

Original article

Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling

Mutasem O. Taha ^{a,*}, Amjad M. Qandil ^b, Dhia D. Zaki ^a, Murad A. AlDamien ^c

^a Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan

^b Department of Medicinal Chemistry and Pharmacognosy, Faculty of Pharmacy, Jordan University of Science and Technology, P.O. Box 3030, Irbid 22110, Jordan

^c Department of Chemistry, Faculty of Science, University of Jordan, Amman, Jordan

Received 15 July 2004; accepted 11 October 2004

Available online 04 February 2005

Abstract

The flexibility of activated factor X (fXa) binding site was assessed employing ligand-based pharmacophore modeling combined with genetic algorithm (GA)-based QSAR modeling. Four training subsets of wide structural diversity were selected from a total of 199 direct fXa inhibitors and were employed to generate different fXa pharmacophoric hypotheses using CATALYST® software over two subsequent stages. In the first stage, high quality binding models (hypotheses) were identified. However, in the second stage, these models were refined by applying variable feature weight analysis to assess the relative significance of their features in the ligand-target affinity. The binding models were validated according to their coverage (capacity as a three-dimensional (3D) database search queries) and predictive potential as three-dimensional quantitative structure–activity relationship (3D-QSAR) models. Subsequently, GA and multiple linear regression (MLR) analysis were employed to construct different QSAR models from high quality pharmacophores and explore the statistical significance of combination models in explaining bioactivity variations across 199 fXa inhibitors. Three orthogonal pharmacophoric models emerged in the optimal QSAR equation suggesting they represent three binding modes accessible to ligands in the binding pocket within fXa.
© 2005 Elsevier SAS. All rights reserved.

Keywords: Active site flexibility; Activated factor X inhibitors; Pharmacophore modeling; Genetic algorithm; Multiple linear regression; QSAR

1. Introduction

Thrombosis-related ischemic diseases are leading causes of death in the world. Unfortunately, these diseases are still treated by relatively antiquated drugs. However, an exciting new wave of antithrombotic compounds has recently emerged in clinical trials. A particularly attractive new class of antithrombotic agents is the direct fXa inhibitors, which appear to provide an enhanced risk-to-benefit margin compared to conventional anticoagulation therapies [1,2].

Factor X (fX), which is responsible for the activation of prothrombin, has long been recognized to play a significant role in hemostasis [3]. Studies on activated factor X (fXa) deficiency have indicated that fXa activity must be less than

5% of its normal levels in order to result in spontaneous bleeding tendencies [4]. This suggests that fXa activity can be suppressed markedly without affecting hemostasis, which is a desirable characteristic for anticoagulants of clinical potential. Selective fXa inhibitors appear to reduce thrombin generation at the thrombus [5], to prevent thrombus growth [6] and even to dissolve formed thrombi, i.e. “dethrombosis” [1,7]. A growing number of reviews [8–13] and publications [14–25] have demonstrated the recent interest in potent and selective fXa inhibitors.

fXa is a two-chain macromolecule, consisting of a γ -carboxyglutamic acid (Gla) domain followed by two epidermal growth factor (EGF)-like repeats (the light chain) joined by a disulphide bond to a serine protease domain (the heavy chain) [12,26–28].

The binding site of fXa includes two main regions named the S1 and S4 pockets and several side pockets involved in substrate recognition. The S1 pocket is a narrow cleft that

* Corresponding author. Tel.: +962 6 535 5000x2505; fax: +962 6 533 9649.

E-mail addresses: mutasem@ju.edu.jo (M.O. Taha), drqandil@just.edu.jo (A.M. Qandil).

extends about 8 Å deep into the core of the protein. It is bordered by planar hydrophobic walls and terminates with a negatively charged aspartate [27,28]. On the other hand, the S4 pocket consists of a surface cleft characterized by hydrophobic aromatic (HbicArom) floor and walls. It usually accommodates hydrophobic residues, however, it has been suggested that it might participate in favorable interactions with a positive charge [28–31].

A typical direct fXa inhibitor includes three molecular fragments. (i) A positively charged group (P1) intended to fit into the S1 pocket [14,31–33]. (ii) An aromatic fragment (P4) intended to interact with the S4 pocket [14,26,34,35]. (iii) A central linker designed to project the substituents appropriately into their corresponding pockets [9,31,36–40].

The extensive structural diversity of known direct fXa inhibitors [9,32,36,41–43] suggests the existence of discrete multiple binding modes within fXa binding site to provide complementary pockets for such diverse ligands. Furthermore, a recently published crystallographic study revealed the existence of two binding modes adopted by four different ligands within the S4 pocket of fXa [14]. The apparent flexibility of fXa binding site has prompted an attempt to identify corresponding accessible binding modes employing molecular dynamics simulation [44].

Obviously, using a single protein conformation for designing new ligands ignores important dynamic aspects of protein–ligand binding. In particular, the “induced fit” effects are ignored [45,46]. Furthermore, binding site flexibility estimates may be of direct use in ligand design for they indicate which parts of the ligand may tolerate variations in size, substituents and conformations [47]. Moreover, conformational flexibility and strain energy of the ligand play critical roles in affinity interactions. It has been shown that, on average, each freely rotating bond in a ligand reduces binding free energy by 0.7 kcal/mol [48]. Many examples are known where rigidification of a flexible ligand causes a substantial boost in affinity. However, rigidifying a flexible ligand into the wrong conformation will produce substantially less active or inactive compound [49].

Current computational models describing binding site flexibility can be roughly divided into four different categories: (a) use of soft receptors that relax energetic penalties due to steric clashes [49,50]. (b) Selection of few critical rotatable degrees of freedom in the receptor binding site [50–55]. (c) Systematic conformer searches of amino acids’ side chains at the binding site followed by molecular mechanical energy evaluation [56]. (d) Molecular dynamics and free energy calculations conducted on flexible enzyme, ligand structures and explicit solvent molecules [57,58]. (e) Use of multiple crystallographic receptor structures either individually or by combining them using an averaging scheme [50,59].

Nevertheless, the problem of docking or designing fully flexible ligands to match fully flexible receptors remains daunting [50,60]. We believe that the current computational strategies directed towards assessing the flexibility of macromolecular binding sites suffer from two major drawbacks.

Firstly, their computational cost, which reduces their effectiveness in virtual screening and fast docking. Secondly, their complete reliance on crystallographic structures. Although crystallographic data are considered the most reliable structural information that can be used for drug design, they are associated with some serious problems: (i) decisions whether to leave buried water molecules in the binding site or not have to be taken into consideration prior to ligand design and/or docking studies [49]. (ii) Inadequate resolution of crystallographic structures. The most reliable resolution for structure-based drug design has been recently recommended to be below 1.5 Å [61]. Incidentally, reported fXa structures exhibited mediocre resolutions (1.9–2.7 Å) [12,14,27,28,31,62]. (iii) Crystallization-related artifacts, e.g. structural distortions inflicted upon the three-dimensional (3D) structure of the ligand–protein complex during crystallization [63]. (iv) Crystallographic structures lack information on hydrogen atoms, so it should be appropriately assumed whether ionizable moieties exposed in the active site exist in their ionized form or not prior to in-silico design [49,64].

The great recent interest in designing new fXa inhibitors, combined with the drawbacks of structure-based flexibility assessment methods prompted us to evaluate an interesting and novel ligand-based approach as a tool for characterizing fXa binding site flexibility. The proposed approach was carried out over two subsequent stages. Firstly, the pharmacophoric space of the targeted enzyme (i.e. fXa) was extensively explored utilizing the three-dimensional quantitative structure–activity relationship (3D-QSAR) software program CATALYST[®] [65–67]. In the subsequent phase, genetic algorithm (GA) and multiple linear regression (MLR) analyses were employed to select an optimal combination of orthogonal pharmacophores that best explain the observed bioactivities, i.e. best possible QSAR model. Such combination of pharmacophores should correspond to accessible binding modes available for ligands in the fXa binding pocket.

CATALYST[®] models drug–receptor interaction using information derived only from the drug structure [67–69]. It automatically generates pharmacophore models that correlate the biological activity observed for a series of compounds to their chemical structures. Molecules are described as collection of chemical functionalities arranged in 3D space. The conformational flexibility of training ligands is modeled by creating multiple conformers, judiciously prepared to emphasize representative coverage over a specified energy range. A set of chemical features common to the input molecules (training set) that correlates best with the biological activity is determined. This 3D array of chemical features (known as hypothesis) 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 (HBDs) and acceptors, aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionizable groups and aromatic planes. Successful examples involving the use of CATALYST[®] have been reported, wherein the CATALYST[®] derived pharmacophore has been

used efficiently as a query for database searching and 3D-QSAR studies [70–73].

2. Modeling methods

2.1. Software and hardware

The following software packages were utilized in the present research.

- CATALYST 4.6, Accelrys Inc. (<http://www.accelrys.com>), USA.
- Alchemy 2000, version 2.05, Tripos Inc. (<http://www.tripos.com>), USA.
- CS ChemDraw Ultra 6.0, Cambridge Soft Corp. (<http://www.cambridgesoft.com>), USA.
- QSARIS, Scivision Inc. (<http://www.scivision.com>), USA. Now a property of MDL Inc., Germany.

Pharmacophor modeling studies were performed using CATALYST[®] 4.6 installed on a Silicon Graphics O2 desktop workstation equipped with a 300 MHz MIPS R12000 processor (256 MB RAM) running the Irix 6.5 operating system. However, various physicochemical descriptors were calculated employing Alchemy2000[®] and QSARIS[®] installed on a Pentium III personal computer running Microsoft Windows 98.

2.2. Data set

The structures of 199 diverse fXa inhibitors were collected from recently published literature [9,32,41–43]. The corresponding bioactivities are expressed as the dissociation constants of the enzyme-inhibitor complexes (K_i values). Table 1 and Fig. 1 illustrate the collected structures and corresponding K_i values.

2.3. Molecular modeling

The two dimensional (2D) chemical structures of the inhibitors were sketched using ChemDraw Ultra (installed on a Pentium III personal computer) and saved in MDL-molfile format. Subsequently, these 2D structures 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, and were stored in MDL-molfile format for calculating variety of physicochemical properties by QSARIS[®] and Alchemy2000[®].

2.3.1. Conformational analysis

Molecular flexibility was 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 (1–199, Table 1 and Fig. 1) was explored adopting the “Best conformer generation” option within CATALYST[®], which is based on the generalized CHARMM force field implemented in the program. Default parameters were employed in the conformation generation procedure, i.e. conformational ensembles were generated with an energy threshold of 20 kcal/mol from the local minimized structure and a maximum limit of 250 conformers per molecule. Hence, this search procedure will probably identify the best 3D arrangement of chemical functionalities explaining the activity variations among the training set [65,66,80,81].

2.3.2. Generation of pharmacophoric hypotheses

All 199 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 [67,74]. The uncertainty values should minimize regression errors resulting from slight differences in bioassay conditions adopted by different research groups.

Subsequently, four structurally diverse training subsets (Table 2) were carefully selected from the collection for pharmacophore modeling. Typically, CATALYST[®] requires informative training sets that include at least 16 compounds of evenly spread bioactivities over at least four orders of magnitude. Lesser training lists could lead to chance correlation and thus faulty models. On the other hand, large training sets (>45 compounds) may render the computation time excessively lengthy without significant improvements in the resulting models [67,74].

The selected training sets were utilized to conduct nine modeling runs to explore the pharmacophoric space of fXa inhibitors (see Section 3.2). Different hypotheses were generated by altering feature selection and number of allowed features in the resulting pharmacophors (Tables 2 and 3). Furthermore, CATALYST[®] was allowed to explore the weights of individual features in three trials (runs 7, 8 and 9 in Table 3).

Pharmacophor modeling employing CATALYST[®] proceeds through three successive phases: the constructive phase, subtractive phase and optimization phase [67,74]. During the constructive phase, CATALYST[®] generates common conformational alignments among the most-active 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) [67,74]:

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

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 ques-

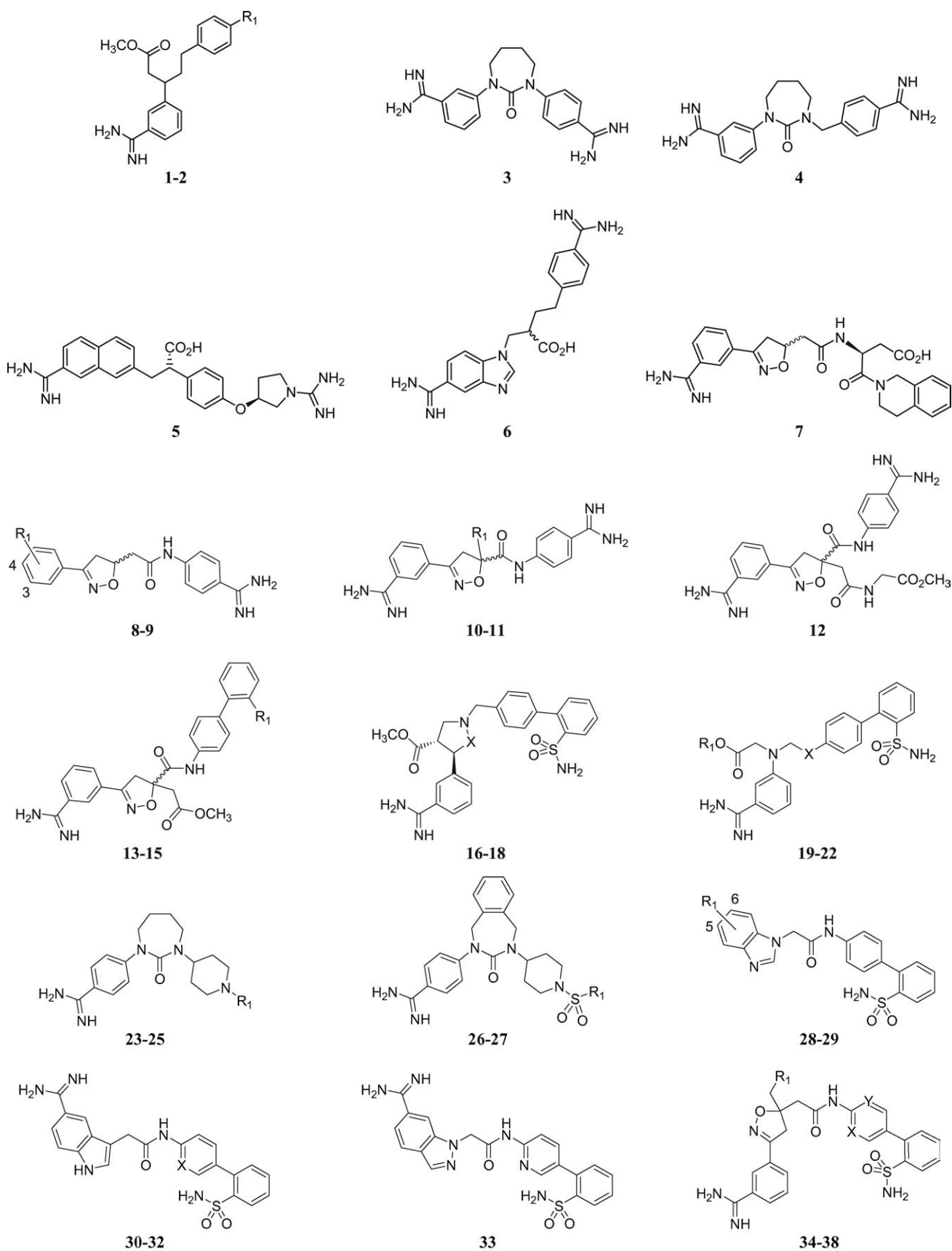


Fig. 1. Mined fXa inhibitors.

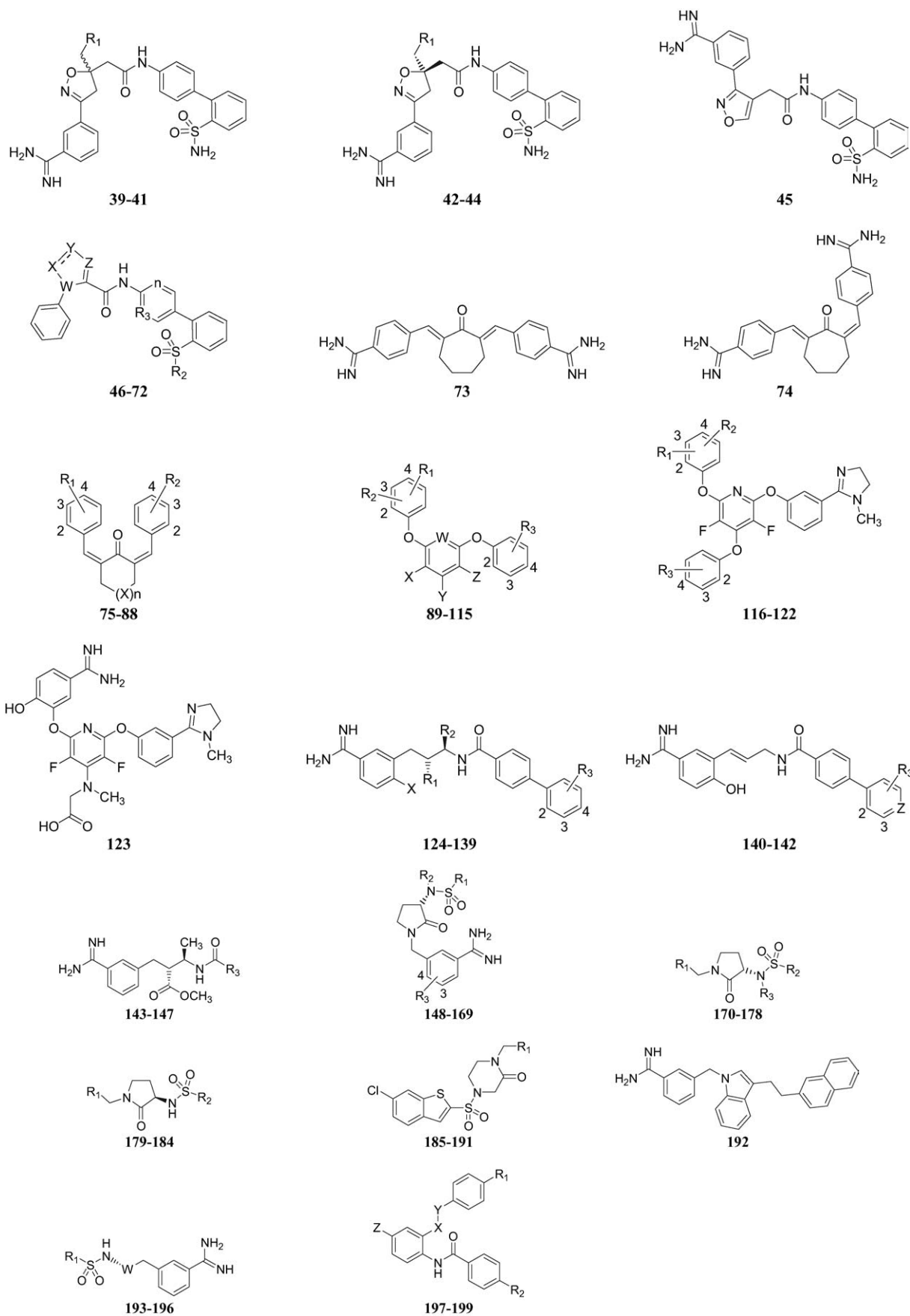


Fig. 1. (continued)

Table 2

Training subsets employed in exploring the pharmacophoric space of fXa inhibitors, numbers correspond to compounds in Table 1 and Fig. 1

| Training set | Most-active subset ^a | Intermediate subset | Least active subset ^b |
|--------------|-------------------------------------|--|---|
| I | 26, 31, 48, 75, 123, 142 | 4, 23, 74, 108, 147, 150, 193, 198, 199 | 85, 89, 111, 159, 192, 197 |
| II | 18, 26, 71, 113, 120, 122, 130, 168 | 4, 5, 24, 28, 100, 102, 110, 126, 148, 149, 163, 169, 172, 175, 177, 182 | 3, 20, 99, 104, 111, 114, 133, 137, 158, 161, 170, 192, 194, 195, 197 |
| III | 24, 31, 72, 113, 117, 121, 134, 142 | 4, 5, 16, 17, 27, 28, 95, 100, 102, 110, 126, 148, 149, 163, 169, 172, 175, 177, 182 | 3, 20, 99, 104, 111, 114, 133, 137, 158, 161, 170, 192, 194, 195, 197 |
| IV | 18, 26, 71, 113, 120, 122, 130, 168 | 4, 5, 24, 28, 49, 68, 93, 95, 100, 102, 105, 110, 126, 136, 144, 148, 149, 160, 163, 167, 169, 172, 175, 177, 182, 187 | 3, 20, 99, 104, 111, 114, 133, 137, 158, 161, 170, 192, 194, 195, 197 |

^a Potency categories as defined by Eqs. (1) and (2).^b Potency categories as defined by Eqs. (1) and (2).

Table 3

Training sets and CATALYST[®] run parameters employed in exploring fXa pharmacophoric space

| Run number | Training set ^a | Number of training compounds | Selected input features: types and ranges ^b | Other run parameters ^{c,d} |
|------------|---------------------------|------------------------------|--|-------------------------------------|
| 1 | I | 21 | HBD (0–3), HBA (0–3), AromRing (0–3), PosIoniz (0–3), HbicArom (0–3) | Min–max: 4–5 |
| 2 | I | 21 | HBD (0–3), HBA (0–3), AromRing (0–3), PosIoniz (0–3), HBicArom (0–3) | Min–max: 5–5 |
| 3 | II | 39 | HBD (0–3), HBA (0–3), AromRing (0–3), PosIoniz (0–3), HbicArom (0–3) | Min–max: 4–5 |
| 4 | II | 39 | HBD (0–3), HBA (0–3), AromRing (0–3), PosIoniz (0–3), HbicArom (0–3) | Min–max: 5–5 |
| 5 | III | 42 | HBD (0–3), HBA (0–3), AromRing (0–3), PosIoniz (0–3), HbicArom (0–3) | Min–max: 4–5 |
| 6 | III | 42 | HBD (0–3), HBA (0–3), AromRing (0–3), PosIoniz (0–3), HbicArom (0–3) | Min–max: 5–5 |
| 7 | IV | 49 | HBA (1–1), HbicArom (2–2), AromRing (1–1), PosIoniz (1–1) | Min–max: 5–5, VW |
| 8 | IV | 49 | HBA (2–2), HbicArom (2–2), PosIoniz (1–1) | Min–max: 5–5, VW |
| 9 | IV | 49 | HBA (2–2), AromRing (1–1), PosIoniz (1–1) | Min–max: 4–4, VW |

^a The numbers correspond to training sets in Table 2.^b HBA: Hydrogen Bond Acceptor, HBD: Hydrogen Bond Donor, AromRing: Aromatic Ring, HbicArom: Hydrophobic Aromatic, PosIoniz: Positive Ionizable. The allowed ranges of input features are in brackets.^c Min–max: allowed minimum and maximum number of output features, VW: allowed variable feature weight.^d The spacing parameter was set to 10 pm in all runs, unmentioned parameters were set to their default values.

tion. However, if there are more than eight most-active inhibitors, only the top eight are used.

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) [67,74]:

$$\log(\text{Act}) - \log(\text{MAct}) > 3.5 \quad (2)$$

However, in the optimization phase, CATALYST[®] applies fine perturbations in the form of vectored feature rotation, adding new feature and/or removing a feature, to selected hypotheses that survived the subtractive phase, in an attempt to find new models of enhanced bioactivity/mapping correlation, i.e. improved 3D-QSAR properties. Eventually, CATALYST[®] selects the highest-ranking models (10 by default) and present them as the optimal pharmacophore hypotheses resulting from the particular automatic modeling run.

The followings are important CATALYST[®] control parameters used for hypotheses generation [66]:

- **MinPoints:** The default value of this parameter is 4, specifying a minimum of 4 individual feature components for a generated hypothesis.
- **MinSubset points:** Only configurations of features in input molecules, with at least the number of points specified by this option, are considered when identifying a candidate hypothesis. The default value is 4.
- **Superposition error, check superposition and tolerance factor:** The three control parameters together check the superposition of compounds for hypothesis generation. All three have a default value of 1. Reducing the value from its default tightens the fit.
- **Feature misses:** This specifies the number of compounds allowed not to map any particular feature in a generated hypothesis. The default value is 1.
- **Mapping coefficient:** This parameter controls the importance of having compounds with similar structure map to a hypothesis in a similar way. Increasing this parameter will penalize the hypotheses that deviate from this behavior. The default value is 0.

Table 1
Mined fXa inhibitors and their corresponding K_i values (nM)

| Compound | R ₁ | R ₂ | R ₃ | n | W | X | Y | Z | Activity K_i (nM) |
|----------|------------------------------------|-----------------|----------------|----|---|---------------------------------|-------------------|-------|---------------------|
| 1 | Amidino | – | – | – | – | – | – | – | 34 |
| 2 | Guanidino | – | – | – | – | – | – | – | 9 |
| 3 | – | – | – | – | – | – | – | – | 800 |
| 4 | – | – | – | – | – | – | – | – | 61 |
| 5 | – | – | – | – | – | – | – | – | 41 |
| 6 | – | – | – | – | – | – | – | – | 140 |
| 7 | – | – | – | – | – | – | – | – | 38,500 |
| 8 | 4-Amidino | – | – | – | – | – | – | – | 1700 |
| 9 | 3-Amidino | – | – | – | – | – | – | – | 1400 |
| 10 | H | – | – | – | – | – | – | – | 270 |
| 11 | CH ₂ CO ₂ Me | – | – | – | – | – | – | – | 94 |
| 12 | – | – | – | – | – | – | – | – | 18 |
| 13 | H | – | – | – | – | – | – | – | 200 |
| 14 | SO ₂ NH ₂ | – | – | – | – | – | – | – | 6.3 |
| 15 | SO ₂ Me | – | – | – | – | – | – | – | 4.2 |
| 16 | – | – | – | – | – | CH ₂ | – | – | 10 |
| 17 | – | – | – | – | – | CH ₂ CH ₂ | – | – | 1.3 |
| 18 | – | – | – | – | – | CH ₂ O | – | – | 0.53 |
| 19 | Me | – | – | – | – | CH ₂ | – | – | 3.4 |
| 20 | H | – | – | – | – | CH ₂ | – | – | 600 |
| 21 | Me | – | – | – | – | O | – | – | 3.1 |
| 22 | <i>i</i> -Pr | – | – | – | – | O | – | – | 1.8 |
| 23 | Benzyl | – | – | – | – | – | – | – | 102 |
| 24 | Benzylsulfonyl | – | – | – | – | – | – | – | 10 |
| 25 | Thiophen-2-sulfonyl | – | – | – | – | – | – | – | 12 |
| 26 | Benzyl | – | – | – | – | – | – | – | 0.8 |
| 27 | 2-Thiophene | – | – | – | – | – | – | – | 3.4 |
| 28 | 5-Amidino | – | – | – | – | – | – | – | 400 |
| 29 | 6-Amidino | – | – | – | – | – | – | – | 2.4 |
| 30 | – | – | – | – | – | CH | – | – | 2.1 |
| 31 | – | – | – | – | – | CBr | – | – | 0.32 |
| 32 | – | – | – | – | – | N | – | – | 0.44 |
| 33 | – | – | – | – | – | – | – | – | 1.2 |
| 34 | CO ₂ Me | – | – | – | – | CH | N | – | 0.83 |
| 35 | CO ₂ Me | – | – | – | – | N | N | – | 2.3 |
| 36 | OEt | – | – | – | – | CH | N | – | 0.8 |
| 37 | <i>N</i> -tetrazolyl | – | – | – | – | CH | CH | – | 0.52 |
| 38 | <i>N</i> -tetrazolyl | – | – | – | – | CH | N | – | 0.11 |
| 39 | <i>N</i> -tetrazolyl | – | – | – | – | – | – | – | 1.6 |
| 40 | Me | – | – | – | – | – | – | – | 11 |
| 41 | OMe | – | – | – | – | – | – | – | 3.4 |
| 42 | <i>N</i> -tetrazolyl | – | – | – | – | – | – | – | 2.3 |
| 43 | Me | – | – | – | – | – | – | – | 1.8 |
| 44 | OMe | – | – | – | – | – | – | – | 0.55 |
| 45 | – | – | – | – | – | – | – | – | 7.2 |
| 46 | 3-Amidino | NH ₂ | CH | CH | C | O | C | N | 2.5 |
| 47 | 3-Amidino | NH ₂ | CH | CH | C | N | C-NH ₂ | O | 0.4 |
| 48 | 3-Amidino | NH ₂ | CH | CH | C | N | NMe | CH | 0.2 |
| 49 | 3-Amidino | NH ₂ | CH | CH | C | N | CCH ₃ | NH | 6 |
| 50 | 3-Amidino | NH ₂ | CH | CH | C | NH | C=O | NH | 2.3 |
| 51 | 3-Amidino | NH ₂ | CH | CH | C | N | CCH ₃ | S | 0.06 |
| 52 | 3-Amidino | NH ₂ | CH | CH | C | N | CH | S | 0.8 |
| 53 | 3-Amidino | NH ₂ | CH | CH | C | N | CH | CH=CH | 1 |
| 54 | 3-Amidino | NH ₂ | CH | CH | C | CH | CH | CH=CH | 8.2 |
| 55 | 3-Amidino | NH ₂ | CH | CH | N | CH | CH | CH | 0.8 |
| 56 | 3-Amidino | NH ₂ | CH | CH | N | N | CH | CH | 0.16 |
| 57 | 3-Amidino | NH ₂ | CH | CH | N | N | N | CH | 0.023 |

(continued on next page)

Table 1
(continued)

| Compound | R ₁ | R ₂ | R ₃ | n | W | X | Y | Z | Activity K _i (nM) |
|----------|-----------------------------------|--------------------------|-----------------------------------|----|----|-----------------------|-------------------|-----------------------|------------------------------|
| 58 | 3-Amidino | NH ₂ | CH | CH | N | N | N | N | 0.037 |
| 59 | 3-Amidino | NH ₂ | CH | CH | N | CH | CH | N | 0.33 |
| 60 | 3-Amidino | NH ₂ | CH | CH | N | CH | N | CH | 1.1 |
| 61 | 3-Amidino | NH ₂ | CH | CH | N | N | C-CH ₃ | CH | 0.01 |
| 62 | 3-Amidino | NH ₂ | N | N | N | N | C-CH ₃ | CH | 0.041 |
| 63 | 3-Amidino | NH ₂ | N | CH | N | N | C-CH ₃ | CH | 0.007 |
| 64 | 3-Amidino | NH ₂ | N | CH | N | N | C-CF ₃ | CH | 0.009 |
| 65 | 3-Amidino | NH ₂ | CH | CH | N | N | C-CF ₃ | CH | 0.015 |
| 66 | 3-Amidino | NH ₂ | CF | CH | N | N | C-CF ₃ | CH | <0.005 |
| 67 | 3-Amidino | Me | CF | CH | N | N | C-CF ₃ | CH | <0.005 |
| 68 | 3-CH ₂ NH ₂ | NH ₂ | CH | CH | N | N | C-CH ₃ | CH | 2.7 |
| 69 | 3-CH ₂ NH ₂ | NH ₂ | CH | CH | N | N | C-CF ₃ | CH | 0.91 |
| 70 | 3-CH ₂ NH ₂ | NH ₂ | CF | CH | N | N | C-CF ₃ | CH | 0.36 |
| 71 | 3-CH ₂ NH ₂ | Me | CF | CH | N | N | C-CF ₃ | CH | 0.15 |
| 72 | 3-Amidino | NH ₂ | CH | CH | C | N | O | CH | 0.15 |
| 73 | – | – | – | – | – | – | – | – | 17,000 |
| 74 | – | – | – | – | – | – | – | – | 200 |
| 75 | 4-Amidino | 4-Amidino | – | 2 | – | CH ₂ | – | – | 0.66 |
| 76 | 4-Amidino | 4-Amidino | – | 1 | – | CH ₂ | – | – | 18 |
| 77 | 4-Amidino | 4-Amidino | – | 3 | – | CH ₂ | – | – | 4.6 |
| 78 | 4-Amidino | 4-Amidino | – | 4 | – | CH ₂ | – | – | 2.8 |
| 79 | 4-Amidino | 4-Amidino | – | 5 | – | CH ₂ | – | – | 110 |
| 80 | 4-Amidino | 4-Amidino | – | 1 | – | CH-CO ₂ H | – | – | 6.9 |
| 81 | 4-Amidino | 4-Amidino | – | 1 | – | CH-CO ₂ Me | – | – | 4.8 |
| 82 | 4-Amidino | 4-Amidino | – | 1 | – | CH-Ph | – | – | 5.2 |
| 83 | 4-Amidino | 4-Amidino | – | 1 | – | N-CO ₂ Et | – | – | 2.5 |
| 84 | 4-C(O)-NH ₂ | 4-C(O)-NH ₂ | – | 2 | – | CH ₂ | – | – | >5000 |
| 85 | 4-C(NH)-NMe ₂ | 4-C(NH)-NMe ₂ | – | 2 | – | CH ₂ | – | – | 1600 |
| 86 | 4-C(NH)-NH ₂ | 4-C(NH)-NHMe | – | 2 | – | CH ₂ | – | – | 12 |
| 87 | 4-C(NH)-NH ₂ | 4-C(NH)-NMe ₂ | – | 2 | – | CH ₂ | – | – | 18 |
| 88 | 4-C(NH)-NH ₂ | 4-C(O)-NH ₂ | – | 2 | – | CH ₂ | – | – | 50 |
| 89 | 4-Amidino | H | 4-Amidino | – | N | H | H | H | 400 |
| 90 | 4-Amidino | H | 4-Amidino | – | N | Cl | H | Cl | 620 |
| 91 | 4-Amidino | H | 4-Amidino | – | CH | Cl | H | Cl | 1100 |
| 92 | 3-Amidino | H | 3-Amidino | – | N | H | H | H | 53 |
| 93 | 3-Amidino | H | 3-Amidino | – | N | Cl | H | Cl | 130 |
| 94 | 3-Amidino | H | 3-Amidino | – | N | F | H | F | 24 |
| 95 | 3-Amidino | H | 3-Amidino | – | N | F | CH ₃ | F | 13 |
| 96 | 3-Amidino | H | 3-Amidino | – | N | F | OCH ₃ | F | 23 |
| 97 | 3-Amidino | H | 3-Amidino | – | N | H | H | CO ₂ H | 95 |
| 98 | 3-Amidino | H | 3-Amidino | – | N | H | H | CO ₂ Et | 190 |
| 99 | 3-Amidino | H | 3-Amidino | – | N | H | H | C(O)-NH ₂ | 600 |
| 100 | 3-Amidino | H | 3-Amidino | – | N | H | H | C(O)-NHMe | 230 |
| 101 | 3-Amidino | H | 3-Amidino | – | N | H | H | C(O)-NMe ₂ | 32 |
| 102 | 3-Amidino | H | 3-Amidino | – | N | H | H | NH ₂ | 53 |
| 103 | 3-Amidino | H | H | – | N | F | CH ₃ | F | >5000 |
| 104 | 3-Amidino | H | 3-CO ₂ H | – | N | F | CH ₃ | F | 8000 |
| 105 | 3-Amidino | H | 3-NMe ₂ | – | N | F | CH ₃ | F | 160 |
| 106 | 3-Amidino | H | 3-CH ₂ NH ₂ | – | N | F | CH ₃ | F | 1400 |
| 107 | 3-Amidino | H | 3-NEt ₂ | – | N | F | CH ₃ | F | 400 |
| 108 | 3-Amidino | H | 3-C(O)NMe ₂ | – | N | F | CH ₃ | F | 80 |

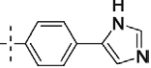
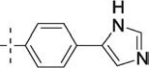
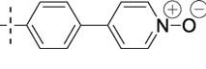
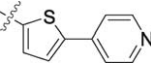
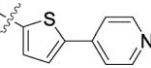
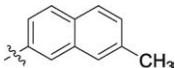
(suite page suivante)

Table 1
(continued)

| Compound | R ₁ | R ₂ | R ₃ | n | W | X | Y | Z | Activity K _i (nM) |
|----------|------------------------------------|-----------------|----------------|----|---|---------------------------------|-------------------|-------|------------------------------|
| 1 | Amidino | – | – | – | – | – | – | – | 34 |
| 2 | Guanidino | – | – | – | – | – | – | – | 9 |
| 3 | – | – | – | – | – | – | – | – | 800 |
| 4 | – | – | – | – | – | – | – | – | 61 |
| 5 | – | – | – | – | – | – | – | – | 41 |
| 6 | – | – | – | – | – | – | – | – | 140 |
| 7 | – | – | – | – | – | – | – | – | 38,500 |
| 8 | 4-Amidino | – | – | – | – | – | – | – | 1700 |
| 9 | 3-Amidino | – | – | – | – | – | – | – | 1400 |
| 10 | H | – | – | – | – | – | – | – | 270 |
| 11 | CH ₂ CO ₂ Me | – | – | – | – | – | – | – | 94 |
| 12 | – | – | – | – | – | – | – | – | 18 |
| 13 | H | – | – | – | – | – | – | – | 200 |
| 14 | SO ₂ NH ₂ | – | – | – | – | – | – | – | 6.3 |
| 15 | SO ₂ Me | – | – | – | – | – | – | – | 4.2 |
| 16 | – | – | – | – | – | CH ₂ | – | – | 10 |
| 17 | – | – | – | – | – | CH ₂ CH ₂ | – | – | 1.3 |
| 18 | – | – | – | – | – | CH ₂ O | – | – | 0.53 |
| 19 | Me | – | – | – | – | CH ₂ | – | – | 3.4 |
| 20 | H | – | – | – | – | CH ₂ | – | – | 600 |
| 21 | Me | – | – | – | – | O | – | – | 3.1 |
| 22 | <i>i</i> -Pr | – | – | – | – | O | – | – | 1.8 |
| 23 | Benzyl | – | – | – | – | – | – | – | 102 |
| 24 | Benzylsulfonyl | – | – | – | – | – | – | – | 10 |
| 25 | Thiophen-2-sulfonyl | – | – | – | – | – | – | – | 12 |
| 26 | Benzyl | – | – | – | – | – | – | – | 0.8 |
| 27 | 2-Thiophene | – | – | – | – | – | – | – | 3.4 |
| 28 | 5-Amidino | – | – | – | – | – | – | – | 400 |
| 29 | 6-Amidino | – | – | – | – | – | – | – | 2.4 |
| 30 | – | – | – | – | – | CH | – | – | 2.1 |
| 31 | – | – | – | – | – | CBr | – | – | 0.32 |
| 32 | – | – | – | – | – | N | – | – | 0.44 |
| 33 | – | – | – | – | – | – | – | – | 1.2 |
| 34 | CO ₂ Me | – | – | – | – | CH | N | – | 0.83 |
| 35 | CO ₂ Me | – | – | – | – | N | N | – | 2.3 |
| 36 | OEt | – | – | – | – | CH | N | – | 0.8 |
| 37 | <i>N</i> -tetrazolyl | – | – | – | – | CH | CH | – | 0.52 |
| 38 | <i>N</i> -tetrazolyl | – | – | – | – | CH | N | – | 0.11 |
| 39 | <i>N</i> -tetrazolyl | – | – | – | – | – | – | – | 1.6 |
| 40 | Me | – | – | – | – | – | – | – | 11 |
| 41 | OMe | – | – | – | – | – | – | – | 3.4 |
| 42 | <i>N</i> -tetrazolyl | – | – | – | – | – | – | – | 2.3 |
| 43 | Me | – | – | – | – | – | – | – | 1.8 |
| 44 | OMe | – | – | – | – | – | – | – | 0.55 |
| 45 | – | – | – | – | – | – | – | – | 7.2 |
| 46 | 3-Amidino | NH ₂ | CH | CH | C | O | C | N | 2.5 |
| 47 | 3-Amidino | NH ₂ | CH | CH | C | N | C-NH ₂ | O | 0.4 |
| 48 | 3-Amidino | NH ₂ | CH | CH | C | N | NMe | CH | 0.2 |
| 49 | 3-Amidino | NH ₂ | CH | CH | C | N | CCH ₃ | NH | 6 |
| 50 | 3-Amidino | NH ₂ | CH | CH | C | NH | C=O | NH | 2.3 |
| 51 | 3-Amidino | NH ₂ | CH | CH | C | N | CCH ₃ | S | 0.06 |
| 52 | 3-Amidino | NH ₂ | CH | CH | C | N | CH | S | 0.8 |
| 53 | 3-Amidino | NH ₂ | CH | CH | C | N | CH | CH=CH | 1 |
| 54 | 3-Amidino | NH ₂ | CH | CH | C | CH | CH | CH=CH | 8.2 |
| 55 | 3-Amidino | NH ₂ | CH | CH | N | CH | CH | CH | 0.8 |
| 56 | 3-Amidino | NH ₂ | CH | CH | N | N | CH | CH | 0.16 |
| 57 | 3-Amidino | NH ₂ | CH | CH | N | N | N | CH | 0.023 |

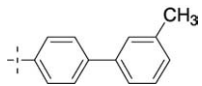
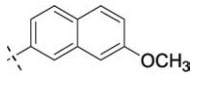
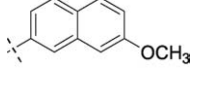
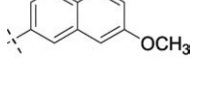
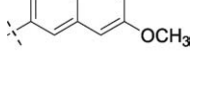
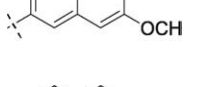
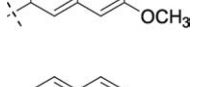
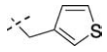
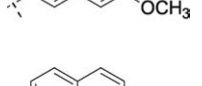
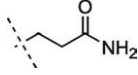
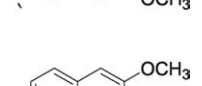
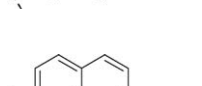
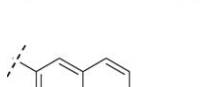
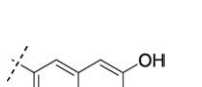
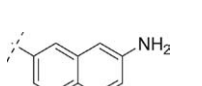
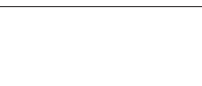
(continued on next page)

Table 1
(continued)

| Compound | R ₁ | R ₂ | R ₃ | n | W | X | Y | Z | Activity K _i (nM) |
|----------|---|--------------------|---|---|---|----|-------------------|----|------------------------------|
| 109 | 3-Amidino | H | 3-C(O)NEt ₂ | – | N | F | CH ₃ | F | 400 |
| 110 | 3-Amidino | H | 3-C(O)NH ₂ | – | N | F | CH ₃ | F | 280 |
| 111 | 3-Amidino | 6-OMe | 3-C(O)NMe ₂ | – | N | F | CH ₃ | F | 740 |
| 112 | 3-Amidino | 6-NH ₂ | 3-C(O)NMe ₂ | – | N | F | CH ₃ | F | 14 |
| 113 | 3-Amidino | 6-OH | 3-C(O)NMe ₂ | – | N | F | CH ₃ | F | 1.2 |
| 114 | 3-Amidino | 2-OH | 3-C(O)NMe ₂ | – | N | F | CH ₃ | F | >5000 |
| 115 | 3-Amidino | 6-OH | 3-C(O)NMe ₂ | – | N | H | CO ₂ H | H | 3.9 |
| 116 | 3-Amidino | 2-OH | 2-OMe, 5-COO ₂ H | – | – | – | – | – | 0.18 |
| 117 | 3-Amidino | 2-OH | 2-OCH ₃ | – | – | – | – | – | 0.56 |
| 118 | 3-Amidino | 2-OH | 2-OMe, 4-CO ₂ H | – | – | – | – | – | 0.33 |
| 119 | 3-Amidino | 2-OH | 2-OH, 4-CO ₂ H | – | – | – | – | – | 0.12 |
| 120 | 3-Amidino | 2-OH | 3-CONH ₂ , 5-CO ₂ H | – | – | – | – | – | 0.22 |
| 121 | 3-Amidino | 2-OH | 2,6-Me, 4-CO ₂ H | – | – | – | – | – | 0.25 |
| 122 | 3-Amidino | 2-OH | 2-Cl, 4-CO ₂ H | – | – | – | – | – | 0.3 |
| 123 | – | – | – | – | – | – | – | – | 0.11 |
| 124 | COOMe | CH=CHPh | H | – | – | H | – | – | 9.4 |
| 125 | COOMe | CH ₂ Ph | H | – | – | H | – | – | 11 |
| 126 | COOMe | CH ₃ | H | – | – | H | – | – | 5.3 |
| 127 | COOMe | CH ₃ | 2-NO ₂ | – | – | H | – | – | 1.1 |
| 128 | COOMe | CH ₃ | 3-NO ₂ | – | – | H | – | – | 1.1 |
| 129 | COOMe | CH ₃ | 3, 4-OMe | – | – | H | – | – | 1.3 |
| 130 | COOMe | CH ₃ | 3-CH ₂ NH ₂ | – | – | H | – | – | 0.9 |
| 131 | COOMe | CH ₃ | 3-CH ₂ NMe ₂ | – | – | H | – | – | 1 |
| 132 | COOMe | CH ₃ | 3-CH ₂ NMe ₃ ⁺ | – | – | H | – | – | 1 |
| 133 | COOMe | CH ₃ | 3-CH ₂ NHCO ₂ Et | – | – | H | – | – | >1200 |
| 134 | COOMe | CH ₃ | 3-CONH ₂ | – | – | H | – | – | 0.5 |
| 135 | COOMe | CH ₃ | 4-CONH ₂ | – | – | H | – | – | 2 |
| 136 | COOMe | CH ₃ | 4-CO ₂ H | – | – | H | – | – | 69 |
| 137 | H | H | H | – | – | H | – | – | 1000 |
| 138 | H | H | H | – | – | OH | – | – | 88 |
| 139 | H | H | 3-CONH ₂ | – | – | OH | – | – | 7 |
| 140 | – | – | H | – | – | – | – | CH | 51 |
| 141 | – | – | 3-CONH ₂ | – | – | – | – | CH | 5 |
| 142 | – | – | 3-CONH ₂ | – | – | – | – | N | 0.75 |
| 143 | – | – |  | – | – | – | – | – | 3 |
| 144 | – | – |  | – | – | – | – | – | 46 |
| 145 | – | – |  | – | – | – | – | – | 0.4 |
| 146 | – | – |  | – | – | – | – | – | 12 |
| 147 | – | – |  | – | – | – | – | – | 42 |
| 148 |  | H | H | – | – | – | – | – | 100 |

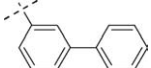
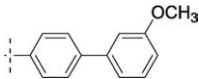
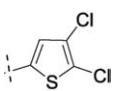
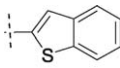
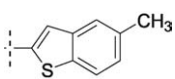
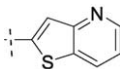
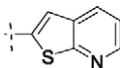
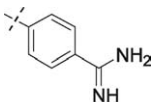
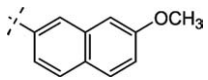
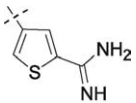
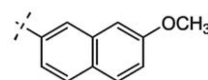
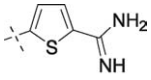
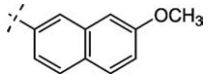
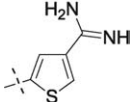
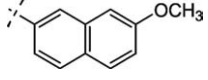
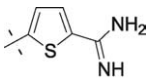
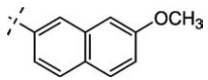
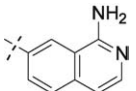
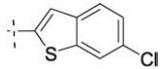
(continued on next page)

Table 1
(continued)

| Compound | R ₁ | R ₂ | R ₃ | n | W | X | Y | Z | Activity K _i (nM) |
|----------|---|---|----------------|---|---|---|---|---|------------------------------|
| 149 |  | H | H | – | – | – | – | – | 120 |
| 150 |  | H | H | – | – | – | – | – | 47 |
| 151 |  | CH ₃ | H | – | – | – | – | – | 22 |
| 152 |  | CH ₃ | H | – | – | – | – | – | 440 |
| 153 |  | CH ₂ CH ₃ | H | – | – | – | – | – | 43 |
| 154 |  | CH ₂ Ph | H | – | – | – | – | – | 41 |
| 155 |  |  | H | – | – | – | – | – | 19 |
| 156 |  |  | H | – | – | – | – | – | 56 |
| 157 |  | H | H | – | – | – | – | – | 3 |
| 158 |  | H | H | – | – | – | – | – | 1000 |
| 159 |  | H | H | – | – | – | – | – | 280 |
| 160 |  | H | H | – | – | – | – | – | 70 |
| 161 |  | H | H | – | – | – | – | – | 500 |
| 162 |  | H | H | – | – | – | – | – | 120 |

(continued on next page)

Table 1
(continued)

| Compound | R ₁ | R ₂ | R ₃ | n | W | X | Y | Z | Activity K _i (nM) |
|----------|---|---|-----------------|---|---|---|---|---|------------------------------|
| 163 |  | H | 4-OH | – | – | – | – | – | 440 |
| 164 |  | H | H | – | – | – | – | – | 100 |
| 165 |  | H | H | – | – | – | – | – | 47 |
| 166 |  | H | H | – | – | – | – | – | 19 |
| 167 |  | H | H | – | – | – | – | – | 78 |
| 168 |  | H | 4-OH | – | – | – | – | – | 0.7 |
| 169 |  | H | H | – | – | – | – | – | 88 |
| 170 |  |  | H | – | – | – | – | – | 970 |
| 171 |  |  | H | – | – | – | – | – | 11 |
| 172 |  |  | H | – | – | – | – | – | 22 |
| 173 |  |  | H | – | – | – | – | – | 83 |
| 174 |  |  | CH ₃ | – | – | – | – | – | 7 |
| 175 |  |  | H | – | – | – | – | – | 170 |

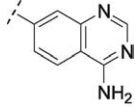
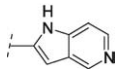
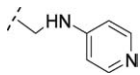
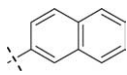
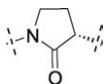
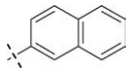

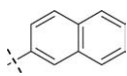

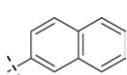
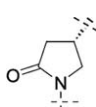
(continued on next page)

Table 1
(continued)

| Compound | R ₁ | R ₂ | R ₃ | n | W | X | Y | Z | Activity K _i (nM) |
|----------|----------------|----------------|----------------|---|---|---|---|---|------------------------------|
| 176 | | | H | – | – | – | – | – | 22 |
| 177 | | | H | – | – | – | – | – | 210 |
| 178 | | | H | – | – | – | – | – | 41 |
| 179 | | | – | – | – | – | – | – | 4 |
| 180 | | | – | – | – | – | – | – | 43 |
| 181 | | | – | – | – | – | – | – | 17 |
| 182 | | | – | – | – | – | – | – | 340 |
| 183 | | | – | – | – | – | – | – | 42 |
| 184 | | | – | – | – | – | – | – | 18 |
| 185 | | – | – | – | – | – | – | – | 18 |
| 186 | | – | – | – | – | – | – | – | 1.3 |
| 187 | | – | – | – | – | – | – | – | 58 |
| 188 | | – | – | – | – | – | – | – | 200 |

(continued on next page)

Table 1
(continued)

| Compound | R ₁ | R ₂ | R ₃ | n | W | X | Y | Z | Activity K _i (nM) |
|----------|---|------------------|----------------|---|---|-----|-----|----------------------|------------------------------|
| 189 |  | – | – | – | – | – | – | – | 0.8 |
| 190 |  | – | – | – | – | – | – | – | 0.004 |
| 191 |  | – | – | – | – | – | – | – | 0.022 |
| 192 | – | – | – | – | – | – | – | – | 900 |
| 193 |  | – | – | – |  | – | – | – | 230 |
| 194 |  | – | – | – |  | – | – | – | 7000 |
| 195 |  | – | – | – |  | – | – | – | 1300 |
| 196 |  | – | – | – |  | – | – | – | 9800 |
| 197 | OCH ₃ | OCH ₃ | – | – | – | NH | C=O | H | 770 |
| 198 | CH ₃ | t-Bu | – | – | – | C=O | NH | H | 294.1 |
| 199 | OCH ₃ | t-Bu | – | – | – | C=O | NH | NHSO ₂ Me | 74.1 |

- Spacing: This parameter controls the minimum allowed inter-feature distance in the resulting hypotheses. The default value is 297 picometers. However, in the current study, the spacing was set to 10 picometers to generate more meaningful hypotheses that reflect the feature-rich nature of fXa inhibitors [66].
- Weight variation: In a CATALYST[®] generated pharmacophore model each feature represents certain orders of magnitude of activity of the compound. The extent to which this magnitude is explored by the hypothesis generator is controlled by the parameter weight variation [75]. By default, CATALYST[®] keeps the value of function weights close to 2. However, in the current study, CATALYST[®] was allowed to explore variable feature weights in runs 7, 8 and 9 (see Section 3.2 and Table 3).

2.3.3. Assessment of the generated hypotheses

2.3.3.1. Automatic assessment. When generating hypotheses, CATALYST[®] tries to minimize a cost function consisting of three terms: weight cost, error cost and configuration

cost [66,67,74–76]. Weight cost is a value that increases as the feature weight in a model deviates from an ideal value of 2. The deviation between the estimated activities of the training set and their experimentally determined values adds to the error cost. The third term penalizes the complexity of the hypothesis, i.e. the configuration cost. This is a fixed cost, which is equal to the entropy of the hypothesis space. The more the numbers of features (a maximum of 5) in a generated hypothesis, the higher is 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 normally distributed 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 [66]. Table 4 shows the total costs and null costs for some selected superior fXa inhibitor pharmacophore hypotheses.

Table 4

The performance of representative fXa pharmacophore hypotheses as predictive 3D-QSARs and database search queries (see Section 2.3.3)

| Run | Hypotheses ^a | Pharmacophoric features in generated hypotheses | Total cost | Cost of null hypothesis | Residual cost | R ^b | SCC ^c | Coverage |
|-----|-------------------------|---|------------|-------------------------|---------------|----------------|------------------|----------|
| 1 | 2 ^d | HBA, HBD, AromRing, PosIoniz | 119.1 | 154.2 | 35.1 | 0.83 | 0.41 | 52 |
| | 3 | HBA, HBD, HbicArom, PosIoniz | 122.7 | 154.2 | 31.5 | 0.83 | 0.53 | 69 |
| | 6 | HBA, HBD, AromRing, PosIoniz | 125.3 | 154.2 | 28.9 | 0.78 | 0.39 | 74 |
| 2 | 1 | HBA, HBD, 2 × HbicArom, PosIoniz | 109.7 | 154.2 | 44.5 | 0.91 | 0.45 | 89 |
| | 2 | 2 × HBA, 2 × HbicArom, PosIoniz | 114.9 | 154.2 | 39.3 | 0.83 | 0.52 | 121 |
| | 4 | HBA, HBD, 2 × HbicArom, PosIoniz | 119.3 | 154.2 | 34.9 | 0.81 | 0.32 | 117 |
| | 8 | HBA, HBD, 2 × HbicArom, PosIoniz | 123.1 | 154.2 | 31.1 | 0.78 | 0.40 | 130 |
| | 9 | HBA, HBD, 2 × HbicArom, PosIoniz | 124.2 | 154.2 | 30.0 | 0.76 | 0.39 | 52 |
| 3 | 1 | 2 × HBA, AromRing, PosIoniz | 212.0 | 281.6 | 69.6 | 0.81 | 0.66 | 74 |
| | 3 | 2 × HBA, AromRing, PosIoniz | 241.2 | 281.6 | 40.4 | 0.67 | 0.64 | 126 |
| | 4 | HBA, 2 × AromRing, PosIoniz | 243.5 | 281.6 | 38.1 | 0.64 | 0.59 | 87 |
| | 6 | HBA, 2 × AromRing, PosIoniz | 245.0 | 281.6 | 36.6 | 0.66 | 0.60 | 66 |
| | 9 | 2 × HBA, AromRing, PosIoniz | 248.7 | 281.6 | 32.9 | 0.62 | 0.59 | 112 |
| 4 | 1 | 2 × HBA, 2 × HbicArom, PosIoniz | 213.4 | 281.6 | 68.2 | 0.79 | 0.63 | 93 |
| | 2 | 2 × HBA, 2 × HbicArom, PosIoniz | 238.7 | 281.6 | 42.9 | 0.64 | 0.65 | 78 |
| | 7 | HBA, 2 × HbicArom, AromRing, PosIoniz | 250.4 | 281.6 | 31.2 | 0.57 | 0.63 | 132 |
| | 10 | HBA, 3 × HbicArom, PosIoniz | 254.0 | 281.6 | 27.6 | 0.55 | 0.60 | 117 |
| 5 | 6 | 2 × HBA, AromRing, PosIoniz | 253.3 | 299.3 | 46.0 | 0.65 | 0.54 | 100 |
| | 8 | HBA, HBD, AromRing, PosIoniz | 255.3 | 299.3 | 44.0 | 0.66 | 0.47 | 49 |
| | 9 | HBA, HBD, AromRing, PosIoniz | 257.0 | 299.3 | 42.3 | 0.62 | 0.49 | 69 |
| 6 | 1 | HBA, HBD, 2 × HbicArom, PosIoniz | 213.1 | 299.3 | 86.2 | 0.85 | 0.43 | 71 |
| | 5 | HBA, HBD, 2 × HbicArom, PosIoniz | 244.2 | 299.3 | 55.1 | 0.68 | 0.39 | 32 |
| | 8 | 2 × HBA, 2 × HbicArom, PosIoniz | 246.6 | 299.3 | 52.7 | 0.66 | 0.42 | 126 |
| 7 | 1 | HBA, 2 × HbicArom, AromRing, PosIoniz | 275.2 | 324.4 | 49.2 | 0.71 | 0.59 | 103 |
| | 4 | HBA, 2 × HbicArom, AromRing, PosIoniz | 289.5 | 324.4 | 34.9 | 0.64 | 0.66 | 102 |
| | 8 | HBA, 2 × HbicArom, AromRing, PosIoniz | 301.6 | 324.4 | 22.8 | 0.58 | 0.44 | 140 |
| | 9 | HBA, 2 × HbicArom, AromRing, PosIoniz | 302.4 | 324.4 | 22.0 | 0.57 | 0.44 | 144 |
| | 10 | HBA, 2 × HbicArom, AromRing, PosIoniz | 304.4 | 324.4 | 20.0 | 0.56 | 0.52 | 56 |
| 8 | 1 | 2 × HBA, 2 × HbicArom, PosIoniz | 269.8 | 324.4 | 54.6 | 0.74 | 0.65 | 84 |
| | 2 | 2 × HBA, 2 × HbicArom, PosIoniz | 294.2 | 324.4 | 30.2 | 0.61 | 0.66 | 90 |
| | 3 | 2 × HBA, 2 × HbicArom, PosIoniz | 300.5 | 324.4 | 23.9 | 0.57 | 0.54 | 93 |
| | 4 | 2 × HBA, 2 × HbicArom, PosIoniz | 302.8 | 324.4 | 21.6 | 0.56 | 0.65 | 76 |
| | 9 | 2 × HBA, 2 × HbicArom, PosIoniz | 310.4 | 324.4 | 14.0 | 0.53 | 0.67 | 80 |
| 9 | 1 | 2 × HBA, AromRing, PosIoniz | 270.1 | 324.4 | 54.3 | 0.76 | 0.60 | 43 |
| | 2 | 2 × HBA, AromRing, PosIoniz | 279.0 | 324.4 | 45.4 | 0.72 | 0.55 | 14 |
| | 3 | 2 × HBA, AromRing, PosIoniz | 281.6 | 324.4 | 42.8 | 0.71 | 0.54 | 99 |
| | 6 | 2 × HBA, AromRing, PosIoniz | 289.9 | 324.4 | 34.5 | 0.65 | 0.61 | 90 |
| | 8 | 2 × HBA, AromRing, PosIoniz | 291.0 | 324.4 | 33.4 | 0.64 | 0.61 | 72 |

^a High ranking representative fXa hypotheses (represent their corresponding clusters). Representative models of very low predictive (SCC) values were omitted from the table.

^b The correlation coefficients between bioactivity estimates and bioactivities of corresponding training compounds.

^c The SCC determined for each hypothesis against the activities of 199 inhibitors.

^d Serial numbers given by CATALYST[®] for individual hypotheses and reported in the log book of the corresponding automatic run.

2.3.3.2. *Evaluation of hypotheses as database search queries and predictive 3D-QSAR models.* The cost-based hypotheses ranking concept implemented in CATALYST[®] depends heavily on the training compounds employed in the respective modeling run. Accordingly, it is fairly inaccurate to compare two or more CATALYST[®] pharmacophores derived from different training sets according to their total costs. The current project involves exploring and evaluating various pharmacophores originating from different training sets. Therefore, a standard validation approach was developed and employed to allow effective ranking of pharmacophoric models regardless of the training sets that yielded them. The proposed validation proceeded as follows. Hypotheses

created by each automatic CATALYST[®] run were clustered into five groups utilizing the hierarchical average linkage method available in CATALYST[®]. Thereafter, the highest-ranking hypothesis within each cluster (i.e. of lowest cost) was selected to represent the corresponding cluster in subsequent assessment. Representative models were then evaluated as database search queries, and as predictive 3D-QSARs models against the complete list of fXa inhibitors (1–199, Table 1 and Fig. 1). Accordingly, these compounds (and associated conformational ensembles) were enlisted in a database that was subjected to “Best Flexible Database Search” against the particular pharmacophore undergoing validation. Pharmacophoric coverage is defined as the total number of

Table 5
The highest-ranking pharmacophore models according to their SCC

| Pharmacophore model ^a | Coverage | Features of resulting pharmacophores | SCC |
|----------------------------------|----------|---|-------|
| Hypo8/9 | 80 | HBA (2); HbicArom (2); 1 PosIoniz (1) | 0.674 |
| Hypo8/2 | 90 | HBA (2); HbicArom (2); 1 PosIoniz (1) | 0.663 |
| Hypo7/4 | 102 | HBA (1); HbicArom (2); AromRing (1); 1 PosIoniz (1) | 0.659 |
| Hypo3/1 | 74 | HBA (2); AromRing (1); 1 PosIoniz (1) | 0.655 |
| Hypo8/4 | 76 | HBA (2); HbicArom (2); 1 PosIoniz (1) | 0.652 |
| Hypo4/2 | 78 | HBA (2); HbicArom (2); 1 PosIoniz (1) | 0.648 |
| Hypo8/1 | 84 | HBA (2); HbicArom (2); 1 PosIoniz (1) | 0.647 |
| Hypo3/3 | 126 | HBA (2); AromRing (1); 1 PosIoniz (1) | 0.644 |
| Hypo4/1 | 93 | HBA (2); HbicArom (2); 1 PosIoniz (1) | 0.633 |
| Hypo4/7 | 132 | HBA (1); HbicArom (2); AromRing (1); 1 PosIoniz (1) | 0.628 |
| Hypo9/6 | 90 | HBA (2); AromRing (1); 1 PosIoniz (1) | 0.614 |
| Hypo9/8 | 72 | HBA (2); AromRing (1); 1 PosIoniz (1) | 0.612 |
| Hypo3/6 | 66 | HBA (1); AromRing (2); 1 PosIoniz (1) | 0.603 |
| Hypo9/1 | 43 | HBA (2); AromRing (1); 1 PosIoniz (1) | 0.595 |
| Hypo7/1 | 103 | HBA (1); HbicArom (2); AromRing (1); 1 PosIoniz (1) | 0.591 |
| Hypo3/9 | 112 | HBA (2); AromRing (1); 1 PosIoniz (1) | 0.590 |

^a Hypo8/9 means hypothesis number 9 generated in run number 8 (Table 4).

inhibitors captured by the respective model. Table 4 shows the pharmacophoric coverage values observed for the representative pharmacophores. On the other hand, the predictive potential of any representative model was assessed by statistically correlating bioactivity estimates, produced by fitting the 199 fXa inhibitors (Table 1) against the pharmacophore under validation, and the corresponding experimental bioactivities. The activity of any compound can be estimated from a particular hypothesis through Eq. (3) [65,66].

$$\log (\text{Estimated activity}) = 10 (I - \text{Fit}) \quad (3)$$

where, I = 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) [66].

$$\text{Fit} = \sum \text{mapped hypothesis features} \times W \quad (4)$$

$$\left[1 - \sum \text{hypothesis feature} (\text{disp/tol})^2 \right]$$

where, W weight of the corresponding hypothesis feature spheres, adaptively determined; disp = distance of the mapping function on the molecule to the hypothesis feature sphere center; tol = radius of the hypothesis feature sphere (known as Tolerance and it is related to the uncertainty values of reported bioactivities). Practically, the predictive capacity of any hypothesis was determined by using the SCORE command within “Generate Hypothesis” workbench of CATALYST[®] against the 199 inhibitors. Hence we termed the resulting correlation coefficient as the scoring correlation coefficient (SCC). Table 4 illustrates the SCC values determined for representative fXa pharmacophore models. Interestingly, although some hypotheses were assigned high correlation coefficients against their training sets (i.e. low total costs) by CATALYST[®], they exhibited poor coverage and SCC values, e.g. Hypo1/2, Hypo2/1, Hypo2/4 and Hypo6/1.

2.3.4. QSAR modeling

2.3.4.1. Molecular descriptors. QSAR modeling was performed employing QSARIS[®] [78]. The 3D energy-minimized chemical structures of fXa inhibitors (1–199, Table 1) were imported into QSARIS[®] from CATALYST[®] in MDL-molfile format. Thereafter, QSARIS[®] was utilized to compute a variety of molecular descriptors for the inhibitors. A total of 228 descriptors in five categories were calculated: (i) simple and valence connectivity indices (52 indices), (ii) electrotopological indices (126 E-state indices), (iii) Kappa Shape Indices (eight parameters), (iv) general molecular properties (29 descriptors, e.g. $\log P$, molecular weight, number of different elements, HBDs/acceptors, volume, surface, ovality, etc...) and (v) molecular moment descriptors (13 parameters) [77]. Thereafter, the 3D structures (1–199, Table 1) were imported into Alchemy2000[®] to calculate a variety of electronic descriptors employing semi-empirical quantum mechanical methods (MOPAC, PM3 model) without further energy minimization. The following descriptors were calculated employing Alchemy2000[®]: energies of highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO), heats of formation, dipole moments and ionization potentials. Finally, the logarithmic transformations of bioactivity estimates produced by fitting the 199 inhibitors against best pharmacophores (17 models of $\text{SCC} \geq 0.59$, Table 5) were imported into QSARIS[®] and combined with other molecular descriptors. Then, the GA module implemented within QSARIS[®] was utilized to select an optimal combination of descriptors for MLR. Statistical outliers were identified from experimental-versus-predicted plots and were subsequently removed (i.e. compounds 20, 73, 49, 75, 114, 133 and 171, Table 1).

2.3.4.2. Genetic algorithm selection of molecular descriptors. GA techniques rely on the evolutionary operations of “crossover and mutation” to select optimal combination of

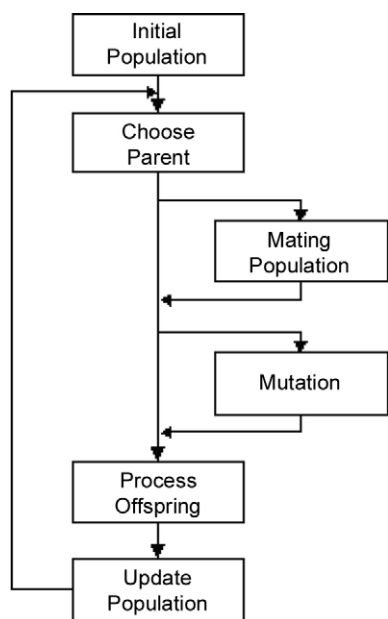


Fig. 2. A flowchart of the GA processing for QSAR analysis.

descriptors capable of explaining bioactivity variation among training compounds. GA operates through a cycle of four stages: (i) encoding mechanism; (ii) definition of a fitness function; (iii) creating a population of chromosomes; (iv) genetic manipulation of chromosomes [77,78]. The coding scheme used in QSARIS[®] is gene-based. In this scheme, the possible regression models (chromosomes) differ from one another by the set of independent variables (descriptors) that comprise each model. If the general number of independent variables (descriptors) is equal to P (in this particular case, $P = 250$ variable), the chromosome corresponding to any model consists of a string of P binary digits (bits) called “genes”. Each value in the string represents an independent variable (0 = absent, 1 = present). 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); Eqs. (5) and (6):

$$\text{Fit} = \frac{1}{\text{LOF}} \quad (5)$$

$$\text{LOF} = \frac{\text{RSS}_p/N}{\left(1 - \frac{d \cdot (p+1)}{N}\right)^2} \quad (6)$$

where, p is the number of independent variables in the model; N is the number of samples in the data set; RSS_p is a residual sum of squares based on the regression model using p independent variables; d is a smoothing factor to be set by the user according to practical considerations (from 1 up to 4). The other operations of the genetic methods are depicted in the flowchart in Fig. 2. The followings are important QSA-RIS[®] control parameters used in the GA-based selection of optimal descriptors [77,78]:

- Creating an initial population: The user must specify a number of initial random chromosomes. In the current research, we decided to start with 100 initial random chromosomes, which correspond to the maximal limit of initial chromosomes in QSARIS[®].
 - Choosing a parent: Parent selection in GA aims at providing more reproductive chances (mating) for the fittest chromosomes. Our diagnostic trials indicated that the Linear Ranking Parent Selection Scheme yielded optimal QSAR models for the current fXa inhibitors. In this scheme the probability of selecting each parent is directly proportional to its ordinal fitness number in the population [78,79].
 - Mating population: Mating is an operation during which two parents’ chromosomes are combined to generate new solutions (offspring). For a couple of parents two parameters are to be configured: (i) the probability of mating, which can take values between 0.0 and 1.0 (set to 0.90 in the current project). (ii) The number of offspring chromosomes from the same parents (set to 2 in the present work). QSARIS[®] offers three possible crossover operators for mating: (a) one-point crossover, (b) two-point crossover, and (c) uniform crossover. Diagnostic trials performed on the current fXa data indicated that uniform crossover yielded superior QSAR models. In uniform crossover each gene for a given offspring can be independently chosen from one parent or the other. The other offspring receives the complementary value [78,79].
 - Mutation operator: This operator modifies any single chromosome with a given probability, which can take values between 0.0 and 1.0. A mutation operator changes one or more bits in the chromosome to its complement. It is possible to define single or two-point randomly chosen mutations. In addition, uniform mutation is possible, where at least one gene is changed. Diagnostic trials employing the current fXa data have shown that uniform mutation with a 0.7 probability yielded the best results [78,79].
 - The offspring process: Displacing an existing member with better offspring. The following variants are possible in QSARIS[®]: (i) replace weakest members of the population with the best offspring, and (ii) derive the next population from the best solutions only. In the present case we achieved optimal QSAR models through replacing the weakest parents with best offspring.
 - Maximum number of generations: This is needed for an exit from a basic cycle and completion of the algorithm.
- Table 6 summarizes the optimal GA parameters employed in the development of fXa inhibitors’ QSAR model.

2.3.4.3. *Cross-validation of QSAR model.* The final statistical model was automatically cross-validated utilizing the leave-one-out technique. In this technique, each observation is systematically deleted from the dataset and predicted by re-computing the regression model from the remaining observations. The resulting correlation coefficient (R^2) is called

multiple- Q^2 (or R^2_{press}). This cross-validation approach often gives an indication of potential for prediction [78].

3. Results and discussion

CATALYST[®] enables automatic pharmacophore construction by using a collection of molecules with activities ranging over a number of orders of magnitude. The pharmacophore (hypothesis) produced by CATALYST[®] explains 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 (HBA) and donors, hydrophobic regions, etc...). Different hypotheses were generated for a series of fXa inhibitors. A total of 199 compounds were used in this study (Table 1). Four training subsets were selected from the collection. Each subset consisted of inhibitors of wide structural diversity. The biological activity in each group spanned nearly four orders of magnitude. GA and MLR statistical analysis were subsequently employed to select an optimal combination of complementary pharmacophores capable of explaining bioactivity variations among all 199 fXa inhibitors.

3.1. Data mining and conformational coverage

An extensive literature search was conducted to collect structurally diverse set of fXa inhibitors (1–199, Table 1). Evidently, effective pharmacophore modeling requires training sets of adequate structural diversity to elucidate the common functional features responsible for ligand-receptor affinity across wide ligand diversity.

The conformational space of each fXa inhibitor (compounds 1–199, Table 1) was extensively sampled utilizing the poling algorithm of CATALYST[®]. Poling promotes confor-

mational variation via employing molecular mechanical force field algorithm that penalizes similar conformers, and thus improves the coverage of the conformational space [80,81]. For each of the training set compounds, a conformational database was generated using the ‘best conformer generation’ option. This option considers the arrangement in space of chemical features rather than simply the arrangement of atoms and uses the poling function. The algorithm intends to optimize the conformational coverage versus the size of the assembly [80,81]. Effective conformation coverage guarantees minimal conformation-related noise during pharmacophore generation and validation. Pharmacophore modeling and pharmacophore-based database search procedures are known to be sensitive to inadequate conformational sampling of training compounds [74,75,82].

3.2. Ligand-based exploration of fXa pharmacophoric space

In order to effectively explore the flexibility of fXa binding site, it is necessary to identify all possible pharmacophoric modes assumed by various ligands within the binding pocket of fXa. However, CATALYST[®] implements an optimization algorithm that evaluates large number of potential models within the pharmacophoric “space” of a particular target through fine perturbations to hypotheses that survived the subtractive and constructive phases (see Section 2.3.2) [74]. 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 CATALYST[®] run not to exceed 17 (corresponding to 2^{17} hypotheses to be assessed by CATALYST[®]), otherwise a thorough analysis of all models cannot be guaranteed [75]. Undoubtedly, the size of the investigated pharmacophoric space is a function of training compounds, selected input chemical features and other CATALYST[®] control parameters such as feature tolerances and weights [75]. Conceivably, restricting the extent of explored pharmacophoric space might improve the efficiency of optimization via allowing effective evaluation of limited number of pharmacophoric models. On the other hand, severe restrictions imposed on the pharmacophoric space might reduce the possibility of discovering optimal pharmacophoric hypotheses, particularly if they occur outside the “boundaries” of the constrained space. Consequently, it was reasoned that in order to discover the pharmacophores that represent fXa binding site flexibility, it is essential to initially identify optimal “area(s)” within the pharmacophoric space of fXa inhibitors. Subsequent thorough evaluation of the identified optimal pharmacophoric area(s) should uncover significant models that collectively represent ligand flexibility within the binding site. In order to achieve that, a two-stage optimization process was employed. In the first stage (runs 1–6 in Table 3), the interest was to identify an optimal training set that provides access to promising areas within the pharmacophoric space of fXa inhibitors. Three training sets were selected for stage I exploration:

Table 6

GA control parameters employed in the QSAR modeling of fXa inhibitors

| Parameter | Description | |
|-----------|--|--|
| 1 | Initial population | 100 |
| 2 | Mating | Uniform crossover |
| 3 | Probability of mating | 0.9 |
| 4 | Mutation | Uniform mutation |
| 5 | Mutation parameter | 0.7 |
| 6 | Choosing parents | Linear ranking selection |
| 9 | Number of offspring from the same parents | 2 |
| 7 | Number of the all generated offspring for population update | 60 |
| 8 | Number of replaced worst parent solutions for best offspring solutions | 6 |
| 10 | Probability for a variable to be included | 0.05 |
| 11 | Total number of generations | 4000 |
| 12 | Fitness function | Freidman's LOF (smoothing parameter = 2) |

I, II and III in Table 2. The three sets were selected in such away to guarantee maximal 3D diversity and continuous bioactivity spread over four logarithmic cycles. The “boundaries” of the investigated pharmacophoric space in stage I were set by selecting input pharmacophore features in agreement with published SAR studies and crystallographic data. For example, the fact that crystallographic information suggested the involvement of a negatively charged aspartate (within S1 pocket) in salt-bridge formation with some ligands [26] prompted us to select positive ionizable (PosIoniz) functionality as possible pharmacophoric feature. In the same manner, the following features were fed into CATALYST[®] during stage I modeling trials: HBA, HBD, HbicArom, aromatic ring (RingArom) and PosIoniz. Moreover, CATALYST[®] was restricted to evaluate 4- and 5-featured models only, which are assumed to reflect the feature-rich nature of fXa inhibitors (Table 3).

On the other hand, the boundaries of the investigated pharmacophoric space (i.e. in stage I) were extended as wide as possible by allowing the number of investigated features of any particular type to vary from 0 to 3 (Table 3), i.e. the software was allowed to investigate hypotheses that include 0, 1, 2 or 3 features of any given type (e.g. HBD, hydrophobic, etc...).

Upon careful validation of stage I hypotheses (see Section 2.3.3, Table 4), it became clearly evident that the highest-ranking 4- and 5-featured models were Hypo3/3 (run 3-model 3) and Hypo4/7 (run 4-model 7, Table 4) respectively. Both models combine superior predictive properties (SCC > 0.60) and coverage (>120 hits out of the 199 inhibitors). Interestingly, both models originated from the same set of training compounds, i.e. set II (Table 2). Presumably, the emergence of high quality models from this set is related to the optimal 3D diversity of its member compounds.

The superiority of Hypo3/3 and Hypo4/7 suggests that they represent “actual” binding features. However, being extracted from wide pharmacophoric space (by exploring 0–3 features of any particular type), both models were assumed to exhibit rough 3D properties, i.e. inter-feature distances and angles. This assumption is supported by the observed high Config. cost values of stage I models (>16.4, Table 7). Therefore, it was reasoned that further exploration is recommended to

search for models of superior 3D characters within the vicinity of Hypo3/3 and Hypo4/7, i.e. pharmacophores of identical feature types and numbers and enhanced inter-feature distances and angles. Accordingly, the pharmacophoric features of Hypo3/3 and Hypo4/7 were employed in “focused” stage II modeling (runs 7 and 9, Table 3). Furthermore, we were prompted to enroll the pharmacophoric features of Hypo4/1 and Hypo4/2 in stage II exploration (run 8, Table 3) due to their superior predictive properties (SCC ≈ 0.64, Table 4).

Consequently, the pharmacophoric features of Hypo4/1, Hypo4/2 and Hypo4/7 were employed to explore 5-featured space (runs 8 and 7, Table 3), while those of Hypo3/3 (and Hypo3/1) were utilized to investigate four-featured pharmacophores (run 9, Table 3). Additionally, stage II runs were configured to explore models of variable feature weights (Table 3). Feature weight designates the relative significance of the particular feature exerted on ligand-target affinity. Allowing the weight to vary in CATALYST[®] could create pharmacophores more suited to encode the subtle nature of the receptor-ligand interactions [75]. However, because variable weight analysis was accompanied by explosive expansion in the Config. cost, as evident in Table 7, variable weight configuration was restricted to stage II runs only. This allowed CATALYST[®] to explore variable feature weights without indulging in additional assessment of variable features.

It remains to be mentioned that pharmacophore exploration in stage II relied entirely on training set IV (Table 2), which is an extended version of II. Extending II to IV aimed at providing abundance of data-points for the regression analysis (i.e. error cost analysis) performed during CATALYST[®]'s optimization phase, which is expected to improve the quality of emerging hypotheses. Table 7 illustrates the gains in SCC values upon transition from stage I to stage II.

3.3. QSAR modeling and assessment of fXa binding site flexibility

The central goal of this research effort is to define optimal combination of orthogonal pharmacophores capable of explaining bioactivity variations among numerous, structurally diverse fXa inhibitors of bioactivities spanning several logarithmic cycles. Such orthogonal pharmacophoric combination should reflect the flexibility of fXa binding site.

GA and MLR analysis were employed to explore various pharmacophoric combinations and evaluate their statistical properties as predictive QSAR models. However, only high-ranking hypotheses of SCC ≥ 0.59 were enrolled in the analysis (17 models, Table 5) to avoid overwhelming GA–MLR with numerous inferior descriptor pharmacophores. Generally, overloading GA–MLR with poor independent variables may allow the emergence of less-than-optimal regression models [78].

Accordingly, the logarithmic transformations of bioactivity estimates derived from candidate hypotheses were enrolled as independent variables (genes) in a cycle of GA–MLR over

Table 7

The influence of transition from stage I to stage II pharmacophore exploration on the configuration costs of modeling runs and the SCC of representative high-ranking pharmacophores

| Stage | Modeling run | Config. cost | SCC of representative models |
|-------|--------------|--------------|------------------------------|
| I | 1 | 20.2274 | 0.53 (Hypo1/3) |
| | 2 | 17.4524 | 0.52 (Hypo2/2) |
| | 3 | 19.8275 | 0.66 (Hypo3/1) |
| | 4 | 16.3946 | 0.65 (Hypo4/2) |
| | 5 | 18.9709 | 0.54 (Hypo5/6) |
| | 6 | 17.2121 | 0.43 (Hypo6/1) |
| II | 7 | 28.4783 | 0.66 (Hypo7/4) |
| | 8 | 27.5906 | 0.67 (Hypo8/9) |
| | 9 | 29.3985 | 0.61 (Hypo9/6) |

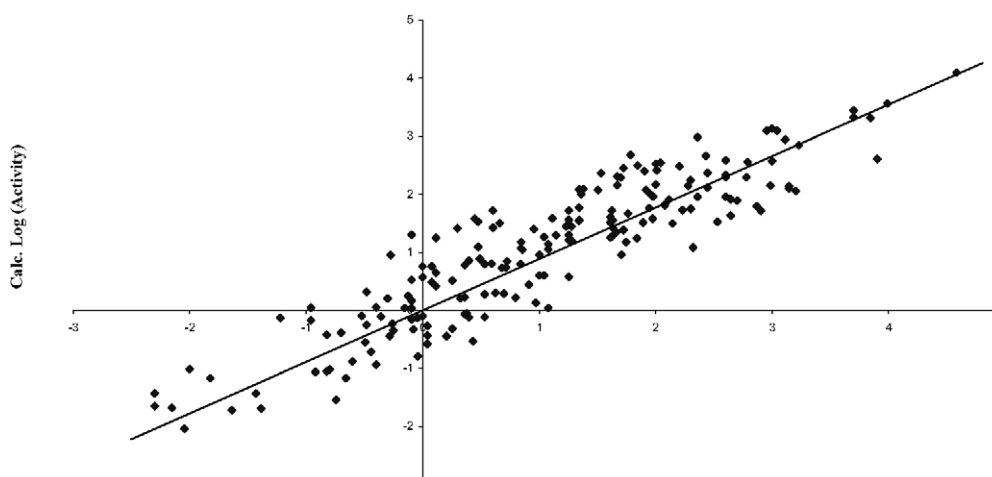


Fig. 3. A scatter plot illustrating experimental vs. estimated bioactivities of 192 fXa inhibitors (without outliers) derived from Eq. (8).

4000 iterations employing Friedman's LOF fitness criterion (see experimental, [78]). Genetic evolution yielded Eq. (7) as optimal QSAR model after removing seven statistical outliers:

$$\begin{aligned} \log(K_i) = & -0.864 [\pm 0.322] \\ & +0.237 [\pm 0.164] \times \log(\text{Hypo4/1}) \\ & +0.402 [\pm 0.195] \times \log(\text{Hypo4/2}) \\ & +0.589 [\pm 0.225] \times \log(\text{Hypo9/6}) \end{aligned}$$

$$r = 0.76, F = 86.17, n = 192 \quad (7)$$

where, r is the correlation coefficient, F is Fisher statistical parameter and n is the number of observations. The variables $\log(\text{Hypo4/1})$, $\log(\text{Hypo4/2})$ and $\log(\text{Hypo9/6})$ represent the logarithmic transformations of bioactivity estimates determined employing Hypo4/1, Hypo4/2 and Hypo9/6 respectively. The 95% confidence limits (CL) of different regression coefficients are shown in brackets ($[\pm\text{CL}]$).

The emergence of three orthogonal pharmacophoric models (see Table 8) in Eq. (8) suggests they represent three sig-

nificant binding modes accessible to ligands within the binding pocket of fXa. Figs. 4–6 show the three pharmacophores and how they map to potent, moderate and weak fXa inhibitors. Fig. 7 illustrates three superimposed conformers of the potent fXa inhibitor **71** (Table 1, Fig. 1) produced by fitting it against Hypo4/1, Hypo4/2 and Hypo9/6. This superimposition illustrates the similarity between the three pharmacophores in the way they fit the benzylamino fragment and pyrazole ring of **71** (i.e. the P1 fragment). However, the three pharmacophores exhibit significant spatial differences in the way they fit the bisphenyl part of **71** (P4 fragment). The 3D coordinates and feature weights of Hypo4/1, Hypo4/2 and Hypo9/6 are shown in Table 9.

Despite the undisputed significance of pharmacophoric hypotheses in understanding ligand-macromolecule affinity, their predictive value as 3D-QSAR models is generally limited by steric shielding, and/or bioactivity-enhancing auxiliary groups [76]. Accordingly, it is anticipated that their value would increase significantly if complemented with addi-

Table 8

Cross-correlation coefficients (R^2) among bioactivity estimates determined for the highest-ranking hypotheses (Table 5) against compounds **1–199** in Table 1. Cross-correlation coefficients of collinear models (threshold $R^2 \leq 0.65$) are highlighted in gray

| Model | Hypo 3/1 | Hypo 3/3 | Hypo 3/6 | Hypo 3/9 | Hypo 4/1 | Hypo 4/2 | Hypo 4/7 | Hypo 7/1 | Hypo 7/4 | Hypo 8/1 | Hypo 8/2 | Hypo 8/4 | Hypo 8/9 | Hypo 9/1 | Hypo 9/6 | Hypo 9/8 |
|---------|----------|-------------|-------------|----------|-------------|-------------|-------------|----------|-------------|-------------|-------------|------------|----------|-------------|----------|----------|
| Hypo3/1 | 1 | | | | | | | | | | | | | | | |
| Hypo3/3 | 0.58 | 1 | | | | | | | | | | | | | | |
| Hypo3/6 | 0.46 | 0.52 | 1 | | | | | | | | | | | | | |
| Hypo3/9 | 0.47 | 0.51 | 0.47 | 1 | | | | | | | | | | | | |
| Hypo4/1 | 0.56 | 0.54 | 0.42 | 0.6 | 1 | | | | | | | | | | | |
| Hypo4/2 | 0.59 | 0.61 | 0.58 | 0.59 | 0.64 | 1 | | | | | | | | | | |
| Hypo4/7 | 0.47 | 0.57 | 0.69 | 0.48 | 0.51 | 0.64 | 1 | | | | | | | | | |
| Hypo7/1 | 0.46 | 0.44 | 0.51 | 0.42 | 0.42 | 0.58 | 0.56 | 1 | | | | | | | | |
| Hypo7/4 | 0.52 | 0.58 | 0.61 | 0.59 | 0.59 | 0.69 | 0.74 | 0.55 | 1 | | | | | | | |
| Hypo8/1 | 0.59 | 0.61 | 0.57 | 0.63 | 0.68 | 0.78 | 0.58 | 0.57 | 0.68 | 1 | | | | | | |
| Hypo8/2 | 0.6 | 0.68 | 0.61 | 0.6 | 0.64 | 0.83 | 0.62 | 0.56 | 0.67 | 0.8 | 1 | | | | | |
| Hypo8/4 | 0.55 | 0.61 | 0.62 | 0.62 | 0.61 | 0.84 | 0.7 | 0.57 | 0.76 | 0.79 | 0.82 | 1 | | | | |
| Hypo8/9 | 0.58 | 0.67 | 0.69 | 0.61 | 0.62 | 0.84 | 0.76 | 0.61 | 0.77 | 0.78 | 0.85 | 0.9 | 1 | | | |
| Hypo9/1 | 0.48 | 0.48 | 0.45 | 0.47 | 0.53 | 0.56 | 0.43 | 0.39 | 0.5 | 0.62 | 0.56 | 0.54 | 0.56 | 1 | | |
| Hypo9/6 | 0.47 | 0.44 | 0.25 | 0.36 | 0.36 | 0.35 | 0.22 | 0.37 | 0.28 | 0.43 | 0.41 | 0.38 | 0.37 | 0.41 | 1 | |
| Hypo9/8 | 0.53 | 0.52 | 0.53 | 0.55 | 0.58 | 0.62 | 0.49 | 0.47 | 0.53 | 0.67 | 0.62 | 0.62 | 0.64 | 0.77 | 0.45 | 1 |

Table 9

Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of Hypo4/1, Hypo4/2 and Hypo9/6

| Model | Definitions | Chemical features | | | | | | | |
|----------------|-------------|-------------------|-------|------------|-------|-----------------|-----------------|-----------------|-------|
| Hypo4/2 | Weights | HBA | | HBA | | HbicArom | HbicArom | PosIoniz | |
| | | 2.47711 | | 2.47711 | | 2.47711 | 2.47711 | 2.47711 | |
| | Tolerances | 1.6 | 2.2 | 1.6 | 2.2 | 1.6 | 1.6 | 1.6 | |
| | Coordinates | X | -7.65 | -8.63 | 5.2 | 7.69 | -6.18 | -2.06 | 2.47 |
| | | Y | 0.99 | 3.79 | 1.82 | 3.32 | -2.62 | 0.1 | -2.78 |
| Z | | -1.96 | -2.42 | 0.08 | -0.68 | 0.64 | 0.2 | -2.44 | |
| Hypo4/1 | Weights | 3.04062 | | 3.04062 | | 3.04062 | 3.04062 | 3.04062 | |
| | Tolerances | 1.6 | 2.2 | 1.6 | 2.2 | 1.6 | 1.6 | 1.6 | |
| | Coordinates | X | -4.13 | -4.57 | 5.25 | 7.56 | -5.88 | 5.69 | 4.27 |
| | | Y | 0.39 | 3.01 | 2.14 | 3.02 | -3.42 | -0.98 | -3.71 |
| | | Z | 4.67 | 6.05 | 1.3 | -0.43 | 1.9 | -0.86 | -1.44 |
| Hypo9/6 | Weights | 2.49146 | | 3.2389 | | 1.74402 | | 2.49146 | |
| | Tolerances | 1.6 | 2.2 | 1.6 | 2.2 | 1.6 | 1.6 | 1.6 | |
| | Coordinates | X | 2.17 | 2.62 | -4.02 | -3.07 | 1.3 | 4.2 | 5.38 |
| | | Y | -4.93 | -7.44 | 2.66 | 5.41 | 1.08 | 1.45 | 0.51 |
| | | Z | -5.34 | -6.98 | 2.17 | 3.07 | 1.82 | 2.47 | -4.13 |

tional descriptors within the context of traditional QSAR analysis. Therefore, we were prompted to enroll additional molecular descriptors, calculated using Alchemy2000[®] and QSARIS[®], as independent variables in the GA-MLR-QSAR modeling. The most significant QSAR model was achieved after 4000 iterations of GA-MLR analysis followed by excluding statistical outliers [78]. The resulting QSAR model was cross-validated automatically using leave-one-out cross-validation [78]. Eq. (8) shows the resulting QSAR model, while Fig. 3 shows the corresponding scatter plot of experimental versus estimated bioactivities of 192 fXa inhibitors (without outliers).

$$\log(K_i) = -6.586 [\pm 3.131] + 0.194 [\pm 0.129] \times \log(\text{Hypo4/1}) + 0.373 [\pm 0.1587] \times \log(\text{Hypo4/2}) + 0.389 [\pm 0.177] \times \log(\text{Hypo9/6}) - 0.3189 [\pm 0.096] \times \text{SdsCH} + 0.219 [\pm 0.108] \times \text{SssssC} - 0.0530 [\pm 0.035] \times \text{SaaN} - 0.183 [\pm 0.105] \times \text{SsCH}_3\text{-acnt} - 13.570 [\pm 4.230] \times \text{MaxHp} + 0.395 [\pm 0.156] \times \text{SHBint2_Acnt} - 0.111 [\pm 0.050] \times \text{SHBint8_Acnt} - 0.0827 [\pm 0.049] \times \text{SHBint10_Acnt} - 0.200 [\pm 0.073] \times \text{SHHBd} + 0.537 [\pm 0.197] \times \text{Ioniz Pot} - 0.721 [\pm 0.306] \times \text{HOMO} + 0.322 [\pm 0.155] \times \text{Py} + 49.520 [\pm 15.125] \times \chi^{10}_{\text{ch}} + 1.628 [\pm 0.277] \times \text{knotpv}$$

$$r = 0.91, F = 49.64, \text{Multiple } Q^2 = 0.79, n = 192 \quad (8)$$

where, multiple Q^2 is the leave-one-out correlation coefficient, SdsCH is the sum of all =CH- topological E-state values in a molecule, SssssC is the sum of all E-state values of quaternary carbon atoms in a molecule, SaaN is the sum of all E-state values of aromatic nitrogens, SsCH₃-acnt is the count of all CH₃- groups in a molecule, MaxHp is the largest positive charge on a hydrogen atom, SHBint2_Acnt, SHBint8_Acnt, SHBint10_Acnt are the sums of E-state descriptors of strength for potential hydrogen bonds of path lengths 2, 8 and 10, respectively, SHHBd is the E-state indices for HBDs, Ioniz Pot is the ionization potential, HOMO is

the energy of the highest occupied molecular orbital and Py is the component of the dipole moment along the inertial Y-axis, χ^{10}_{ch} is the simple 10th order chain chi index and knotpv is the difference between chi valence cluster-3 and chi valence path/cluster-4 [78]. Reemergence of Hypo4/1, Hypo4/2 and Hypo9/6 in Eq. (8), despite evolutionary competition imposed by the newly added physicochemical descriptors, strongly emphasizes the statistical significance of this pharmacophoric combination. However, it is clearly evident from Eq. (8) that the new descriptors improved the overall statistical criteria of the model (i.e. compared to Eq. 7). In particular, Eq. (8) illustrated sufficient robustness to withstand the leave-one-out cross-validation procedure contrary to Eq. (7). It remains to be mentioned that the optimal pharmacophoric combination in Eq. (8) incorporates only one hypothesis from stage II modeling trials (i.e. Hypo9/6, see Section 3.2), while it includes two models from stage I trials (Hypo4/1 and Hypo4/2). Nevertheless, despite the apparent inefficiency of stage II pharmacophore exploration, our overall two-stage explorative strategy provided abundance of high quality models for evaluation by GA-MLR-QSAR modeling, and thus allowed access to better characterization of ligand flexibility within fXa binding site.

3.4. Comparison of pharmacophore model with the active site of fXa

Despite the problems associated with crystallographic structures (see Section 1), pharmacophore features obtained by CATALYST[®] can be compared with crystallographic fXa active site to identify probable residues important for activity [83]. However, one should not use these pharmacophores as receptor maps [84]. The features in Hypo4/1, Hypo4/2 and Hypo9/6 as well as the alignment of compound **71** as proposed by these hypotheses (Fig. 4a, Fig. 5a and Fig. 6a) were compared with the structure of **71** docked within the binding

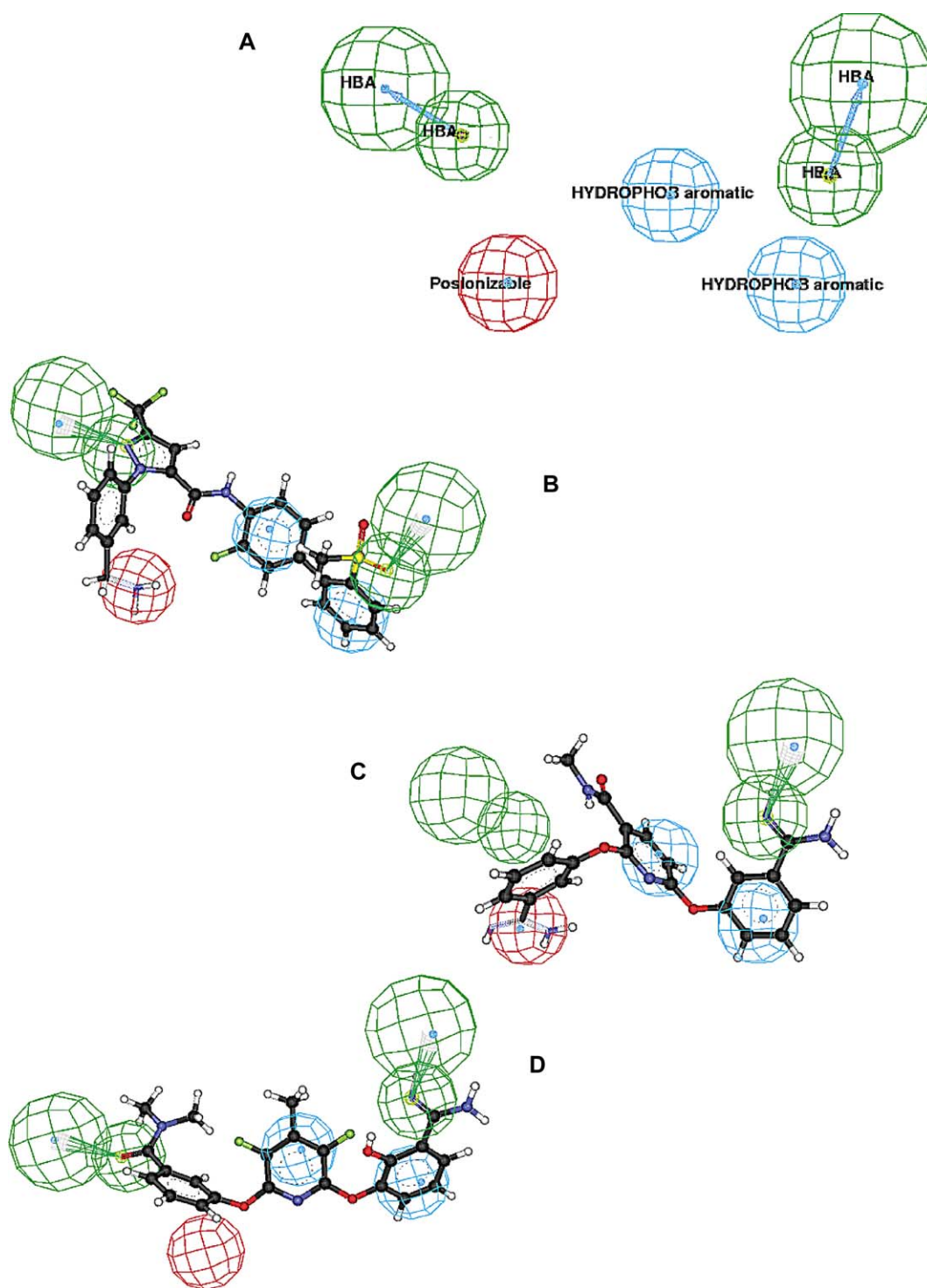


Fig. 4. (A) Hypo4/2 mapped against (B) **71** (potent, $K_i = 0.15$ nM); (C) **100** (moderately potent, $K_i = 230$ nM); and (D) **114** (poorly potent, $K_i > 5000$ nM). (HBA: Hydrogen Bond Acceptor, PosIonizable: Positive Ionizable).

site of fXa (Fig. 8). The docking experiment was performed employing Cerius2.LIGANDFIT [85] (Accelrys Inc.). A marked similarity was observed between features proposed by the pharmacophore models and ligand binding features in the docked structure. The benzylamine part (P1 fragment) of the docked conformer (likely to be protonated at physiological pH) is fitted within a cavity (S1 pocket) formed primarily by the amino acid residues ASP 189, TRP 215, CYS 191,

CYS 220 and GLN 192. This pocket has a negative electrostatic potential because of the acidic amino acid ASP 189 (Fig. 8) and hence is a probable site for binding of the PosIoniz benzylamine group. Furthermore, the hydrophobic amino acids TRP 215, CYS 191 and CYS 220 furnish suitable environment for hydrophobic interactions with the aromatic ring of the benzylamine fragment. The hydroxyl moiety of SER 195 seems to provide excellent opportunity for hydrogen

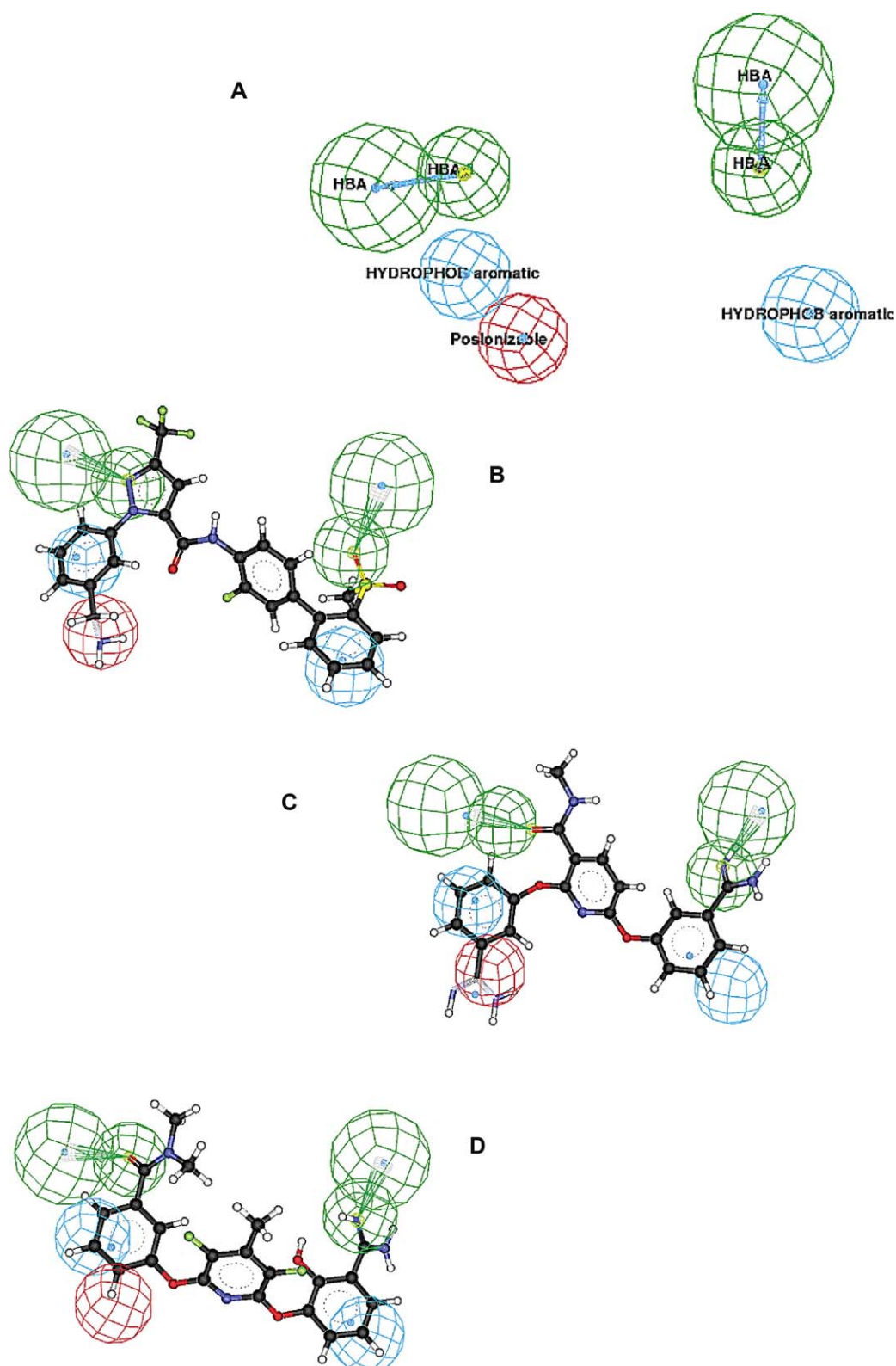


Fig. 5. (A) Hypo4/1 mapped against (B) **71** (potent $K_i = 0.15$ nM); (C) **100** (moderately potent, $K_i = 230$ nM), and (D) **114** (poorly potent, $K_i > 5000$ nM). (HBA: Hydrogen Bond Acceptor, PosIonizable: Positive Ionizable).

bonding with the nitrogen atom of the pyrazole linker of **71**. Finally, the bisphenyl fragment of **71** (P4 group) fits nicely within the aromatic environment of the S4 pocket furnished by PHE 174, TRP 215 and TYR 99 (Fig. 8). The mappings of

compound **71** against Hypo4/1, Hypo4/2 and Hypo9/6 are shown in Fig. 4a, Fig. 5a and Fig. 6a. Each pharmacophore illustrates a HBA feature and a PosIoniz function corresponding to the P1 fragment and the nearby pyrazole linker of com-

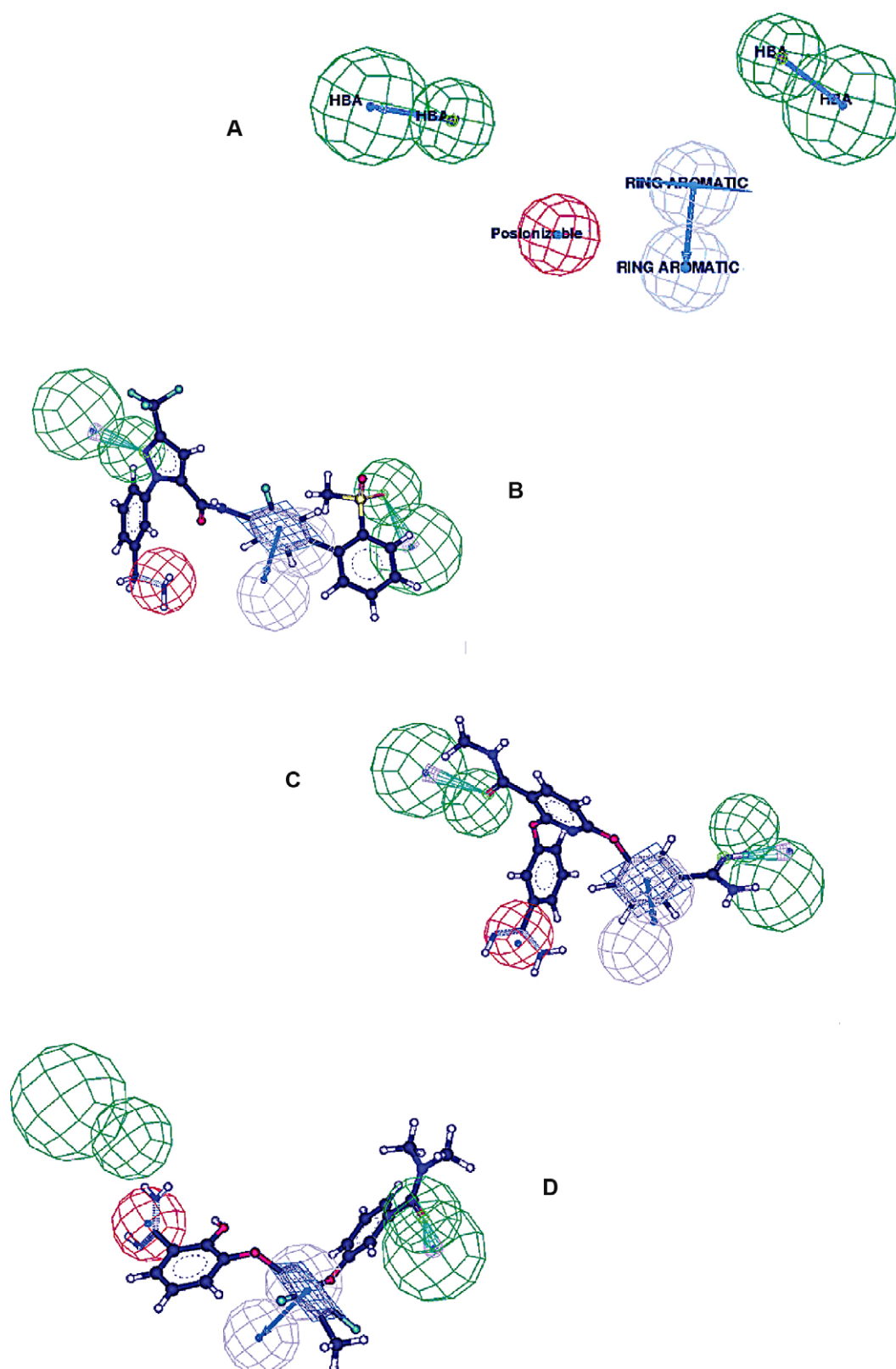


Fig. 6. (A) Hypo9/6 mapped against (B) **71** (potent $K_i = 0.15$ nM); (C) **100** (moderately potent, $K_i = 230$ nM); and (D) **114** (poorly potent, $K_i > 5000$ nM). (HBA: Hydrogen Bond Acceptor, PosIonizable: Positive Ionizable).

compound **71**. Furthermore, Hypo4/1 has one hydrophobic feature corresponding to the aromatic ring of the P1 fragment of **71**. Moreover, the pharmacophoric combination exhibits

hydrophobic features (Hypo4/1 and Hypo4/2) or aromatic ring function (Hypo9/6) corresponding to the HbicArom environment of the S4 pocket. Interestingly, each of the three phar-

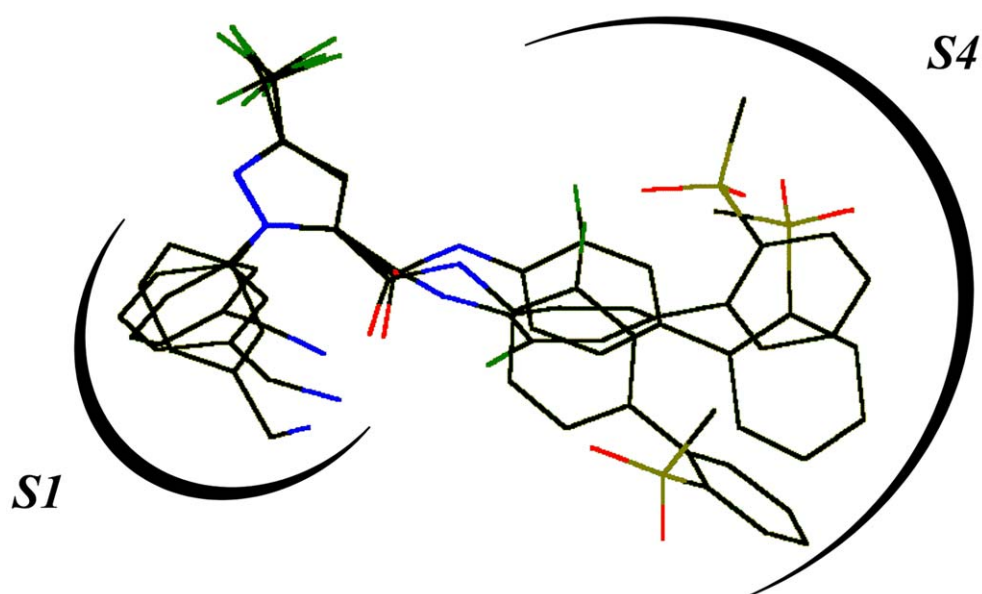


Fig. 7. A wire-frame view illustrating three conformers of inhibitor **71** (Table 1) fitted against Hypo4/1, Hypo4/2 and Hypo9/6. The figure shows the molecular fragments of compound **71** that correspond to S1 and S2 pockets within human fXa binding site. The superimposition clearly suggests flexible binding mode (blue: nitrogen, red: oxygen, black: carbon, green: fluorine, yellow: sulfur).

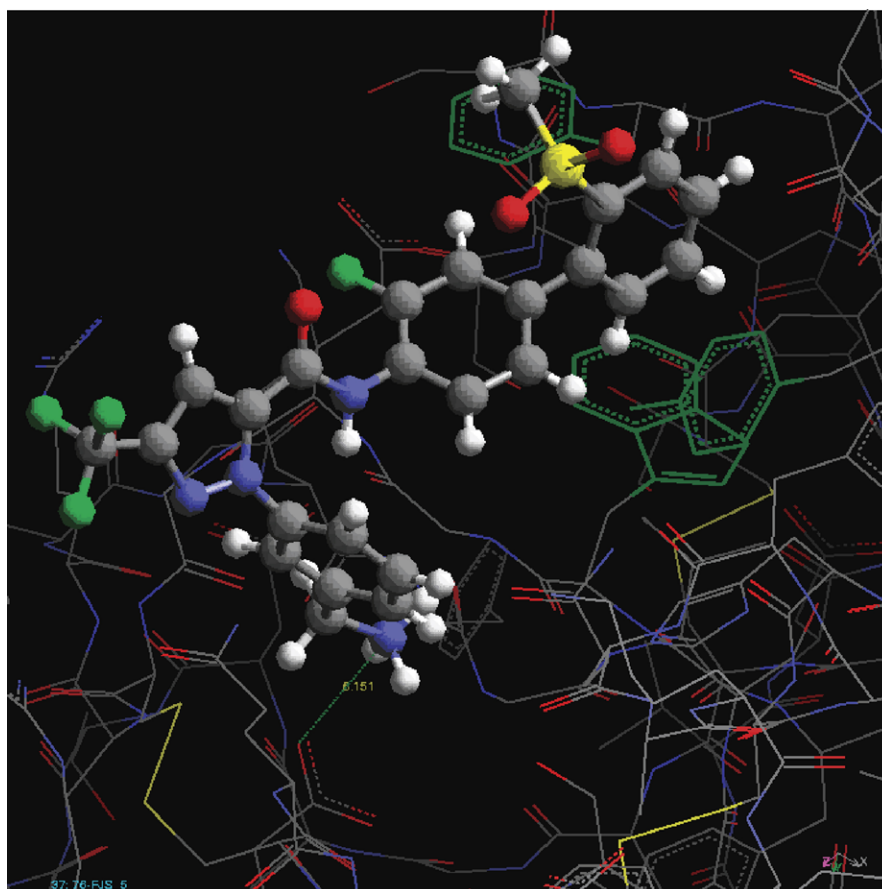


Fig. 8. A wire-frame view illustrating compounds **71** docked to the binding site of fXa (PDB code: 1FJS, resolution = 1.92 Å). The docking experiment was performed employing C2.LIGANDFIT[®] software (from Accelrys Inc.). Hydration water molecules and protein hydrogen atoms were hidden for clarity. The green aromatic rings resemble S4 pocket, while the carboxylate anion connected to the ligand's ammonium group is the aspartate moiety of S1 pocket (ASP 189).

macophors illustrates one hydrogen-bond acceptor feature corresponding to the sulfone moiety of compound **71**. Apparently, this feature refers to a hydrogen-bonding interaction between one of the sulfone oxygens in compound **71** and a water molecule attached to TYR 99. Thus the 3D array of features is consistent with the proposed docking of compound **71** in the crystal structure of fXa.

4. Conclusion

This presented effort includes extensive exploration of the pharmacophoric space of fXa inhibitors utilizing CATALYST[®]. The exploration process included evaluating variety of training sets, input pharmacophoric features and run parameters. The hypothetical pharmacophors were assessed as a predictive 3D-QSAR models and database search queries. The validation process identified 17 plausible pharmacophore models. However, subsequent QSAR modeling utilizing GA and MLR analysis identified three complementary pharmacophores as representatives of the accessible binding modes within fXa binding site. The significance of this approach originates from the fact that the proposed binding modes were derived from the experimental affinities of respective fXa ligands and not rigid fXa crystallographic structures. Furthermore, CATALYST[®] pharmacophores are more appropriate for searching massive 3D compound libraries [67] than the time consuming flexible docking techniques such as molecular dynamics simulation.

Acknowledgments

The authors wish to thank Hamdi-Mango Research Fund and the Deanship of Scientific Research at Jordan University for providing funds towards purchasing O2 and Octane2 Sgi workstations and CATALYST[®] and CERIU2[®] software packages. The authors are also grateful to Professor Mohammad Zgoul for his support.

References

- [1] R.J. Leadley, *Curr. Top. Med. Chem.* 1 (2001) 151–159.
- [2] K. Sato, T. Kawasaki, Y. Taniuchi, F. Hirayama, H. Koshio, Y. Matsumoto, *Eur. J. Pharmacol.* 339 (1997) 141–146.
- [3] F. Koller, *Thromb. Diath. Haemorr.* 4 (1959) 58–65.
- [4] D.N. Cooper, D.S. Millar, A. Wacey, S. Pemberton, E.G.D. Tuddenham, *Thromb. Haemost.* 78 (1997) 161–172.
- [5] K.G. Mann, *Trends Biochem. Sci.* 12 (1987) 229–233.
- [6] P.R. Eisenberg, B.E. Sobel, A.S. Jaffe, *J. Am. Coll. Cardiol.* 19 (1992) 1065–1069.
- [7] W. Wysokinski, R. McBane, J.H. Chesebro, W.G. Owen, *In Vivo Thromb., Haemost.* 76 (1996) 1108–1113.
- [8] F. Al-Obeidi, J.A. Ostrem, *Expert Opin. Ther. Pat.* 74 (1999) 635–639.
- [9] W.R. Ewing, H.W. Pauls, A.P. Spada, *Drugs Future* 24 (1999) 771–787.
- [10] B.-Y. Zhu, R.M. Scarborough, *Annu. Rep. Med. Chem.* 35 (2000) 83–102.
- [11] J.M. Fevig, R.R. Wexler, *Annu. Rep. Med. Chem.* 34 (1999) 81–100.
- [12] R. Rai, P.A. Sprengler, K.C. Elrod, W.B. Young, *Curr. Med. Chem.* 8 (2001) 101–119.
- [13] W.R. Ewing, V. Mikol, *Curr. Top. Med. Chem.* 1 (2001) 161–174.
- [14] H. Matter, E. Defossa, U. Heinelt, P.M. Blohm, D. Schneider, A. Müller, et al., *J. Med. Chem.* 45 (2002) 2749–2769.
- [15] D.J.P. Pinto, M.J. Orwat, S. Wang, J.M. Fevig, M.L. Quan, E. Amparo, et al., *J. Med. Chem.* 44 (2001) 566–578.
- [16] M.L. Quan, A.Y. Liauw, C.D. Ellis, J.R. Pruitt, D.J. Carini, L.L. Bostrom, et al., *J. Med. Chem.* 42 (1999) 2752–2759.
- [17] M.L. Quan, C.D. Ellis, A.Y. Liauw, R.S. Alexander, R.M. Knabb, G.N. Lam, et al., *J. Med. Chem.* 42 (1999) 2760–2773.
- [18] R. Kucznierz, F. Grams, H. Leinert, K. Marzenell, R.A. Engh, W. Von der Saal, *J. Med. Chem.* 41 (1998) 4983–4994.
- [19] C.K. Marlowe, S. Uma, A.C. Gunn, R.M. Scarborough, *Bioorg. Med. Chem. Lett.* 10 (2000) 13–16.
- [20] Y. Gong, H.W. Pauls, A.P. Spada, M. Czeka, G. Liang, V. Chu, et al., *Bioorg. Med. Chem. Lett.* 10 (2000) 217–221.
- [21] W. He, B. Hanney, M.R. Myers, A.R. Spada, K. Brown, D. Colussi, et al., *Bioorg. Med. Chem. Lett.* 15 (2000) 1737–1739.
- [22] D.A. Dudley, A.M. Bunker, L. Chi, W.L. Cody, D.R. Holland, D.P. Ignasiak, et al., *J. Med. Chem.* 43 (2000) 4063–4070.
- [23] J.J. Masters, J.B. Franciskovich, J.M. Tinsley, C. Campbell, J.B. Campbell, T.J. Craft, et al., *J. Med. Chem.* 43 (2000) 2087–2092.
- [24] D.K. Herron, T. Goodson, M.R. Wiley, L.C. Weir, J.A. Kyle, Y.K. Yee, et al., *J. Med. Chem.* 43 (2000) 859–872.
- [25] M.R. Wiley, L.C. Weir, S. Briggs, N.A. Bryan, J. Buben, C. Campbell, et al., *J. Med. Chem.* 43 (2000) 883–899.
- [26] S. Maignan, V. Mikol, *Curr. Top. Med. Chem.* 1 (2001) 161–174.
- [27] K. Padmanabhan, K.P. Padmanabhan, A. Tulinsky, C.H. Park, W. Bode, R. Huber, et al., *J. Mol. Biol.* 232 (1993) 947–966.
- [28] H. Brandstetter, A. Kuhne, W. Bode, R. Huber, W. Von der Saal, K. Wirthensohn, et al., *J. Biol. Chem.* 271 (1996) 29988–29992.
- [29] D.A. Dougherty, *Science* 271 (1996) 163–168.
- [30] J.P. Gallivan, D.A. Dougherty, *Proc. Natl. Acad. Sci. USA* 96 (1999) 9459–9464.
- [31] S. Maignan, J.P. Guilloteau, S. Pouzieux, Y.M. Choi-Sledeski, M.R. Becker, S.I. Klein, et al., *J. Med. Chem.* 43 (2000) 3226–3232.
- [32] Y.K. Yee, A.L. Tebbe, J.H. Linebarger, D.W. Beight, T.J. Craft, D. Gifford-Moore, et al., *J. Med. Chem.* 43 (2000) 873–882.
- [33] B.O. Buckman, R. Mohan, S. Koovakkat, A. Liang, L. Trinh, M.M. Morrissey, *Bioorg. Med. Chem. Lett.* 8 (1998) 2235–2237.
- [34] R.A. Galemno Jr., B.L. Wells, K.A. Rossi, R.S. Alexander, C. Dominguez, T.P. Maduska, et al., *Bioorg. Med. Chem. Lett.* 10 (2000) 301–304.
- [35] Q. Han, C. Dominguez, P.F. Stouten, J.M. Park, D.E. Duffy, R.A. Galemno Jr., et al., *J. Med. Chem.* 43 (2000) 4398–4415.
- [36] Y.M. Choi-Sledeski, D.G. McGarry, D.M. Green, H.J. Mason, M.R. Becker, R.S. Davis, et al., *J. Med. Chem.* 42 (1999) 3572–3587.
- [37] K.J. Shaw, W.J. Guilford, J.L. Dallas, S.K. Koovakkaat, M.A. McCarrick, A. Liang, et al., *J. Med. Chem.* 41 (1998) 3551–3556.
- [38] W.J. Guilford, K.J. Shaw, J.L. Dallas, S. Koovakkat, W. Lee, A. Liang, et al., *J. Med. Chem.* 42 (1999) 5415–5425.
- [39] M. Renatus, W. Bode, R. Huber, J. Sturzebecher, M.T. Stubbs, *J. Med. Chem.* 41 (1998) 5445–5456.
- [40] G. Phillips, B.O. Buckman, D.V. Davey, K.A. Eagen, W.J. Guilford, J. Hinchman, et al., *J. Med. Chem.* 41 (1998) 3557–3562.
- [41] M.L. Quan, R.R. Wexler, *Curr. Top. Med. Chem.* 1 (2001) 137–149.
- [42] D.R. Light, W.J. Guilford, *Curr. Top. Med. Chem.* 1 (2001) 121–136.
- [43] H.W. Pauls, W.R. Ewing, *Curr. Top. Med. Chem.* 1 (2001) 83–100.
- [44] X. Daura, E. Haaksma, W.F. Van Gunsteren, *J. Comput. Aided Mol. Des.* 14 (2000) 507–529.
- [45] D.E. Koshland, *Proc. Natl. Acad. Sci. USA* 44 (1958) 98–104.
- [46] W.L. Jorgensen, *Science* 254 (1991) 951–955.
- [47] A. Morton, B.W. Matthews, *Biochemistry* 34 (1995) 8576–8588.
- [48] P.R. Andrews, D.J. Craik, J.L. Martin, *J. Med. Chem.* 27 (1984) 1648–1657.

- [49] N.C. Cohen, *Guidebook on Molecular Modeling in Drug Design*, Academic Press, UK, 1996.
- [50] M.L. Teodoro, L.E. Kavraki, *Curr. Pharm. Des.* 9 (2003) 1635–1648.
- [51] R.L. Dunbrack, M. Karplus, *J. Mol. Biol.* 230 (1993) 543–574.
- [52] G. Vriend, C. Sander, P.F.W. Stouten, *Prot. Eng.* 7 (1994) 23–29.
- [53] H. Shrauber, F. Eisenhaber, P. Argos, *J. Mol. Biol.* 230 (1993) 592–612.
- [54] A.R. Leach, I.D. Kuntz, *J. Comput. Chem.* 13 (1992) 730–748.
- [55] A.R. Leach, *J. Mol. Biol.* 235 (1994) 345–356.
- [56] F. Eisenmenger, P. Argos, R. Abagyan, *J. Mol. Biol.* 231 (1993) 849–860.
- [57] P. Kollman, *Curr. Opin. Struct. Biol.* 4 (1994) 240–245.
- [58] J.A. McCammon, S.C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, 1987.
- [59] R.M.A. Knechtel, I.D. Kuntz, C.M. Oshiro, *J. Mol. Biol.* 266 (1997) 424–440.
- [60] J.A. Erickson, M. Jalaie, D.H. Robertson, R.A. Lewis, M. Vieth, *J. Med. Chem.* 47 (2004) 45–55.
- [61] N.R.A. Beeley, C. Sage, *Targets* 2 (2003) 19–25.
- [62] M. Adler, D.D. Davey, G.B. Phillips, S.H. Kim, J. Jancarik, G. Rumennik, et al., *Biochemistry* 39 (2000) 12534–12542.
- [63] A.L. Harvey, *Advances in Drug Discovery Techniques*, Wiley, UK, 1998.
- [64] R.A. Silverman, *The Organic Chemistry of Drug Design and Drug Action*, Academic Press, San Diego, 1991.
- [65] C.A.T.A.L.Y.S.T., *Computational Results Obtained Using Software Program, 3.1*, Molecular Simulations Inc., San Diego, CA, 1996.
- [66] CATALYST, 4.6 users' manual, Accelrys Inc., San Diego, CA, 2002.
- [67] Y. Kurogi, O.F. Güner, *Curr. Med. Chem.* 8 (2001) 1035–1055.
- [68] P.W. Sprague, R. Hoffmann, in: H. Van de Waterbeemd, B. Testa, G. Folkers (Eds.), *Computer Assisted Lead Finding and Optimization—Current Tools for Medicinal Chemistry*, VHCA, Basel, 1997, pp. 230–240.
- [69] D. Barnum, J. Greene, A. Smellie, P. Sprague, *J. Chem. Inf. Comput. Sci.* 36 (1996) 563–571.
- [70] P.A. Greenidge, B. Carlson, L. Bladh, M. Gillner, *J. Med. Chem.* 41 (1998) 2503–2512.
- [71] T. Langer, R.D. Hoffmann, *J. Chem. Inf. Comput. Sci.* 38 (1998) 325–330.
- [72] P.A. Keller, M. Bowman, K.H. Dang, J. Garner, S.P. Leach, R. Smith, et al., *J. Med. Chem.* 42 (1999) 2351–2357.
- [73] R.G. Karki, V.M. Kulkarni, *Eur. J. Med. Chem.* 36 (2001) 147–163.
- [74] H. Li, J. Sutter, R. Hoffmann, in: O.F. Güner (Ed.), *Pharmacophor Perception, Development, and Use in Drug Design*, International University Line, California, 2000, pp. 173–189.
- [75] J. Sutter, O. Güner, R. Hoffmann, H. Li, M. Waldman, in: O.F. Güner (Ed.), *Pharmacophor Perception, Development, and Use in Drug Design*, International University Line, California, 2000, pp. 501–511.
- [76] I.B. Bersuker, S. Bahçeci, J.E. Boggs, in: O.F. Güner (Ed.), *Pharmacophor Perception, Development, and Use in Drug Design*, International University Line, California, 2000, pp. 457–473.
- [77] L.H. Hall, L.B. Kier, M.L. Hall, *QSAR Development with QSARIS. The Guide for Development of QSAR Models with QSARIS*, Academic Press, UK, 2001.
- [78] *QSARIS® Reference Guide: Statistical Analysis and Molecular Descriptors SciVision*, Academic Press, Massachusetts, 2000.
- [79] P.J. Angeline, *IEEE Expert Intell. Syst. Appl.* 10 (1995) 6–10.
- [80] A. Smellie, S.D. Kahn, S.L. Teig, *J. Chem. Inf. Comput. Sci.* 35 (1995) 285–294.
- [81] A. Smellie, S. Teig, P. Towbin, *J. Comp. Chem.* 16 (1995) 171–187.
- [82] R.P. Sheridan, S.K. Kearsley, *DDT* 7 (2002) 903–910.
- [83] S.S. Kulkarni, V.M. Kulkarni, *J. Med. Chem.* 42 (1999) 373.
- [84] R.D. Cramer III, D.E. Patterson, J.D. Bunce, *J. Am. Chem. Soc.* 110 (1988) 5959.
- [85] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, *J. Mol. Graph. Model.* 21 (2003) 289–307.