight are used. Therefore, in this case, the most potent

**Table 1** (continued)

| No | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | X | Y | Ki values (μM) against β-D-glucosidase |
|---|---|---|---|---|---|---|---|---|---|
| **13** | (*S*)OH | (*S*)OH | (*S*)OH | H | H |  | C | N | 0.05 |
| **14** | (*S*)OH | (*S*)OH | (*R*)OH | H | H |  | C | N | 0.14 |
| **15** | (*S*)OH | (*S*)OH | (*S*)OH | H | H | $CH_3$ | C | N | 0.19 |
| **16** | (*S*)OH | (*S*)OH | (*S*)OH | H | H | $CH_2OH$ | C | N | 0.10 |
| **17** | (*S*)OH | (*S*)OH | (*S*)OH | H | $CH_2OH$ |  | C | N | 0.31 |
| **18** | (*S*)OH | (*S*)OH | (*S*)OH | H |  |  | C | N | 4.00 |
| **19** | (*S*)OH | (*S*)OH | (*S*)OH | H |  |  | C | N | NA[a] |
| **20** | (*S*)OH | (*S*)OH | (*S*)OH | H | $CH_2CH_2$ | $CH_2CH_2$ | C | N | 350 |
| **21*** | (*R*)OH | (*R*)OH | (*R*)OH | H | H | H | C | N | NA |
| **22** | (*S*)OH | (*R*)OH | (*R*)OH | H | H | H | C | N | 20 |
| **23** | (*S*)OH | (*S*)OH | (*R*)OH | H | H | H | C | N | 17 |
| **24** | (*R*)OH | (*S*)OH | (*R*)OH | H | H | H | C | N | NA |

category of training compounds included **9**, **10**, **13**, **14** and **15**.

In the subsequent subtractive phase, CATALYST eliminates some hypotheses that fit inactive training compounds. A particular training compound is defined as being inactive if it satisfies Eq. 2 [30, 32, 34]:

$$\mathrm{Log(Act)} - \mathrm{log(MAct)} > 3.5 \qquad (2)$$

Accordingly, compounds **21**, **24**, **25**, **27**, **37**, **38** and **39** are considered least active and therefore employed in the subtractive phase.

However, in the optimization phase, CATALYST applies fine perturbations in the form of vectored feature rotation, adding new features and/or removing a feature, to selected hypotheses that survived the subtractive phase, in an attempt to find new models of enhanced bioactivity/

**Table 1** (continued)

| No | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | X | Y | Ki values (μM) against β-D-glucosidase |
|---|---|---|---|---|---|---|---|---|---|
| **25** | (*R*)OH | (*S*)OH | (*S*)OH | H | H | H | C | N | NA |
| **26** | (*R*)OH | (*R*)OH | (*S*)OH | H | H | H | C | N | NA |
| **27** | (*S*)OH | (*R*)OH | (*S*)OH | H | H | H | C | N | NA |
| **28*** | (*R*)OH | (*R*)OH | (*R*)OH | (*S*)$CH_2OH$ | H | H | N | C | NA |
| **29** | (*S*)OH | (*R*)OH | (*S*)OH | (*R*)$CH_2OH$ | H | H | N | C | NA |
| **30** | (*R*)OH | (*S*)OH | (*R*)OH | (*R*)$CH_2OH$ | H | H | N | C | 230 |
| **31** | (*R*)OH | H | (*S*)OH | (*R*)$CH_2OH$ | H | H | N | C | NA |
| **32*** | (*R*)OH | (*R*)OH | (*R*)$CH_2OH$ | - | - | - | C | N | NA |
| **33** | (*S*)OH | (*R*)OH | (*R*)$CH_2OH$ | - | - | - | C | N | NA |
| **34*** | (*S*)OH | (*S*)OH | (*R*)$CH_2OH$ | - | - | - | C | N | 64 |
| **35** | (*R*)OH | (*S*)OH | (*R*)$CH_2OH$ | - | - | - | C | N | NA |
| **36** | (*S*)OH | (*S*)OH | (*S*)$CH_2OH$ | - | - | - | C | N | NA |

mapping correlation, i.e., improved 3D-QSAR properties. Eventually, CATALYST selects the highest-ranking models (ten by default) and presents them as the optimal pharmacophore hypotheses resulting from the particular automatic modeling run. In conclusion, our pharmacophore exploration efforts, which included eight automatic runs, culminated in 80 pharmacophore models of variable qualities.

## Assessment of the generated hypotheses

During pharmacophore modeling, CATALYST attempts to minimize a cost function consisting of three terms: weight cost, error cost and configuration cost [26, 30–34]. Weight cost is a value that increases as the feature weight in a model deviates from an ideal value of 2. The deviation

**Table 1** (continued)

| No | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | X | Y | Ki values (μM) against β-D-glucosidase |
|---|---|---|---|---|---|---|---|---|---|
| **37** | (*R*)OH | (*S*)OH | (*S*)$CH_2OH$ | - | - | - | C | N | 100 |
| **38** | (*R*)OH | (*R*)OH | (*S*)$CH_2OH$ | - | - | - | C | N | NA |
| **39*** | (*S*)OH | (*R*)OH | (*S*)$CH_2OH$ | - | - | - | C | N | NA |
| **40** | (*S*)$CH_2OH$ | (*S*)OH | (*R*)$CH_2OH$ | - | - | - | N | C | NA |
| **41** | (*R*)OH | (*S*)OH | (*R*)OH | - | - | - | N | C | NA |

*These compounds were employed as the external test set in QSAR modeling
[a]NA = Not active

between the estimated activities of the training set and their experimentally determined values adds to the error cost. The activity of any compound can be estimated from a particular hypothesis through Eq. 3 [26].

$$\text{Log}(\text{Estimated Activity}) = \text{I} + \text{Fit} \quad (3)$$

Where, I is the intercept of the regression line obtained by plotting the log of the biological activity of the training set compounds against the Fit values of the training compounds. The Fit value for any compound is obtained automatically employing Eq. 4 [26].

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

Where, Σmapped hypothesis features represents the number of pharmacophore features that successfully superimpose (i.e., map or overlap with) corresponding chemical moieties within the fitted compound, and W is the weight of the corresponding hypothesis feature spheres (this value is fixed at 1.0 in CATALYST-generated models), disp is the distance between the center of a particular pharmacophoric sphere (feature centroid) and the center of the corresponding superimposed chemical moiety of the fitted compound, tol is the radius of the pharmacophoric feature sphere (known as Tolerance, equals 1.6 Å by default) and $\Sigma(\text{disp}/\text{tol})^2$ is the summation of $(\text{disp}/\text{tol})^2$ values for all pharmacophoric features that successfully superimpose corresponding chemical functionalities in the fitted compound [26].

The third term, i.e., the configuration cost, penalizes the complexity of the hypothesis. This is a fixed cost, which is equal to the entropy of the hypothesis space. The greater the numbers of features (a maximum of five) in a generated hypothesis, the higher the entropy with subsequent increase in this cost. The overall cost (total cost) of a hypothesis is calculated by summing over the three cost factors; however, error cost is the main contributor to total cost. CATALYST also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are distributed normally about their mean. Accordingly, the greater the difference from the null hypothesis cost, the more likely that the hypothesis does not reflect a chance correlation. In a successful automatic modeling run, CATALYST ranks the generated models according to their total costs [26]. Table B in the Electronic Supplementary Material summarizes the success criteria of best representative pharmacophores for each modeling run.

An additional approach to assessing the quality of CATALYST-HYPOGEN pharmacophores is cross-validation using the Cat-Scramble program implemented in CATALYST. This validation procedure is based on Fisher's randomization test [46]. In this validation test, we selected a 95% confidence level, which instructs CATALYST to generate 19 random spreadsheets by the Cat-Scramble

command. Subsequently, CATALYST-HYPOGEN is challenged to use these random spreadsheets to generate hypotheses using exactly the same features and parameters used in generating the initial unscrambled hypotheses [47]. 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.

## Clustering of the generated pharmacophore hypotheses

The generated pharmacophoric models (80) were clustered utilizing the hierarchical average linkage method available in CATALYST. Subsequently, the highest-ranking representatives, as judged based on the correlation (r) of their best-fit values against the logarithmic transformation of the corresponding inhibitory $K_i$ values of the training compounds, were selected to represent their corresponding clusters in subsequent QSAR modeling.

## QSAR modeling

A set of training inhibitors was selected from the collected list (**1**–**41**, Fig. 1, Table 1) for QSAR modeling. However, since it is essential to access the predictive power of the resulting QSAR models on an external set of inhibitors, eight molecules (ca. 20% of the dataset) were employed as an external test set to validate the QSAR models. The test molecules were selected as follows: the inhibitors were ranked according to their $K_i$ values, subsequently; every fifth compound was selected for the test set starting from the high-potency end. This selection considers the fact that the test molecules should represent a range of biological activities similar to that of the training set. The selected test set for QSAR modeling was: **1**, **7**, **11**, **21**, **28**, **32**, **34**, **39** (numbers as in Fig. 1 and Table 1).

The logarithm of measured $1/K_i$ (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 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 [43]. Furthermore, the training compounds were fitted (using the Best-fit option in CATALYST) against the representative pharmacophores and their fit values were added as additional descriptors. The fit value for any compound is obtained automatically via Eq. 4 [26].

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). GFA techniques rely on the evolutionary operations of "crossover and mutation" to select optimal combinations of descriptors (i.e., chromosomes) capable of explaining bioactivity variation among training compounds from a large pool of possible descriptor combinations, i.e., chromosomes population. However, to avoid overwhelming GFA-MLR with large number of poor descriptor populations, we removed lowest-variance descriptors (20%) prior to QSAR analysis. Each chromosome is associated with a fitness value that reflects how good it is compared to other solutions. The fitness function employed herein is based on Friedman's 'lack-of-fit' (LOF) [48].

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 three to six explanatory terms.

All QSAR models were validated employing leave one-out (LOO) cross-validation ($r_{LOO}^2$), bootstrapping ($r_{BS}^2$) and predictive $r^2$ ($r_{PRESS}^2$) calculated from the test set. The predictive $r_{PRESS}^2$ is defined as [48]:

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

Where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules, and PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set.

## Addition of shape constraints

Our QSAR-based pharmacophoric hypotheses were complemented with ligand-shapes and were employed as 3D search queries. The shape components were introduced employing the CatShape module of CATALYST [49]. Each pharmacophore was mapped against the most potent training inhibitor **10** ($K_i$=0.04 μM). The best fitted conformers were used to generate shape constraints (with tolerance values ranging from 95 to 105%) that were subsequently merged with the corresponding pharmacophores [26].

## In silico screening for new β-D-glucosidase inhibitors

Each shape-complemented pharmacophore model was employed as 3D search query to screen the national cancer institute list (NCI, 238,819 compounds) and our own in-

house database of known drugs and agrochemicals (DAC, 3,005 compounds). In silico screening was performed employing the "Best Flexible Database Search" option implemented within CATALYST. The NCI hits were subsequently filtered based on Lipinski's and Veber's rules [50, 51]. However, DAC hits were processed without subsequent post-filtering. Surviving hits were fitted against each corresponding pharmacophore model using the "best fit" option implemented within CATALYST. The fit values together with the relevant molecular descriptors of each hit were substituted in the corresponding QSAR equations. The highest ranking molecules based on QSAR predictions were acquired and tested in vitro.

### In vitro experimental studies

#### *Materials*

β-D-glucosidase (EC 3.2.1.21) from almond and its corresponding substrate *p*-nitrophenyl-β-D-glucopyranoside were purchased from Sigma-Aldrich (St. Louis, MO). NCI hits were kindly donated from the National Cancer Institute.

#### Preparation of substrate and enzyme solution

*p*-Nitrophenyl-β-D-glucopyranoside aqueous solution (13 mM) was prepared in acetate buffer (pH 5) as substrate solution for β-D-glucosidase. While, β-D-glucosidase was dissolved in acetate buffer (pH 5) to obtain final solution of 0.032 units/μL.

#### Preparation of hit compounds for in vitro assay

The tested compounds were dissolved initially in DMSO to yield stock solutions of 0.01 M. Subsequently, they were diluted to the required concentrations (7, 10, 40 and 100 μM) with the buffer solution.

#### β-D-glucosidase inhibition by hit compounds

The inhibitory potentials of the hit compounds against β-D-glucosidase were evaluated by spectrophotometric assay of released *p*-nitrophenol. The β-D-glucosidase solution (0.16 units) was pre-incubated with 7, 10, 40 and 100 μM of each particular hit compound for 1 h at 25°C. The final concentration of DMSO did not exceed 1.0%. Subsequently, substrate solution (1.3 mM) was added to the reaction mixture and the concentration of released *p*-nitrophenol was monitored at 405 nm every minute within a 5-min period. Substrate concentrations were selected to approximate $K_m$ values [41–45]. The percentage inhibition was determined from the residual activity for each compound by comparing the β-D-glucosidase activity with and without hit compound [41–45].

## Results and discussion

CATALYST enables automatic pharmacophore construction by using a collection of molecules with activities ranging over a number of orders of magnitude. CATALYST pharmacophores (hypotheses) explain the variability of activity of the molecules 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 β-D-glucosidase inhibitors.

### Data mining and conformational coverage

The literature was surveyed extensively to identify as many reported structurally diverse β-D-glucosidase inhibitors as possible (**1**–**41**, Fig. 1 and Table 1). However, we were confined by assay methodologies to 41 compounds assayed via identical bioassay conditions (against almond β-D-glucosidase). The 2D structures of the inhibitors were imported into CATALYST and converted automatically into plausible 3D single conformer representations. The resulting single conformer 3D structures were used as starting point for conformational analysis and in the determination of various molecular descriptors for QSAR modeling.

The conformational space of each inhibitor was sampled extensively utilizing the poling algorithm employed within the CONFIRM module of CATALYST [31]. Conformational coverage was performed employing the "Best" module to ensure extensive sampling of conformational space. Efficient conformational coverage guarantees minimum conformation-related noise during pharmacophore generation and validation stages. Pharmacophore generation and pharmacophore-based search procedures are known for their sensitivity to inadequate conformational sampling within the training compounds [52].

### Exploration of the pharmacophoric space of β-D-glucosidase inhibitors

CATALYST-HYPOGEN enables automatic pharmacophore construction by using a collection of at least 16 molecules with bioactivities spanning over 3.5 orders of magnitude [26, 30–34]. The generated 3D pharmacophores can be used as search queries to screen virtual 3D-structural libraries. As we have an informative list of 41 inhibitors of evenly spread bioactivities over more than 3.5 orders of magnitude, we were prompted to employ HYPOGEN to identify possible pharmacophoric binding modes assumed by inhibitors of β-D-glucosidase. HYPOGEN implements an optimization algorithm that evaluates a large number of

potential models for a particular target through fine perturbations to hypotheses that survive the subtractive and constructive phases [30]. 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 should not exceed 17 (corresponding to $2^{17}$ hypotheses to be assessed by CATALYST) to guarantee thorough analysis of all models [31].

The size of the investigated pharmacophoric space is a function of training compounds, selected input chemical features and other CATALYST control parameters [31]. Restricting the extent of explored pharmacophoric space should improve the efficiency of optimization by allowing effective evaluation of limited number of pharmacophoric models. On the other hand, extensive restrictions imposed on the pharmacophoric space may reduce the possibility of discovering optimal pharmacophoric hypotheses, as they may occur outside the "boundaries" of the pharmacophoric space.

Therefore, we decided to explore the pharmacophoric space of β-D-glucosidase inhibitors under reasonably imposed "boundaries" through eight HYPOGEN automatic runs and a carefully selected training set. Table A (Electronic Supplementary Material) summarizes CATALYST run parameters employed in exploring the pharmacophoric space of β-D-glucosidase inhibitors. The training set (see Materials and methods) was selected in such away to guarantee maximal 3D diversity and continuous bioactivity spread over more than 3.5 logarithmic cycles. Furthermore, the training compounds were selected in such a way that differences in their inhibitory bioactivities are attributable primarily to the presence or absence of pharmacophoric features (e.g., HBA or HBD or Hbic or RingArom) rather than steric shielding and/or bioactivity-enhancing or -reducing auxiliary groups (e.g., electron donating or withdrawing groups). Special emphasis was placed on the 3D diversity of the most active compounds in the training set due to their significant influence on the extent of the evaluated pharmacophore space through the Constructive Phase of the HYPOGEN algorithm.

Guided by our reasonably restricted pharmacophore exploration concept, we restricted the software to explore pharmacophoric models from zero to three HBA, HBD, Hbic, and RingArom features and from zero to one PI features instead of the default range of zero to five. Furthermore, we instructed HYPOGEN to explore only four- and five-featured pharmacophores, i.e., ignore models of lesser number of features (see Table A in Electronic Supplementary Material). The latter restriction has the advantage of narrowing the investigated pharmacophoric space and representing the feature-rich nature of known β-D-glucosidase inhibitors.

In each run, the resulting binding hypotheses were ranked automatically according to their corresponding "total cost" value, defined as the sum of error cost, weight cost and configuration cost [26, 30–34]. Error cost provides the highest contribution to total cost and it is related directly to the capacity of the particular pharmacophore as the 3D-QSAR model, i.e., in correlating the molecular structures to the corresponding biological responses [26, 30–34]. HYPOGEN also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are distributed normally about their mean. Accordingly, the greater the difference from the null hypothesis cost (residual cost, Table B, Electronic Supplementary Material), the more likely that the hypothesis does not reflect a chance correlation [26, 30–34].

An additional validation technique, known as Cat. Scramble, was recently introduced into CATALYST [26]. This procedure is based on Fisher's randomization test [46]. 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 hypotheses (i.e., generated from unscrambled data) is lowered proportional to the number of times the software succeeds in generating binding hypotheses from scrambled data of apparently better cost criteria than the parent hypotheses.

Eventually, 80 pharmacophore models emerged from eight automatic HYPOGEN runs. All models illustrated confidence levels≥95%. Clearly ,from Table B (Electronic Supplementary Material), all models shared comparable features and acceptable statistical success criteria. The emergence of several comparable pharmacophore models suggests the ability of β-D-glucosidase inhibitors to assume multiple pharmacophoric binding modes within the binding pocket for each enzyme. Therefore, to select any particular pharmacophore hypothesis as a sole representative of the binding process is quite challenging.

## QSAR modeling

Pharmacophoric hypotheses are important tools in drug design and discovery as they provide excellent insights into ligand–macromolecule recognition; moreover, they can be used as 3D search queries to look for new and biologically interesting scaffolds. However, their predictive value as 3D-QSAR models is usually limited by steric shielding and auxiliary groups (electron-donating and -withdrawing moieties) [33]. This point, combined with the fact that pharmacophore modeling of β-D-glucosidase inhibitors furnished many binding hypotheses of comparable statistical criteria (Table B, Electronic Supplementary Material) prompted us to employ classical QSAR analysis to search for the best combination of pharmacophore(s) and other 2D descriptors capable of explaining bioactivity variation across the whole list of collected inhibitors **1**–**41** (Fig. 1 and Table 1). We employed GFA and MLR QSAR (GFA-MLR-QSAR) analysis 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 [48]. 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 representative hypotheses against all collected inhibitors **1**–**41** (Fig. 1, Table 1) were enrolled as independent variables (genes) in a cycle of GFA-MLR-QSAR analysis over 30,000 iterations employing Friedman's LOF fitness criterion [48, 53]. However, since it is essential to access the predictive power of the resulting QSAR models on an external set of inhibitors, we randomly selected eight inhibitors (Fig. 1, Table 1) and employed them as an external test set for validating the QSAR models (i.e., $r^2_{PRESS}$). Moreover, all QSAR models were cross-validated automatically using LOO crossvalidation in CERIUS2 [48, 53].

Equations 6, 7 and 8 show the details of the optimal QSAR models. Figures 2, 3, and 4 show the corresponding scatter plots of experimental versus calculated bioactivities. Interestingly, Eqs. 6 and 7 required the pharmacophore fit values to be in spline form to achieve significant extrapolatory prediction against the external test set, while Eq. 8 required quadratic transformation of the pharmacophoric fit values.

$$\begin{aligned}\mathrm{Log}(1/\mathrm{Ki}) &= -2.52 + 0.62[\mathrm{Hypo10}/1 - 7.03] \\ &\quad + 16.67\left[{}^3\chi^{\mathrm{C}} - 0.99\right] - 2.16\left[{}^3\chi^{\mathrm{P}} - 5.60\right] \\ &\quad -9.18[\mathrm{ShadowXYfrac} - 0.60] \\ &\quad +133.01[-0.089 - \mathrm{JursFNSA3}] \\ r^2{}_{33} &= 0.92, \\ &\quad \mathrm{F-statistic} = 58.42,\ r^2{}_{\mathrm{LOO}} = 0.89, \\ &\quad r^2{}_{\mathrm{BS}} = 0.918, r^2{}_{\mathrm{PRESS}} = 0.61\end{aligned} \tag{6}$$

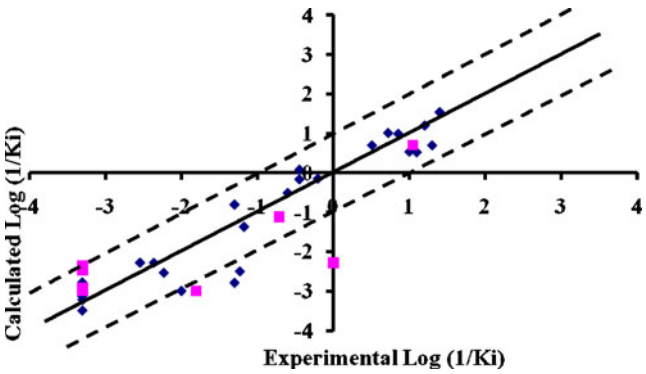



**Fig. 2** Experimental versus fitted (♦, 33 training compounds, $r^2_{LOO}$= 0.889) and predicted (■, 8 compounds, $r^2_{PRESS}$=0.614) bioactivities calculated from quantitative structure-activity relationship (QSAR) Eq. 6 for β-D-glucosidase inhibitors. *Solid lines* Regression lines for the fitted bioactivities, *dotted lines* 1.0 log point error margins

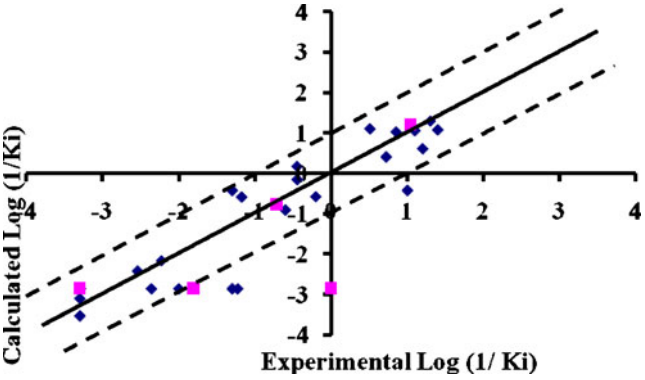



**Fig. 3** Experimental versus fitted (♦, 33 training compounds, $r^2_{LOO}$= 0.814) and predicted (■, 8 compounds, $r^2_{PRESS}$=0.554) bioactivities calculated from QSAR Eq. 7 for β-D-glucosidase inhibitors. *Solid lines* Regression lines for the fitted bioactivities, *dotted lines* 1.0 log point error margins

$$\begin{aligned}\mathrm{Log}(1/\mathrm{Ki}) &= -0.42 + 0.73[\mathrm{Hypo4}/4 - 8.03] \\ &\quad -1.22[3 - \mathrm{AtypeH51}] - 7.72[\mathrm{SaasN} - 1.63] \\ r^2{}_{33} &= 0.87, \mathrm{F-statistic} = 61.16, r^2{}_{\mathrm{LOO}} = 0.81, \\ r^2{}_{\mathrm{BS}} &= 0.87, r^2{}_{\mathrm{PRESS}} = 0.55\end{aligned} \tag{7}$$

$$\begin{aligned}\mathrm{Log}(1/\mathrm{Ki}) &= -2.13 + 0.021[\mathrm{Hypo9}/4]^2 + 6.10\left({}^3\chi^{\mathrm{C}}\right)^2 \\ &\quad -0.16\left({}^2\chi\right)^2 - 8.66\,(\mathrm{ShadowXYfrac})^2 \\ &\quad +2.63 \times 10(\mathrm{JursPPSA3})^2 \\ r^2{}_{33} &= 0.89, \mathrm{F-statistic} = 40.92, r^2{}_{\mathrm{LOO}} = 0.84, \\ r^2{}_{\mathrm{BS}} &= 0.89, r^2{}_{\mathrm{PRESS}} = 0.58\end{aligned} \tag{8}$$

where $r^2_{33}$ is the correlation coefficient against 33 training compounds, $r^2_{LOO}$ is the leave-one-out correlation coefficient, $r^2_{BS}$ is the bootstrapping regression coefficient, and $r^2_{PRESS}$ is the predictive $r^2$ determined for the eight test

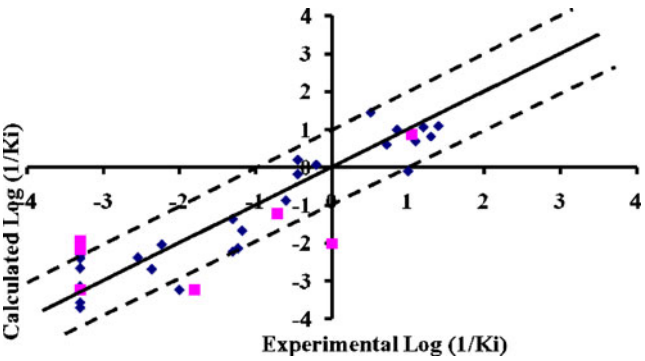



**Fig. 4** Experimental versus fitted (♦, 33 training compounds, $r^2_{LOO}$= 0.837) and predicted (■, 8 compounds, $r^2_{PRESS}$=0.578) bioactivities calculated from QSAR Eq. 8 for β-D-glucosidase inhibitors. *Solid lines* Regression lines for the fitted bioactivities, *dotted lines* 1.0 log point error margins

compounds [48, 53, 54]. Hypo10/1, Hypo4/4 and Hypo9/4 (Figs. 5, 6, 7, Table 2) represent the fit values of the training compounds against these three pharmacophores as calculated from Eqs. 6, 7 and 8. 2χ, 3χC and 3χP are the second order, third order cluster and path connectivity indices, respectively. ShadowXYfrac is one of the Shadow descriptors. Shadow descriptors are geometric descriptors that characterize the shape of the molecules; ShadowXYfrac represents the shadow of the molecule in the XY surface. JursFNSA3 and JursPPSA3 are Jurs charged partial surface

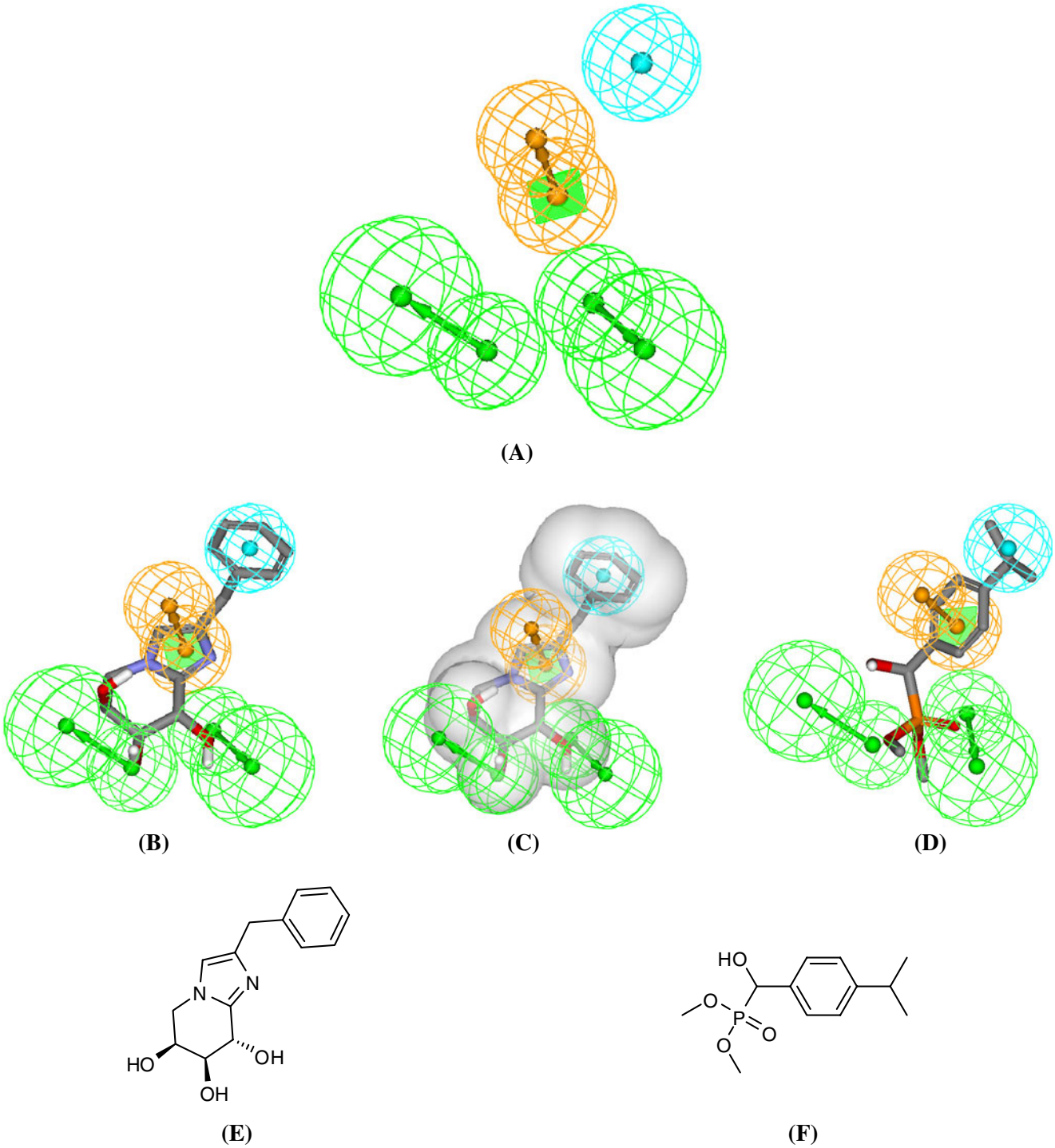



**Fig. 5** **a** Pharmacophoric features of the binding model Hypo10/1: *green vectored spheres* HBA, *blue spheres* Hbic, *orange vectored spheres* RingArom. **b** Hypo10/1 mapping the most potent training inhibitor **10** (Table 1 , Fig. 1, $K_i$=0.04 μM). **c** As **b**, with shape constraints. **d** Hypo10/1 fitted against discovered hit **42**. **e**, **f** Chemical structures of **10** and **42**, respectively

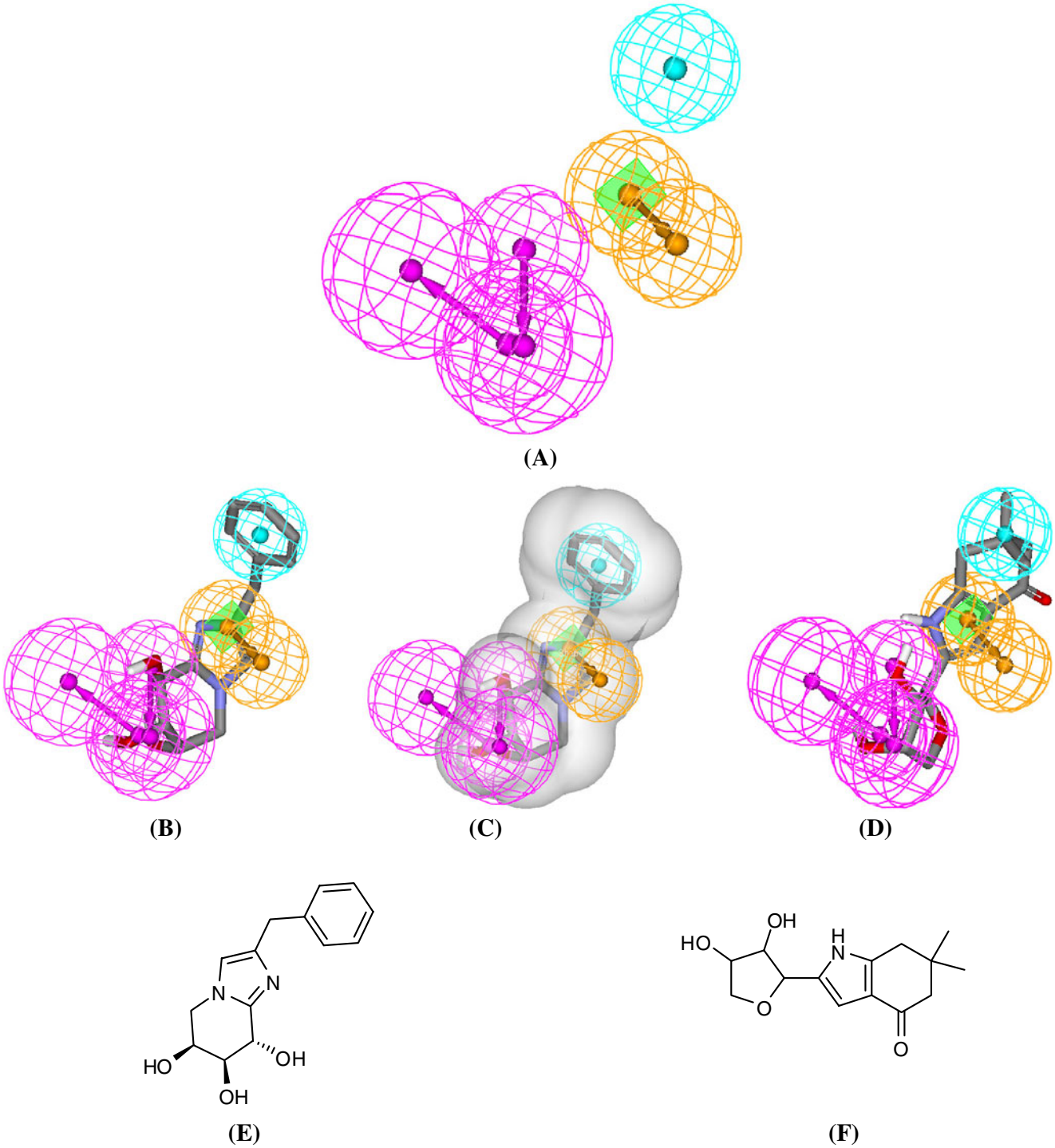



**Fig. 6** **a** Pharmacophoric features of the binding model Hypo4/4: *violet vectored spheres* HBD, *blue spheres* Hbic, *orange vectored spheres* RingArom. **b** Hypo4/4 mapping the most potent training inhibitor **10** (Table 1, Fig. 1, $K_i$=0.04 μM). **c** As **b** with shape constraints. **d** Hypo4/4 fitted against discovered hit **43**. **e**, **f** Chemical structures of **10** and **43**, respectively

area descriptors encoding the fractional charged partial surface area (obtained from) and atomic charge weighted positive surface area (=Σ[solvent-accessible surface area × partial charge for all positively charged atoms]), respectively. SaasN is an electrotopological sum descriptor for trisubstituted aromatic nitrogen atoms. AtypeH51 is atom-type-based AlogP descriptor [48, 55].

Several descriptors emerged in Eqs. 6 and 7 in spline format. The spline terms employed herein are "truncated power splines" and are denoted by bolded brackets ([ ]).

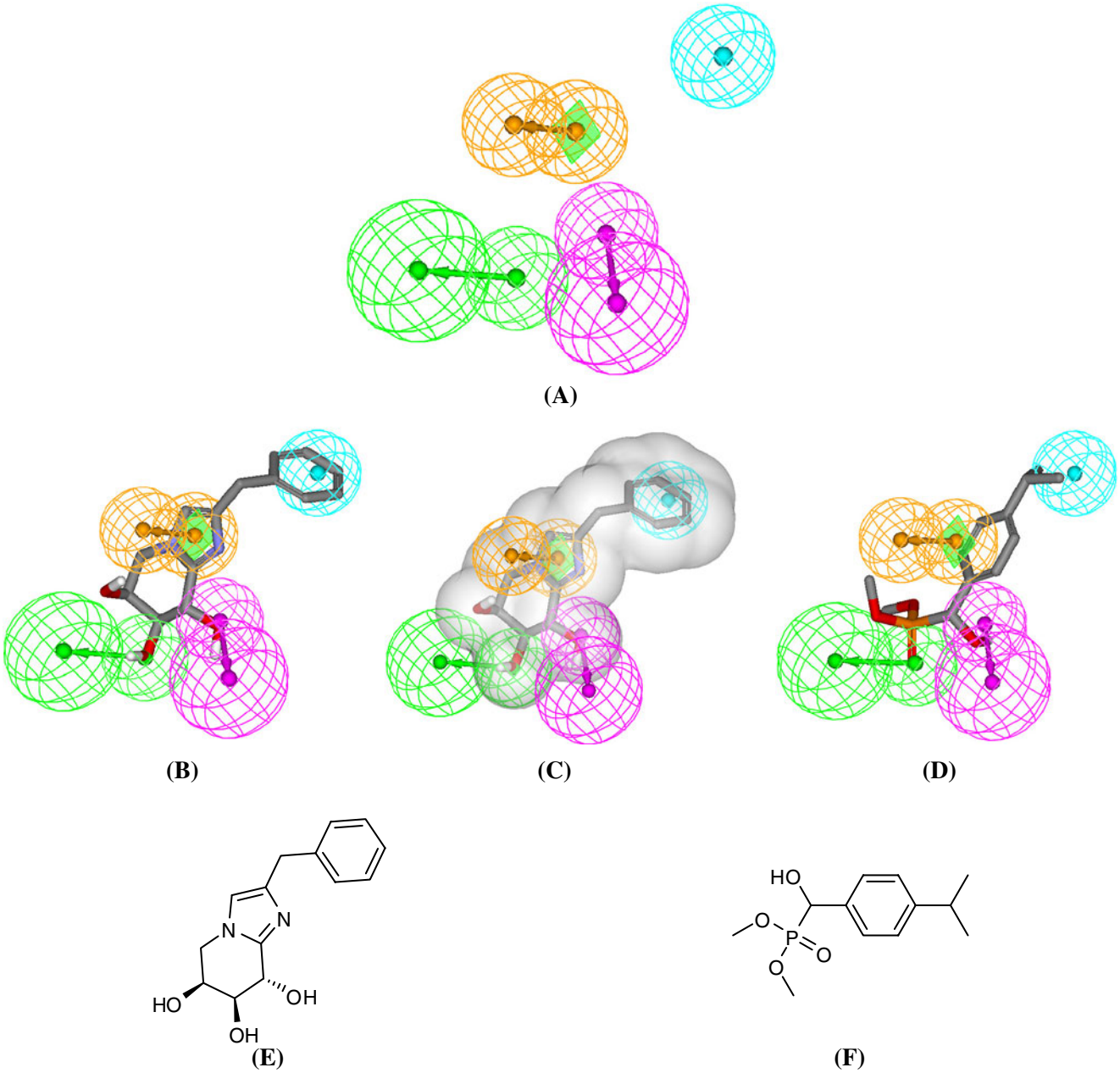



**Fig. 7** **a** Pharmacophoric features of the binding model Hypo9/4: *green vectored spheres* HBA, *violet vectored spheres* HBD, *blue spheres* Hbic, *orange vectored spheres* RingArom. **b** Hypo9/4 mapping the most potent training inhibitor **10** (Table 1, Fig. 1, $K_i$= 0.04 μM). **c** As **b** with shape constraints. **d** Hypo9/4 fitted against discovered hit **42**. **e**, **f** Chemical structures of **10** and **42**, respectively

For example, [f(x)−a] equals zero if the value of (f(x)−a) is negative; otherwise, it equals (f(x)−a) [48].

Interestingly, Hypo10/1 and Hypo4/4 emerged in Eqs. 6 and 7 in spline format, indicating that each binding mode contributes to ligand–β-D-glucosidase affinity only if the fit value of the particular ligand exceeds the corresponding spline threshold. For example, the ability of a certain ligand to map Hypo10/1 will impact its actual affinity to β-D-glucosidase only if its fit value exceeds 7.03 (the spline intercept associated with this pharmacophore in Eq. 6). Since the two spline cutoffs (of both pharmacophores) resemble high overall ligand/pharmacophore mapping (the maximum value is 14.0), it appears that ligand binding to β-D-glucosidase is sensitive to misalignments among the attracting moieties within the complex such that lowering the fits value below 7.03 and 8.03 for Hypo10/1 and Hypo4/4, respectively, nullifies any affinity gains from mapping the pharmacophores. A similar trend is also seen in Eq. 8, albeit in quadratic format. Emergence of Hypo9/4 in quadratic format in Eq. 8 suggests that ligand–β-D-

**Table 2** The pharmacophoric features and corresponding weights, tolerances and 3D coordinates of Hypo10/1, Hypo4/4 and Hypo9/4 generated for β-D-glucosidase inhibitors

| Model | Definitions | | Chemical features | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Hypo10/1[a] | | | HBA | | HBA | | Hbic | RingArom | |
| | Weights | | 2.89844 | | 2.89844 | | 2.89844 | 2.89844 | |
| | Tolerances | | 1.60 2.20 | 2.20 2.20 | 1.60 2.20 | 2.20 2.20 | 1.60 | 1.60 | 1.60 |
| | Coordinates | X | −5.13 | −5.22 | −2.32 | −2.92 | 3.78 | −0.29 | 0.45 |
| | | Y | 0.13 | −2.88 | 2.34 | 2.81 | −0.46 | 0.60 | −1.82 |
| | | Z | −1.46 | −1.13 | −1.09 | −4.02 | −1.24 | 0.65 | −0.96 |
| Hypo4/4[b] | | | HBD | | HBD | | Hbic Hbic | RingArom | |
| | Weights | | 2.98110 | | 2.98110 | | 2.98110 | 2.98110 | |
| | Tolerances | | 1.60 | 2.20 | 1.60 | 2.20 | 1.60 | 1.60 | 1.60 |
| | Coordinates | X | −2.83 | −5.34 | −5.11 | −4.38 | 2.79 | −0.22 | −1.33 |
| | | Y | −1.68 | −1.48 | −1.02 | −3.75 | −0.22 | 0.82 | 1.18 |
| | | Z | −1.18 | −2.81 | −1.03 | −0.02 | −2.17 | −0.38 | −3.15 |
| Hypo9/4[c] | | | HBA | | HBD | | Hbic Hbic | RingArom | |
| | Weights | | 2.70358 | | 2.70358 | | 2.70358 | 2.70358 | |
| | Tolerances | | 1.60 | 2.20 | 1.60 | 2.20 | 1.60 | 1.60 | 1.60 |
| | Coordinates | X | −4.90 | −6.88 | −1.90 | −2.87 | 4.30 | −0.43 | −1.38 |
| | | Y | −1.08 | 0.02 | −2.04 | −4.88 | 0.34 | 0.02 | −0.99 |
| | | Z | −0.91 | 1.06 | −1.32 | −1.33 | −1.28 | 0.79 | 3.45 |

[a] Hypo10/1: hypothesis number 10 generated in run number 1

[b] Hypo4/4: hypothesis number 4 generated in run number 4

[c] Hypo9/4: hypothesis number 9 generated in run number 4

glucosidase affinity is more sensitive to fitting the pharmacophoric models at higher fit values compared to lower values, i.e., misalignment among the attracting moieties within the complex drastically reduce ligand/β-D-glucosidase affinities, suggesting that ligand binding to β-D-glucosidase is extremely sensitive to minor misalignments between attracting moieties within the complex. Emergence of distinct pharmacophoric model in each QSAR equation suggests multiple binding modes assumed by different ligands within the binding pocket. Figures 5–7 show the pharmacophoric features of the binding models that emerged in QSAR Eqs. 6–8 and how they map the most potent training compound and the most potent discovered hit. Table 2 shows the corresponding X, Y, and Z coordinates of each pharmacophore.

Emergence of connectivity, electrotopological and shadow descriptors in Eqs. 6–8 is suggestive of certain role played by ligand topologies in the binding process. However, despite their predictive significance, the information content of these topological descriptors is quite obscure. On the other hand, emergence of JursFNSA3 and JursPPSA3 in Eqs. 6 and 8, respectively, suggests a direct

**Table 3** Performance of QSAR-selected pharmacophores and their shape complemented versions as 3D search queries

| Pharmacophore model | ROC-AUC | ACC | SPC | TPR | FNR |
|---|---|---|---|---|---|
| Hypo10/1 | 85.74 | 96.60 | 96.92 | 0.88 | 0.03 |
| Hypo4/4 | 89.48 | 96.60 | 96.92 | 0.88 | 0.03 |
| Hypo9/4 | 83.59 | 96.60 | 97.36 | 0.75 | 0.03 |
| Shape-complemented Hypo10/1 | 94.71 | 96.60 | 96.92 | 0.88 | 0.03 |
| Shape-complemented Hypo4/4 | 98.29 | 96.60 | 96.92 | 0.88 | 0.03 |
| Shape-complemented Hypo9/4 | 95.15 | 96.60 | 97.36 | 0.75 | 0.03 |

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

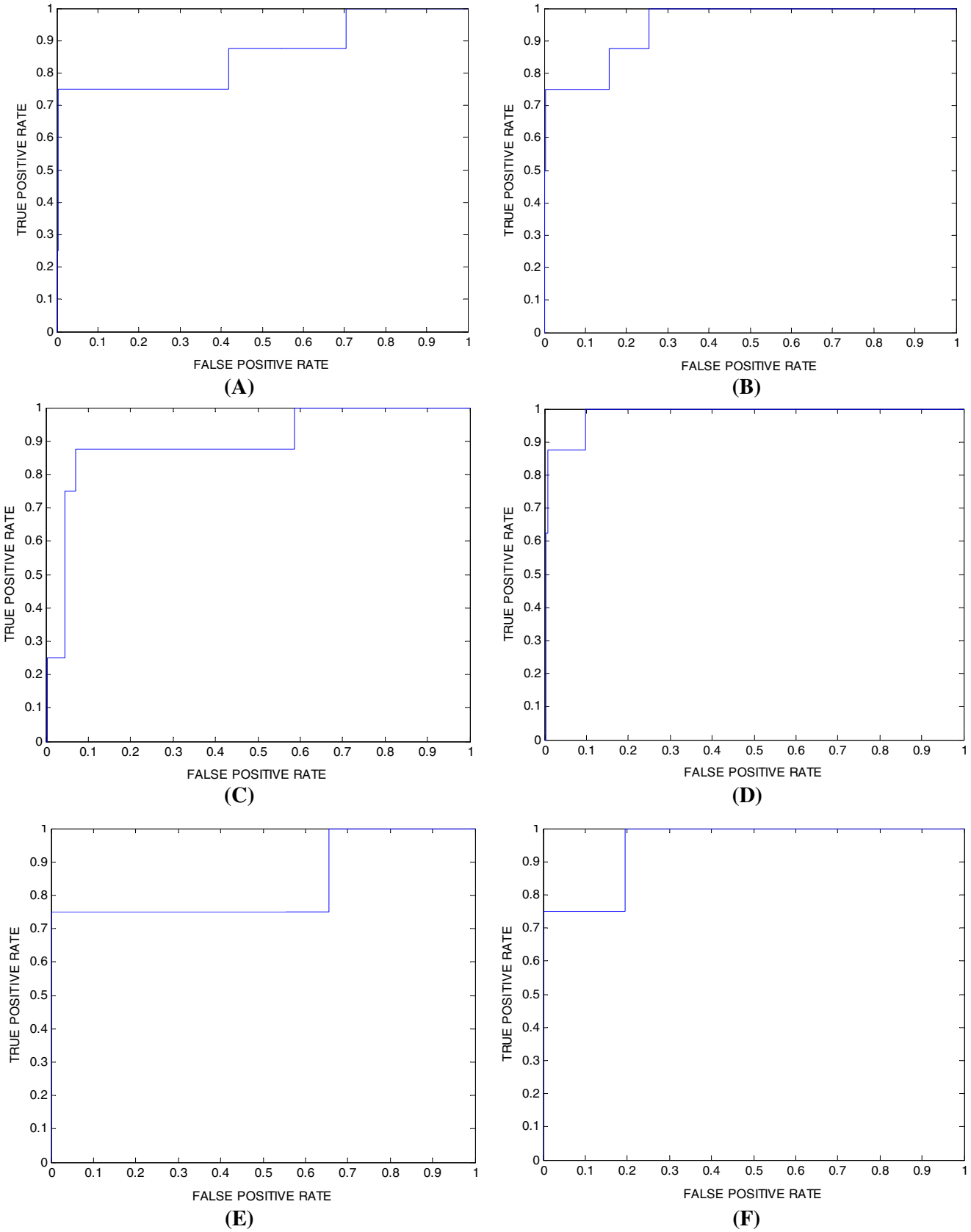



**Fig. 8** Receiver operating characteristic (ROC) curves conducted for QSAR-selected models. **a** Hypo10/1, **b** shape-complemented Hypo10/1 (Shape-Hypo10/1), **c** Hypo4/4, **d** Shape-Hypo4/4, **e** Hypo9/4 and **f** Shape-Hypo9/4

**Table 4** Number of compounds captured by shape-complemented β-D-glucosidase pharmacophores

| 3D Database | Post screening filtering[c] | Pharmacophore models[d] | | |
|---|---|---|---|---|
| | | Hypo10/1 | Hypo4/4 | Hypo9/4 |
| NCI[a] | Before | 1,960 | 441 | 764 |
| | After | 1,869 | 412 | 141 |
| DAC[b] | | 20 | 2 | 3 |

[a] National cancer institute list of available compounds (238,819 structures)

[b] In-house list of established drugs and agrochemicals (3,005 structures)

[c] Post screening filtering employing Lipinski's and Veber's rules. One Lipinski's violation was tolerated

[d] Number of hits captured by the sterically refined versions of the pharmacophore models

relationship between ligand/β-D-glucosidase affinity and ligand charges, which points to the hydrophilic nature of the binding sites.

## ROC curve analysis and shape constraints

To further validate the resulting models (both QSAR and pharmacophores), we subjected our QSAR-selected pharmacophores to ROC analysis, which tests the ability of a particular pharmacophore to classify a set of compounds as actives or decoys (see section on ROC analysis in Electronic Supplementary Material) [56–61]. The number of actives in the ROC testing list was limited to eight while the decoys were extended to 288 (i.e., each active compound was challenged with 36 decoys) to provide a proper challenge for the pharmacophoric models, particu-

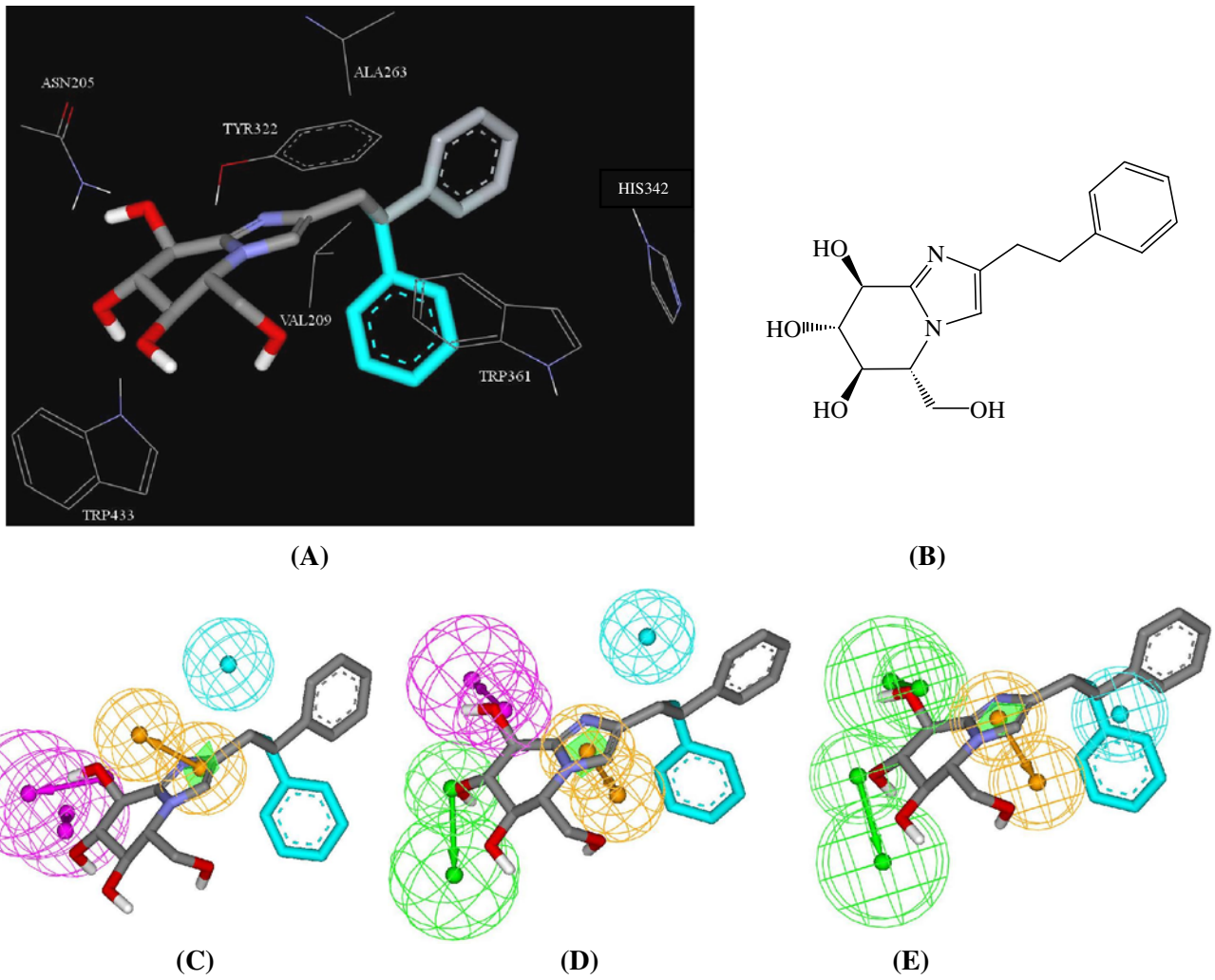



**Fig. 9** **a** Crystallographic structure of phenethylglucoimidazole cocrystallized with β-D-glucosidase (PDB code: 2CER, resolution: 2.2 Å). **b** Chemical structure of phenethyl-glucoimidazole. **c**, **d**, **e** Mapping the phenethyl-glucoimidazole against Hypo4/4, Hypo9/4 and Hypo10/1, respectively. *Blue-colored aromatic ring* indicates the dangling of this ring as reflected by the wide distribution of corresponding electron density

larly as they are required to select sterically and pharmacophorically challenging TSAs during in silico screening.

Table 3 and Fig. 8 show the ROC results from our QSAR-selected pharmacophores. In both the figure and the table, all QSAR-selected models clearly illustrate mediocre overall performance, with area under the curve (AUC) values ranging from 85.74% to 89.48%.

In order to enhance the ROC profiles of the QSAR-selected models, we decided to decorate them with shape-constraints derived from the most potent training inhibitor **10** ($K_i$=0.04 μM). Shape constraints encode the degree of 3D spatial similarity between screened compounds and the template ligand used to build the shape limitations [26, 40, 49]. To generate merged shape-pharmacophore queries, a selected potent training compound **10** was first fitted against the corresponding pharmacophore model; thereafter, the best-fitted conformer of the inhibitor was used to generate shape constraints that were subsequently merged with the pharmacophore. Figures 5, 6 and 7 show the shape-complemented versions of β-D-glucosidase pharmacophores.

Figure 8 and Table 3 show the ROC result of the shape-decorated versions of the QSAR-selected models. Clearly, the performance of shape-complemented models improved significantly as reflected by their ROC-AUC, which ranged from 94.71% to 98.29%.

The generated merged pharmacophore-shape queries were employed as 3D search queries against the NCI, drugs and agrochemical databases (Table 4).

## Comparing pharmacophore models with crystallographic complex

To further emphasize the validity of our pharmacophore/QSAR modeling approach, we compared the crystallographic structure of a β-D-glucosidase/ligand complex [62] (PDB code: 2CER, resolution: 2.2 Å) with Hypo4/4, Hypo9/4 and Hypo10/1. Figure 9 shows the chemical structure of the ligand and

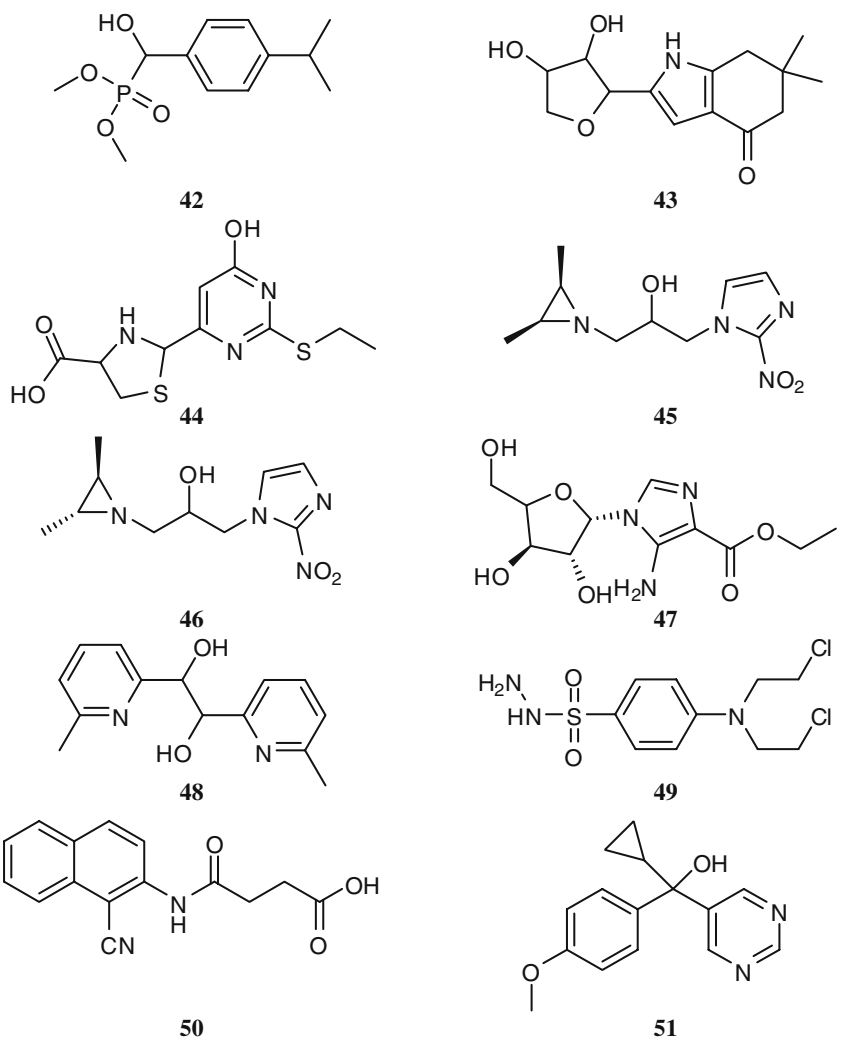



**Fig. 10** Chemical structures of the tested highest-ranking β-D-glucosidase hits (as suggested by the best QSAR models)

compares its β-D-glucosidase complex with the way it maps Hypo4/4, Hypo9/4 and Hypo10/1 employing rigid mapping, i.e., fitting the ligand's bound state against corresponding pharmacophores without conformational adjustments.

Pharmacophore mapping against the three models suggests that only two of the glucoimidazole hydroxyls are involved in hydrogen-bonding within the binding pocket, as in Figs. 9c–e, which seems to correlate with hydrogen-bonding interactions tying these hydroxyls with the amide and indole side chains of ASN205 and TRP433, respectively, as in Fig. 9a. Furthermore, mapping the imidazole ring of the phenethyl-glucoimidazole against RingArom feature in Hypo4/4, Hypo9/4 and Hypo10/1 agrees nicely with π–π stacking involving this ring and the phenolic side chain of TYR322, as in Fig. 9a. Finally, the terminal phenyl ring of the phenethyl-glucoimidazole seems to dangle in hydrophobic/aromatic pocket comprised of VAL209, TRP361, HIS342 and ALA263, as in Fig. 9a, which correlates with hydrophobic features in the three pharmacophores positioned in the same region.

Clearly from the above discussion, the three models, i.e., Hypo4/4, Hypo9/4 and Hypo10/1, represent close binding modes assumed by the ligand within β-D-glucosidase. These models point to a limited number of critical interactions required for high ligand-β-D-glucosidase affinity. In contrast, the crystallographic complex reveals many bonding interactions without highlighting critical ones. Incidentally, Fig. 9a shows only interactions corresponding to pharmacophoric features in the pharmacophores; other binding interactions are hidden for clarity.

## In-silico screening for β-D-glucosidase inhibitors and subsequent in vitro evaluation

Our QSAR-selected, shape-complemented pharmacophores, i.e., Hypo10/1, Hypo4/4 and Hypo9/4 (Eqs. 6–8), were employed as 3D search queries against two available 3D flexible structural databases, namely, the NCI list of compounds (238,819 compounds) and our in-house DAC database (3,005 compounds), to discover new inhibitory leads of alternative scaffolds against β-D-glucosidase.

Table 4 shows the number of hits captured by each pharmacophore model. Hits are defined as those compounds that have their chemical groups spatially overlap (map) with corresponding features within the particular pharmacophoric model. NCI hits were subsequently filtered based on Lipinski's and Veber's rules [50, 51]. Surviving hits were fitted against corresponding pharmacophores (without shape constraints) and their fit values against particular hypothesis were substituted in QSAR Eqs. 6–8 to determine their predicted bioactivities. The fact that the predicted Log ($1/K_i$) values exceeded the upper and lower bioactivity limits of the training compounds prompted us to employ bioactivity predictions merely to rank the corresponding hits in order to minimize the impact of any

**Table 5** QSAR estimated and in vitro bioactivities of β-D-glucosidase hit molecules captured by Hypo10/1, Hypo4/4 and Hypo9/4

| No.[a] | Name or NCI code | Fit values against[b] | | | QSAR-estimates | | Actual affinities[c] |
|---|---|---|---|---|---|---|---|
| | | Hypo10/1 | Hypo4/4 | Hypo9/4 | Log($1/K_i$) | $K_i$ (nM) | % Inhibition |
| **42** | 43458 | – | – | 7.0 | 0.804 | 0.157[h] | 28[d] |
| **43** | 345142 | – | 10.8 | – | −1.393 | 24.743[g] | 27[d] |
| **44** | 154916 | – | – | 9.7 | −3.078 | 1195.695[h] | 22[e] |
| **45** | 600667 | – | – | 9.0 | −2.178 | 150.821[h] | 33[e] |
| **46** | 601351 | – | – | 9.1 | −2.099 | 125.610[h] | 10[e] |
| **47** | 286625 | – | – | 9.6 | −1.598 | 39.610[h] | 10[e] |
| **48** | 60034 | 8.7 | – | – | −0.319 | 2.085[f] | 15[d] |
| **49** | 28578 | – | – | 9.9 | 0.633 | 0.233[h] | 22[d] |
| **50** | 263791 | 9.4 | – | – | 2.398 | 0.004[f] | 22[d] |
| **51** | Ancymidol | 8.5 | – | – | −3.456 | 2,856.145[f] | 16[e] |

[a] Hits shown in Fig. 10

[b] Best-fit values against corresponding binding hypothesis

[c] In vitro enzyme inhibition. Each value (percent inhibition) represents the average of at least two measurements

[d] % inhibition at 40 μM

[e] % inhibition at 100 μM

[f] Values estimated from Eq. 6

[g] Values estimated from Eq. 7

[h] Values estimated from Eq. 8

possible extrapolatory prediction errors on decisions regarding hits that merit subsequent in vitro testing [63].

The ten highest ranking anti-β-D-glucosidase hits were requested for experimental validation. One hit compound was the agrochemical ancymidol.

Figure 10 shows the chemical structures of tested hits, while Table 5 lists the corresponding experimental bioactivities and fit values of the tested hits against corresponding pharmacophore models.

The tested hits were evaluated by measuring the percentage of enzyme inhibition at 7, 10, 40 and 100 μM and by comparing the enzyme activity in the presence and absence of the particular hit.

Hits **42** (28% inhibition at 40 μM) and **43** (27% inhibition at 40 μM), Fig. 10, were most promising. The two compounds represent new anti-β-D-glucosidase inhibitory scaffolds of potential for subsequent optimization. Interestingly, all previously reported non-sugar β-D-glucosidase inhibitors failed to exceed mM inhibitory range, e.g., norbornane derivatives [64]. The significance of **42** and **43** is further supported by the fact that they were assayed against 0.16 units of β-D-glucosidase while previous published assays employed≤0.01 β-D-glucosidase units [65, 66]. Figures 5 and 7 show how **42** maps Hypo10/1 and Hypo9/4, respectively, while Fig. 6 shows how **43** maps Hypo4/4.

## Conclusions

Our results suggest that pharmacophore modeling combined with QSAR analysis can be a useful tool for the discovery of new scaffold of β-D-glucosidase inhibitors. The exploration of the pharmacophoric space of different β-D-glucosidase inhibitors was performed utilizing CATALYST-HYPOGEN to identify high quality binding model(s). Subsequently, QSAR analysis was employed to obtain a model that explains bioactivity variation. Genetic algorithm and MLR analyses were employed to access optimal QSAR model capable of explaining the inhibitory activity variation across 41 collected β-D-glucosidase inhibitors. Three successful β-D-glucosidase inhibitor pharmacophores emerged from three independent equations, suggesting the existence of more than one binding mode accessible to ligands within the β-D-glucosidase pocket. The QSAR equations and the associated pharmacophoric models were validated by ROC curve analysis and experimental identification of several β-D-glucosidase inhibitors retrieved from the NCI database and our own in-house structural database of established drug and agrochemicals.

**Acknowledgment** This work has been financially supported by the Deanship of Scientific Research at the University of Jordan; this support is highly acknowledged.

## References

1. Scofield AM, Witham P, Nash RJ, Kite GC, Fellows LE (1995) Castanospermine and other polyhydroxy alkaloids as inhibitors of insect glycosidases. Comp Biochem Phys 112A:187–196
2. Scofield AM, Witham P, Nash RJ, Kite GC, Fellows LE (1995) Differentiation of glycosidase activity in some Hemiptera and Lepidoptera by means of castanospermine and other polyhydroxy alkaloids. Comp Biochem Phys 112A:197–205
3. Gerber-Lemairer S, Juillerat-Jeanneret L (2006) Glycosylation pathways as drug targets for cancer: glycosidase inhibitors. Mini-Rev Med Chem 6:1043–1052
4. Lillelund VH, Jensen HH, Liang X, Bols M (2002) Recent developments of transition-state analogue glycosidase inhibitors of non-natural product origin. Chem Rev 102:515–553
5. Markad SD, Karanjule NS, Sharma T, Sabharwal SG, Dhavale DD (2006) Synthesis and evaluation of glycosidase inhibitory activity of N-butyl 1-deoxy-D-gluco-homonojirimycin and N-butyl 1-deoxy-L-ido-homonojirimycin. Bioorg Med Chem 14:5535–5539
6. Merrer YL, Gauzy L, Gravier-Pelletier C, Depezay JC (2000) Synthesis of C2-symmetric guanidino-sugars as potent inhibitors of glycosidases. Bioorg Med Chem 8:307–320
7. Robina I, Vogel P (2005) Synthesis of aza-C-disaccharides (dideoxyimino-alditols C-linked to monosaccharides) and analogues. Synthesis 5:675–702
8. Shitara E, Nishimura Y, Kojima F, Takeuchi T (1999) A facile synthesis of D-glucose-type gem-diamine 1-N-iminosugars: a new family of glucosidase inhibitors. Bioorg Med Chem 7:1241–1246
9. Asano A, Nash RG, Molyneuxc RJ, Fleet GWG (2000) Sugar-mimic glycosidase inhibitors: natural occurrence, biological activity and prospects for therapeutic application. Tetrahedron-Asymmetr 11:1645–1680
10. Asano N (2003) Glycosidase inhibitors: update and perspectives on practical use. Glycobiology 13:93–104
11. Berecibar A, Grandjean C, Siriwardena A (1999) Synthesis and biological activity of natural aminocyclopentitol glycosidase inhibitors: mannostatins, trehazolin, allosamidins, and their analogues. Chem Rev 99:779–844
12. Kim JH, Ryu YB, Kang NS, Lee BW, Heo JS, Jeong IY, Park KH (2006) Glycosidase inhibitory flavonoids from Sophora flavescens. Biol Pharm Bull 29:302–305
13. Li H, Schütz C, Favre S, Zhang Y, Vogel P, Sinay P, Blériot Y (2006) Nucleophilic opening of epoxyazepanes: expanding the family of polyhydroxyazepane-based glycosidase inhibitors. Org Biomol Chem 4:1653–1662
14. Pandey G, Dumbre SG, Khan MI, Shabab M (2006) Convergent approach toward the synthesis of the stereoisomers of C-6 homologues of 1-deoxynojirimycin and their analogues: evaluation as specific glycosidase inhibitors. J Org Chem 71:8481–8488
15. Schramm V (2003) Enzymatic transition state poise and transition state analogues. Acc Chem Res 36:588–596
16. Schramm V (2005) Enzymatic transition states and transition state analogues. Curr Opin Struct Biol 15:604–613
17. Amyes T, Richard J (2007) Rational design of transition-state analogues as potent enzyme inhibitors with therapeutic applications. ACS Chem Biol 2:711–714
18. Sutherland J, O'Brien L, Weaver D (2004) Pruned receptor surface models and pharmacophores for three-dimensional database searching. J Med Chem 47:3777–3787
19. Taha MO, Bustanji Y, Al-Ghussein M, Mohammad M, Zalloum H, Al-Masri IM, Atallah N (2008) 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

20. Al-masri IM, Mohammad K, Taha MO (2008) Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening. ChemMedChem 3:1763–1779
21. Taha MO, Dahabiyeh LA, Bustanji Y, Zalloum H, Saleh S (2008) 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
22. Taha MO, Atallah N, Al-Bakri AG, Paradis-Bleau C, Zalloum H, Younis KS, Levesque RC (2008) Discovery of new murf inhibitors via pharmacophore modeling and QSAR analysis followed by in silico screening. Bioorg Med Chem 16:1218–1235
23. 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 Graphics Model 25:870–884
24. Abu Hammad AM, Taha MO (2009) Pharmacophore modeling, quantitative structure-activity relationship analysis, and shape-complemented in silico screening allow access to novel influenza neuraminidase inhibitors. J Chem Inf Model 49:978–996
25. 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
26. Catalyst 4.11 User Guide (2005) Accelrys Software Inc, San Diego, CA
27. 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. VHCA, Basel, pp 223–240
28. Barnum D, Greene J, Smellie A, Sprague P (1996) Identification of common functional configurations among molecules. J Chem Inf Comput Sci 36:563–571
29. Smellie A, Teig S, Towbin P (1995) Poling: promoting conformational variation. J Comput Chem 16:171–187
30. Li H, Sutter J, Hoffmann R (2000) In: Güner OF (ed) Pharmacophore perception, development, and use in drug design. International University Line, La Jolla, CA, pp 173–189
31. Sutter J, Güner OF, 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, La Jolla, CA, pp 501–511
32. Kurogi Y, Güner O (2001) Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. Curr Med Chem 8:1035–1055
33. Bersuker IB, Bahçeci S, Boggs JE (2000) In: Güner OF (ed) Pharmacophore perception, development and use in drug design. International University Line, La Jolla, CA, pp 457–473
34. Poptodorov K, Luu T, Langer T, Hoffmann R (2006) In: Langer T, Hoffmann RD (eds) Methods and principles in medicinal chemistry, pharmacophores and pharmacophores searches, vol 2. WILEY-VCH, Weinheim, pp 17–47
35. Singh J, Chuaqui CE, Boriack-Sjodin PA, Lee WC, Pontz T, Corbley MJ, Cheung HK, 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
36. Taha MO, Qandil AM, Zaki DD, AlDamen MA (2005) Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling. Eur J Med Chem 40:701–727
37. Keller PA, Bowman M, Dang KH, Garner J, Leach SP, Smith R, McCluskey AJ (1999) Pharmacophore development for corticotropin-releasing hormone: new insights into inhibitor activity. J Med Chem 42:2351–2357
38. Karki RG, Kulkarni VM (2001) A feature based pharmacophore for *Candida albicans* MyristoylCoA: protein N-myristoyltransferase inhibitors. Eur J Med Chem 36:147–163
39. Taha MO, Al-Bakri AG, Zalloum WA (2006) Discovery of potent inhibitors of pseudomonal quorum sensing via pharmacophore modeling and in silico screening. Bioorg Med Chem Lett 16:5902–5906
40. Moffat K, Gillet VJ, Whittle M, Bravi G, Leach AR (2008) A comparison of field-based similarity searching methods, Cat-Shape, FBSS, and ROCS. J Chem Inf Model 48:719–729
41. Dubost E, Tschamber T, Streith J (2003) Increasing the inhibitory potency of L-arabino-imidazolo-[1, 2]-piperidinose towards β-D-glucosidase and β-D-galactosidase. Tetrahedron Lett 44:3667–3670
42. Dubost E, Nouën DL, Streith J, Tarnus C, Tschamber T (2006) Synthesis of substituted Imidazolo[1, 2-a] piperidinoses and their evaluation as glycosidase inhibitors. Eur J Org Chem 2006:610–626
43. Frankowski A, Deredas D, Dubost E, Gessier F, Jankowski S, Neuburger M, Seliga C, Tschamber T, Weinberg K (2003) Stereocontrolled synthesis of imidazolo[1, 5]hexopiperidinoses and imidazol-4(5)-yl-C-glycosides. Tetrahedron 59:6503–6520
44. Gessier F, Tschamber T, Tarnus C, Neuburger M, Huber W, Streith J (2001) Synthesis of imidazolo-piperidinopentoses as nagstatine analogues. Eur J Org Chem 2001:4111–4125
45. Tschamber T, Gessier F, Dubost E, Newsome J, Tarnus C, Kohler J, Neuburger M, Streith J (2003) Carbohydrate transition state mimics: synthesis of imidazolo-pyrrolidinoses as potential nectrisine surrogates. Bioorg Med Chem 11:3559–3568
46. Fisher R (1966) The principle of experimentation illustrated by a psycho-physical experiment, 8th edn. Hafner, New York
47. Krovat EM, Langer T (2003) Non-peptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification. J Med Chem 46:716–726
48. CERIUS2, Version 4.10. QSAR Users' Manual (2005) Accelrys Inc, San Diego, CA, pp 221–235
49. Hahn M (1997) Three-dimensional shape-based searching of conformationally flexible compounds. J Chem Inf Comput Sci 37:80–86
50. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26
51. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623
52. Sheridan RP, Kearsley SK (2002) Why do we need so many chemical similarity search methods. Drug Discovery Today 7:903–911
53. Ramsey LF, Schafer WD (1997) The statistical sleuth, 1st edn. Wadesworth, Belmont, CA
54. Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR & Comb Sci 22:69–77
55. Sivakumar PM, Babu SKG, Doble M (2008) Impact of topological and electronic descriptors in the QSAR of pyrazine containing thiazolines and thiazolidinones as antitubercular and antibacterial agents. Chem Biol Drug Des 71:447–463
56. 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
57. 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 Des 22:213–228

58. Irwin JJ, Shoichet BK (2005) ZINC—a free database of commercially available compounds for virtual screening. J Chem Inf Comput Sci 45:177–182
59. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO (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
60. 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
61. Gao H, Williams C, Labute P, Bajorath J (1999) Binary quantitative structure-activity relationship (QSAR) analysis of estrogen receptor ligands. J Chem Inf Comput Sci 39:164–168
62. Gloster TM, Roberts S, Perugino G, Rossi M, Moracci M, Panday N, Terinek M, Vasella A, Davies GJ (2006) Structural, kinetic, and thermodynamic analysis of glucoimidazole-derived glycosidase inhibitors. Biochemistry 45:11879–11884
63. Cronin MTD, Schultz TW (2003) Pitfalls in QSAR. J Mol Struct 622:39–51
64. Buser S, Vasella A (2006) Norbornane mimics of distorted β-D-glucopyranosides inhibitors of β-D-glucopyranosidases. Helv Chim Acta 89:614–621
65. Pabba J, Vasella A (2006) Probing the interaction of the C(4) hydroxy group of lactone-type inhibitors with beta-glucosidases and beta-galactosidases. Helv Chim Acta 89:2006–2019
66. Falshaw A, Hart JB, Tyler PC (2000) New syntheses of 1D- and 1L-1, 2-anhydro-myo-inositol and assessment of their glycosidase inhibitory activities. Carbohydr Res 329:301–308